Business Problem Document

Business Problem Statement

Understanding customer order behavior and revenue distribution is a critical requirement for any e-commerce platform. However, when real user traffic is unavailable, analysts must rely on controlled experiments using synthetic data to validate analytical workflows, business logic, and insight-generation techniques.

In this project, an AI-generated synthetic dataset was used to experiment with customer order and revenue analysis, resembling the type of transactional data an e-commerce website could produce. The purpose of the analysis was not to represent real business performance, but to explore how transactional data can be cleaned, analyzed, and visualized to answer common business questions related to sales, product performance, and customer demographics.

The project focuses on applying real-world analytical methods—data cleaning, SQL-based querying, and dashboarding—to a simulated dataset in order to demonstrate end-to-end data analysis skills.

Primary Analytical Question

How can synthetic transactional data be explored and analyzed to evaluate revenue patterns, product performance, and customer demographics using industry-standard data analytics tools?

Objectives

- Experiment with transactional data resembling e-commerce orders
- Apply data cleaning and feature engineering techniques
- Practice revenue and product performance analysis using SQL
- Explore demographic and geographic patterns in customer orders
- Visualize analytical insights using dashboards
- Validate an end-to-end analytics workflow on non-production data

Deliverables

- **Experimental Data Preparation (Python):** Cleaned and transformed 200 AI-generated synthetic order records, handled missing values, standardized categorical fields, engineered age groups, and calculated net revenue.
- **Exploratory Data Analysis (PostgreSQL):** Stored synthetic data in PostgreSQL and executed queries to explore revenue distribution, product performance, demographic and geographic trends.
- **Data Visualization (Power BI):** Built interactive dashboards visualizing sales and revenue distribution, product performance, and customer demographics.
- **Application-Level Data Generation (Next.js & PostgreSQL):** Built a custom e-commerce platform to generate synthetic transactional data for analytics experimentation.
- **Documentation & Repository:** Maintained a GitHub repository containing synthetic data, Python scripts, SQL queries, dashboards, and documentation.

Project Scope Disclaimer

- The dataset used in this project is synthetic and AI-generated
- No real customers or live business traffic were involved
- Insights are exploratory and experimental in nature
- The project demonstrates analytical methodology, not business outcomes