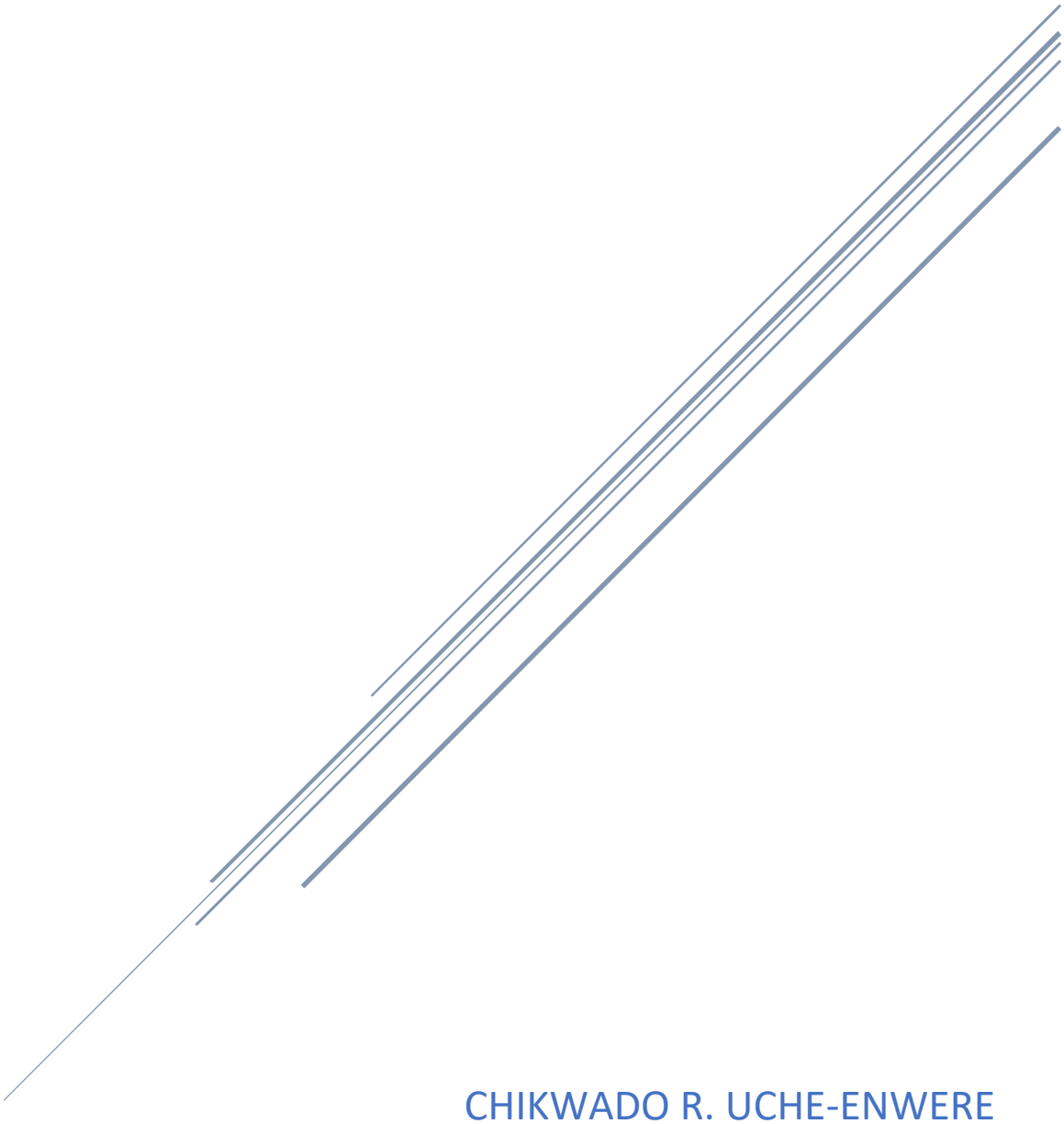
ANALYSIS OF ONLINE RETAIL STORE

PROJECT OVERVIEW



CHIKWADO R. UCHE-ENWERE

PORTFOLIO PROJECT

PROJECT OVERVIEW – Online Retail Data Analysis (SQL Based)

This project is focused on ***cleaning***, ***structuring***, ***normalizing*** and ***analyzing*** transactional online retail data using purely **SQL**. The original dataset is a CSV file that was sourced from Kaggle https://www.kaggle.com/datasets/tunguz/online-retail and contains information on customers transactions over a span of one year (from December 2010 to December 2011). The dataset was originally provided as a flat file that contained information on invoices, stockcode (which is the product code), customer's ID, description of the product, country where the customers reside, price, quantity of purchase and date of invoice. It also contained inconsistencies, missing values, and cancelled/returned orders.

The goal of this project is to transform messy e-Commerce raw file strictly using SQL (PostgreSQL) into a clean data with structured relational schemas where meaningful insights that mimic real life situations can be derived.

PROJECT OBJECTIVE

- Explore and clean the messy data
- Normalize the flat CSV file into structured SQL tables such as ORDERS table, CANCELLED table, INVOICES table, PRODUCTS table, CUSTOMERS table.
- Implement FOREIGN KEYS and Constraints to provide data quality and referential basis.
- Create a CANCELLED table to properly isolate returned/cancelled orders.
- Populate a revenue column with the ORDERS and CANCELLED table.
- Answer key business questions through SQL queries.

PROJECT PHASE

The project would be divided into two phases:

- Phase 1: this would deal with cleaning the data and creating schemas
- Phase 2: This tackles business questions and possible visualization using Power Bi.

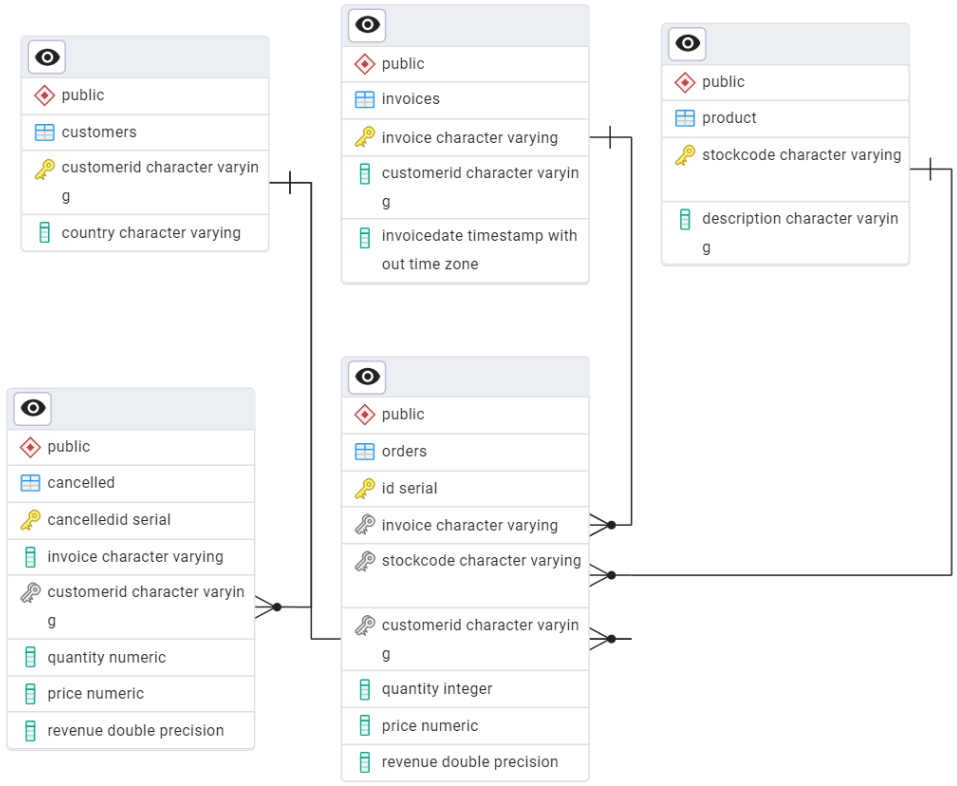
TABLES CREATED

Tables created for this project are

Table Name	Description
Orders	All purchase orders where quantity > 0 and contains invoiceID, stockid, quantity, price, and customerid
Cancelled	All purchase orders where quantity < 0
Customers	Each row contains unique ID of the customer with the most frequent country of that customer

Invoices	Each row contains unique invoice ID with the date the invoices were issued
Products	Each row contains unique invoice ID with the date the invoices were issued

SCHEMA



TOOLS

The tools used are

- ✓ PostgreSQL for data manipulation and querying the database
- ✓ PgAdmin4 GUI.

RESOURCES

- Data source: *https://www.kaggle.com/datasets/tunguz/online-retail*
- SQL for Data Analysis *by Cathy Tanimura*
- Stackoverflow