

**DEBRE BERHAN UNIVERSITY**

**INSTITUTION Of TECHNOLOGY**

**COLLAGE Of COMPUTING**

**DEPARTMENT Of SOFTWARE ENGINEERING**

**FUNDAMENTALOFBIGDATAANALYTICSANDBUSINESS INTELLIGENCY(SEng5112)**

**Prepared by: -**

**Name: Ephrem Tilahun ID No DBUR/0094/13**

**Submitted to: Mr. Derbew Felasman(MSc) February, 2025 Debre Berhan, Ethiopia**

Contents

[1. Introduction 1](#_Toc189924791)

[2. Data Source & Understanding 1](#_Toc189924792)

[3. Data Extraction 1](#_Toc189924793)

[4.Data Cleaning & Transformation 2](#_Toc189924794)

[5.PostgreSQL Schema Design 2](#_Toc189924795)

[6. Loading Data using Python 3](#_Toc189924796)

[7. Visualization and Insights 4](#_Toc189924797)

# **1. Introduction**

This report documents the implementation of an End-to-End Data Pipeline for an e-commerce dataset. The pipeline processes raw transaction data, loads it into a PostgreSQL database, and visualizes key insights using Power BI.

# **2. Data Source & Understanding**

**2.1 Dataset Overview**

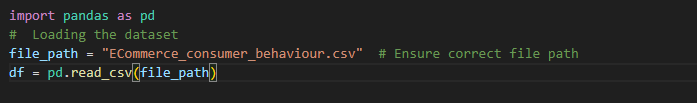
* File Name: **ECommerce\_consumer\_behaviour.csv**
* Size: 1 ,004 ,756 rows
* Columns: order\_id, user\_id, order\_number, order\_dow, order\_hour\_of\_day, days\_since\_prior\_order, product\_id, add\_to\_cart\_order, reordered, department\_id, department, product\_name
* Objective: Analyze customer purchasing behavior and sales trends.

**2.2 Identified Data Issues**

* Missing values in days\_since\_prior\_order
* Duplicate records
* Inconsistent text formatting (department & product names)
* Data types requiring conversion (order\_hour\_of\_day, order\_dow)

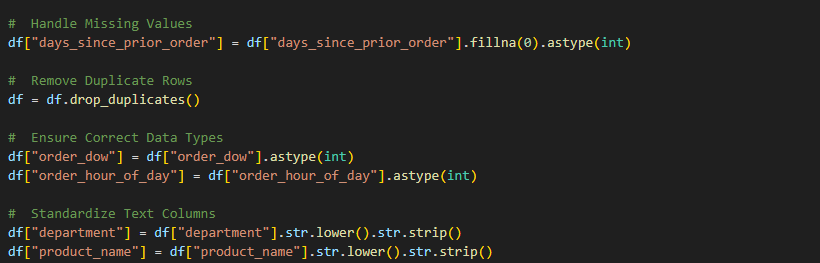
# **3. Data Extraction**

Used pandas to load the CSV file into a DataFrame:

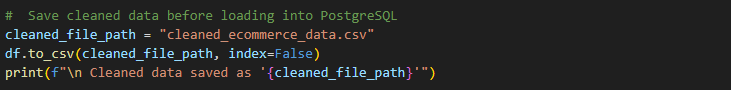


# **4.Data Cleaning & Transformation**

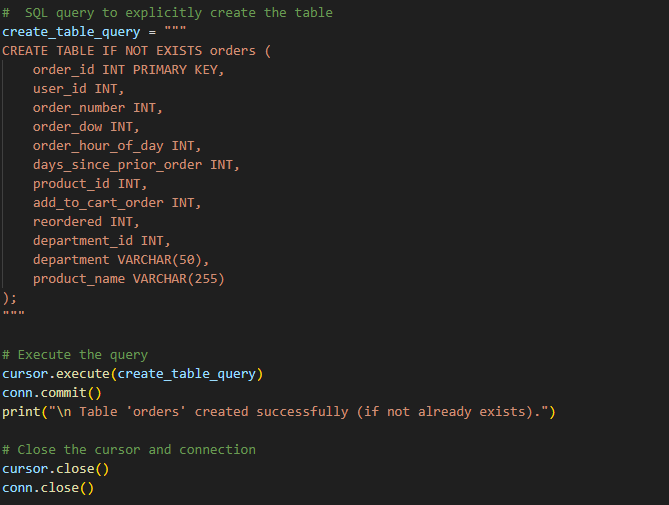
* Handled missing values: Filled days\_since\_prior\_order with 0.
* Removed duplicates: Ensured unique transactions.
* Converted data types: order\_hour\_of\_day, order\_dow to integers.
* Standardized text columns: Lowercased department and product\_name.



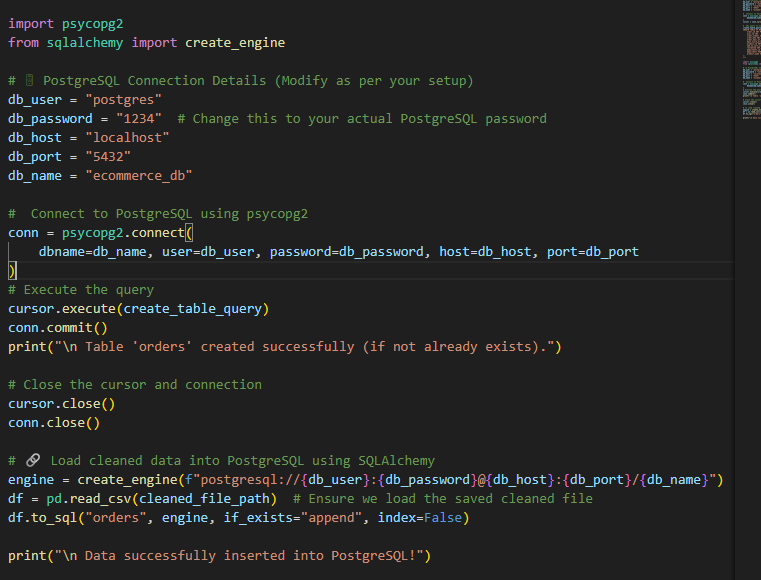
* Saved cleaned data:



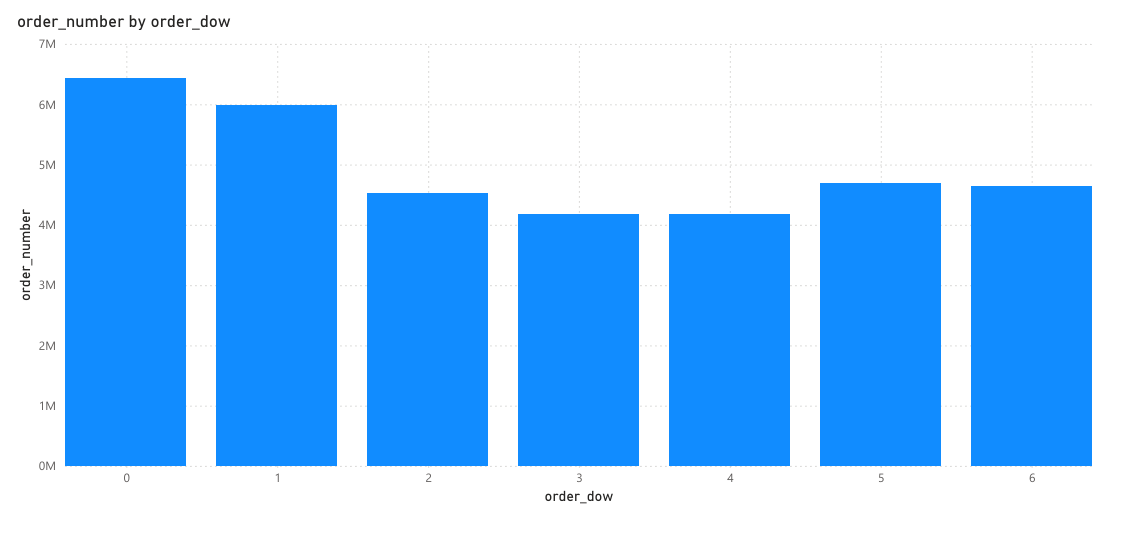
# **5.PostgreSQL Schema Design**

****

# **6. Loading Data using Python**



# **7. Visualization and Insights**



**What It Shows:**

* This bar chart represents the total number of orders placed on each day of the week.
* X-axis: order\_dow (0 = Sunday, 1 = Monday, 2 = Tuesday, 3 = Wednesday ,4 = Thursday, 5 = Friday 6 = Saturday).
* Y-axis: Sum of order\_number (total orders placed).

**Key Insights**

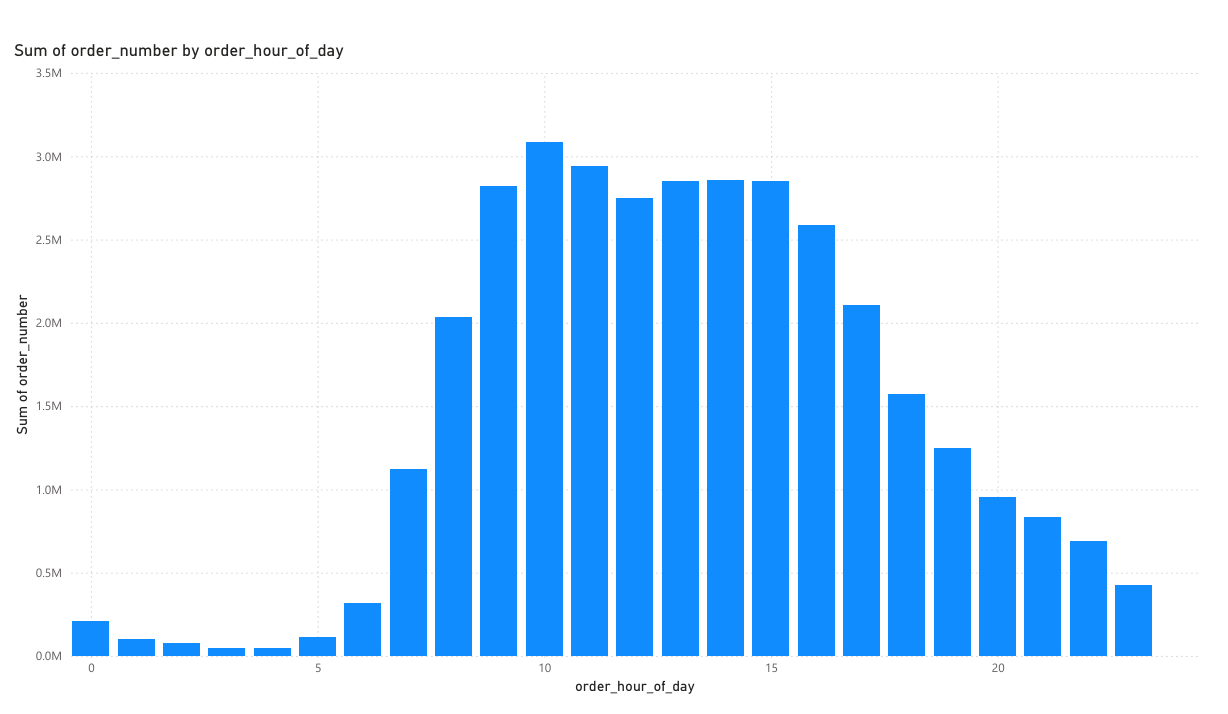
* Some days have significantly higher orders than others.

Based on the barchart Sunday is the day With the most orders.

* Identifies peak shopping days for better inventory and marketing strategies.

**Business Implications**

Schedule promotional campaigns on peak order days.  
 Optimize delivery and logistics to handle high order volumes efficiently.



**Orders by Hour of the Day (order\_hour\_of\_day)**

**What It Shows**

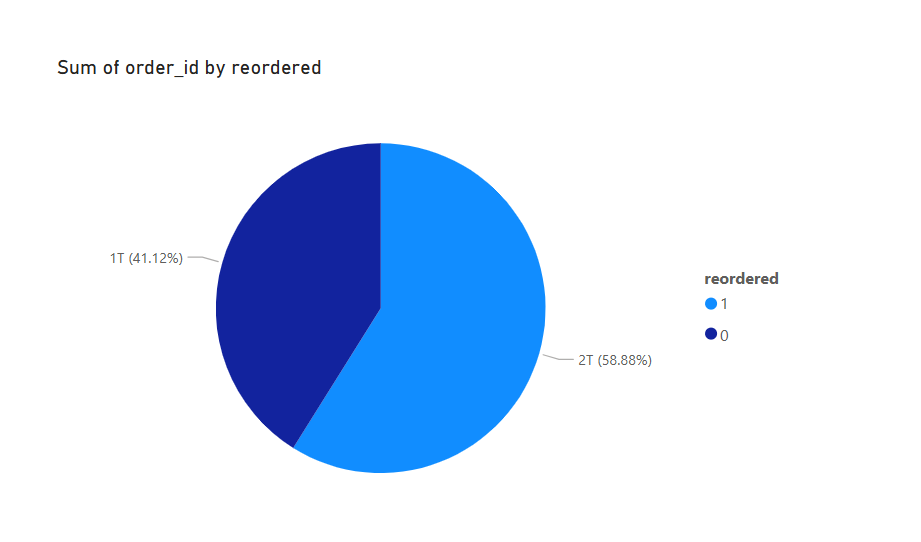
* This bar chart shows the number of orders placed at different hours of the day.
* X-axis: order\_hour\_of\_day (0-23).
* Y-axis: Sum of order\_number (total orders per hour).

**Key Insights**

* Peak order times appear between 9 AM and 3 PM.
* Orders gradually decline in the evening and night.

**Business Implications**

Optimize staffing and delivery logistics around peak hours.  
 Send targeted notifications during low-order hours to boost sales.



**Reordered vs. New Orders (reordered)**

**What It Shows**

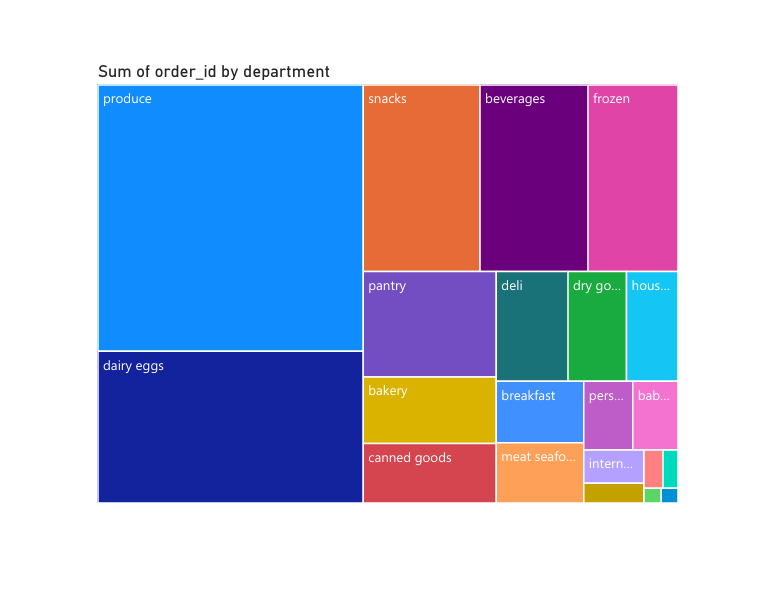
* A pie chart displaying the proportion of new vs. reordered products.
* Segment 1: New orders (reordered = 0).
* Segment 2: Reordered products (reordered = 1).

**Key Insights**

* 58.88% of orders are reorders, meaning a large portion of customers purchase the same items repeatedly.

**Business Implications**

Offer loyalty rewards to frequent reordering customers.  
Use subscription models for frequently reordered products (e.g., dairy, beverages).

****

**What It Shows**

* A bar chart showing the number of orders placed per department.
* X-axis: Different product departments.
* Y-axis: Sum of order\_id (total number of orders per department**).**

**Key Insights**

* The produce department has the highest orders.
* Other major departments include dairy & eggs, snacks, and beverages.

**Business Implications**

Ensure proper inventory levels for high-demand departments.  
Target discounts and promotions in lower-selling departments.