Analysis Report : ABC Ltd. E-Commerce Performance Dashboard

1.Project Overview :

- **Project Name :** ABC Ltd. E-Commerce Performance Dashboard
- **Objective :** To consolidate and analyze raw e-commerce data from 8 separate CSV files. The goal is to build a 4-page Power BI dashboard to provide actionable insights for four key business areas: Executive Summary, Logistics Operations, Marketing, and Seller Management.
- **Dataset:** Marketing Analytics For E-Commerce Market Place Company By Rishi Kumar Dataset. It has 8 distinct CSV files representing a relational e-commerce database, including data on customers, orders, items, payments, reviews, products, sellers, and geolocation.

2. Dataset Summary :

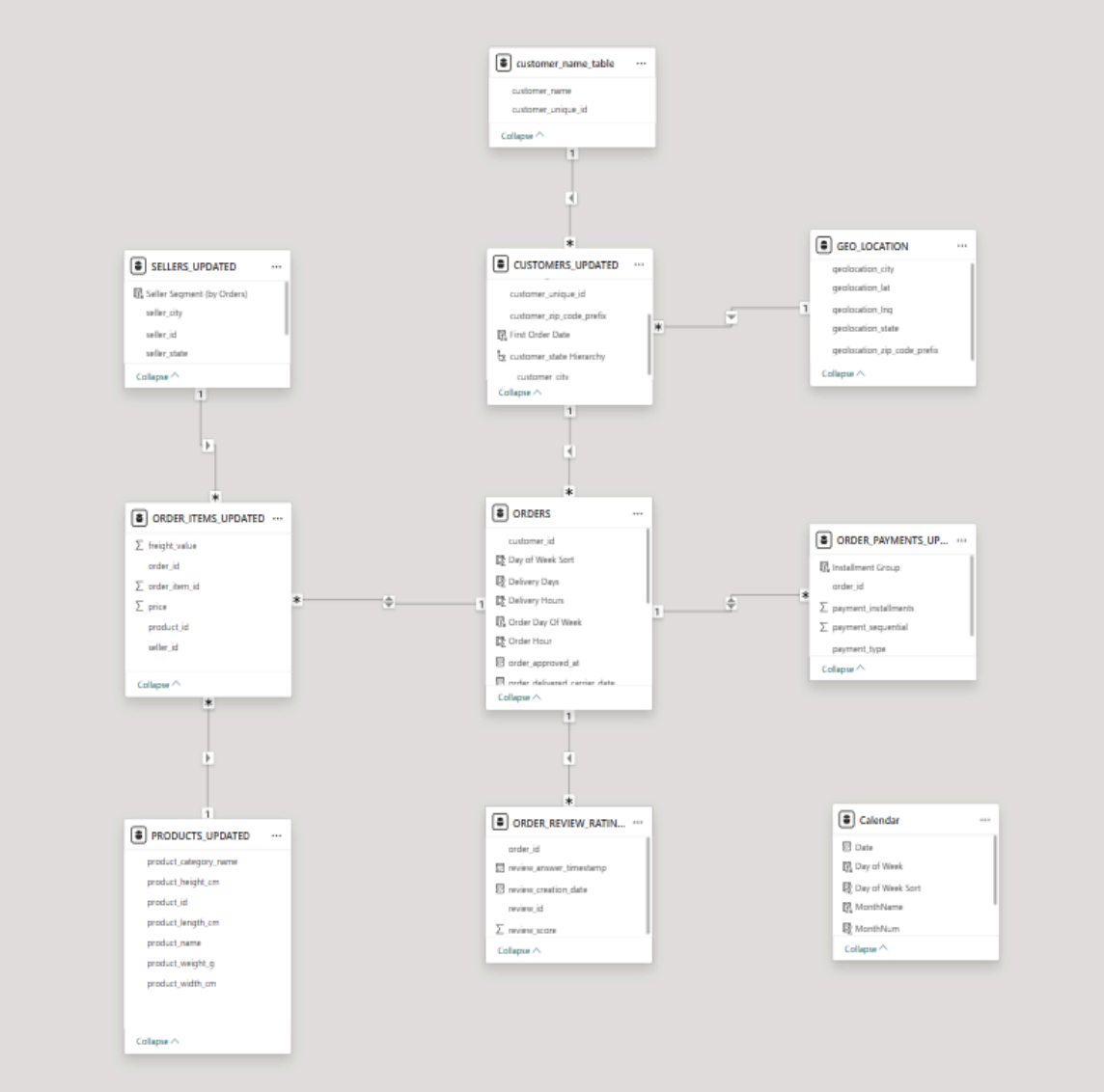
- **Rows/Columns:** The dataset is comprised of 8 tables, with key tables including CUSTOMERS (99,441 rows), ORDERS (96,455 rows), ORDER_ITEMS (112,650 rows), and PRODUCTS (32,327 rows).
- **Key Characteristics:** This is a highly relational dataset perfectly suited for a star schema model. It provides a 360-degree view of the e-commerce marketplace, linking customers to orders, which in turn are linked to products, sellers, payments, and reviews.
- **Cleaning Steps:**
    - **Datatype Conversion:** All date/time columns (e.g., order_purchase_timestamp, order_approved_at, shipping_limit_date) were converted from object (text) to the proper datetime datatype.
    - **Deduplication:** The ORDER_REVIEW_RATINGS table was cleaned by removing 827 duplicate entries based on review_id.
    - **Data Aggregation:** The GEO_LOCATION table, which had multiple entries per zip code, was aggregated to create a clean 1-to-1 lookup table by taking the mean() of latitude and longitude for each geolocation_zip_code_prefix.

3. Column Wise Assessment Summary

- **PRODUCTS:** The most critical finding is the absence of the product_name column. This is a significant data gap. All product-level analysis is restricted to product_category_name, limiting granularity.
- **CUSTOMERS:** customer_unique_id was identified as the true primary key for a customer dimension, while customer_id acts as a foreign key per-order.
- **ORDERS:** All 8 timestamp columns were incorrectly typed as object (text) and required conversion.
- **ORDER_REVIEW_RATINGS:** This table contained significant data duplication (827 duplicate review_ids), which would have skewed all review-based metrics if not cleaned.

4. Data Model Overview

- **Tables**: The data was modeled into a Galaxy Schema.
    - **Dimension Tables**: CUSTOMERS, PRODUCTS, SELLERS, GEO_LOCATION (Aggregated), ORDER_REVIEW_RATING, ORDER_PAYMENTS
    - **Fact Tables** : ORDERS and ORDER_ITEMS
- **Relationships** : One-to-many relationships were established from dimension table primary keys, Between the ORDERS and ORDER_ITEMS table There Stands a One to Many and Bi-directional relations.
- **Diagram** :



- **Key Calculations** :
    - **Measures** : A library of DAX measures was created, including Total Revenue, Total Orders, Average Review Score, Avg. Days to Ship, Avg. Days in Transit, Avg. Days Lagged, and New Customers.

    - **Calculated Columns:** Key columns were created for segmentation, such as Seller Segment (by Orders) and Order Hour and Order Day of Week for seasonality analysis.

5. Analysis & Insights

- **Seller Performance Skew** : Seller performance is not evenly distributed.Most of all sellers have 50 or fewer orders. A small number of sellers represents the Top order receiving (more than 200 orders), making them critical partners.
- **Revenue vs. Quality Correlation On Sellers** : The "Seller Performance Quadrant" (scatter plot) visually confirms a correlation between Total Revenue and Average Review Score. This identifies high revenue generators that have high average review scores.
- **Marketing Misalignment** : The orders seasonality graph reveals that customer purchasing activity peaks time (approx. 10 AM - 10 PM) and helps to find specific days trends. This is a critical insight for marketing teams.
- **Logistics Delivery Time**: The analysis proves the primary operational delay is shipping. The Avg. Days To Transit is the largest contributor to total fulfillment time.

6. Conclusions

The dashboard successfully transforms raw, disconnected data into a powerful analytical tool. The analysis concludes that the business has a solid foundation

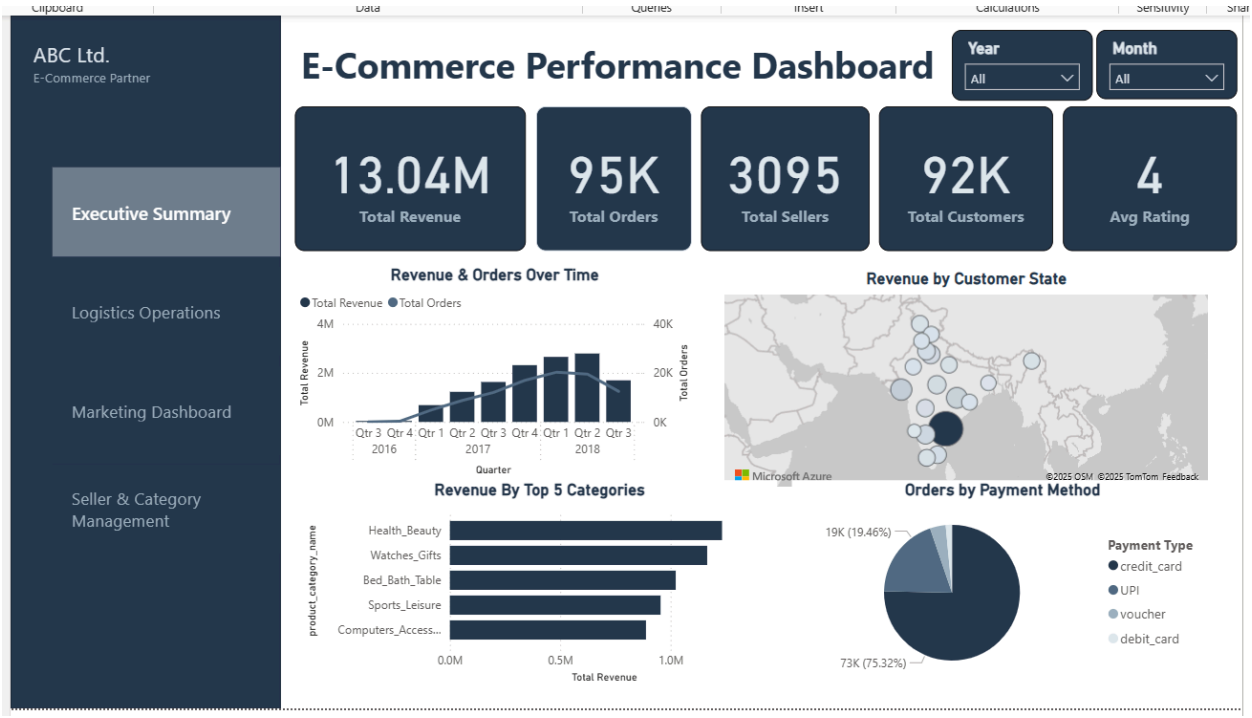
- **Operational Inefficiency:** Growth is constrained by inconsistent shipping partners.
- **Untapped Marketing Potential:** Marketing efforts are likely misaligned with actual customer purchasing behavior, leaving significant revenue opportunities on the table.

7. Recommendations

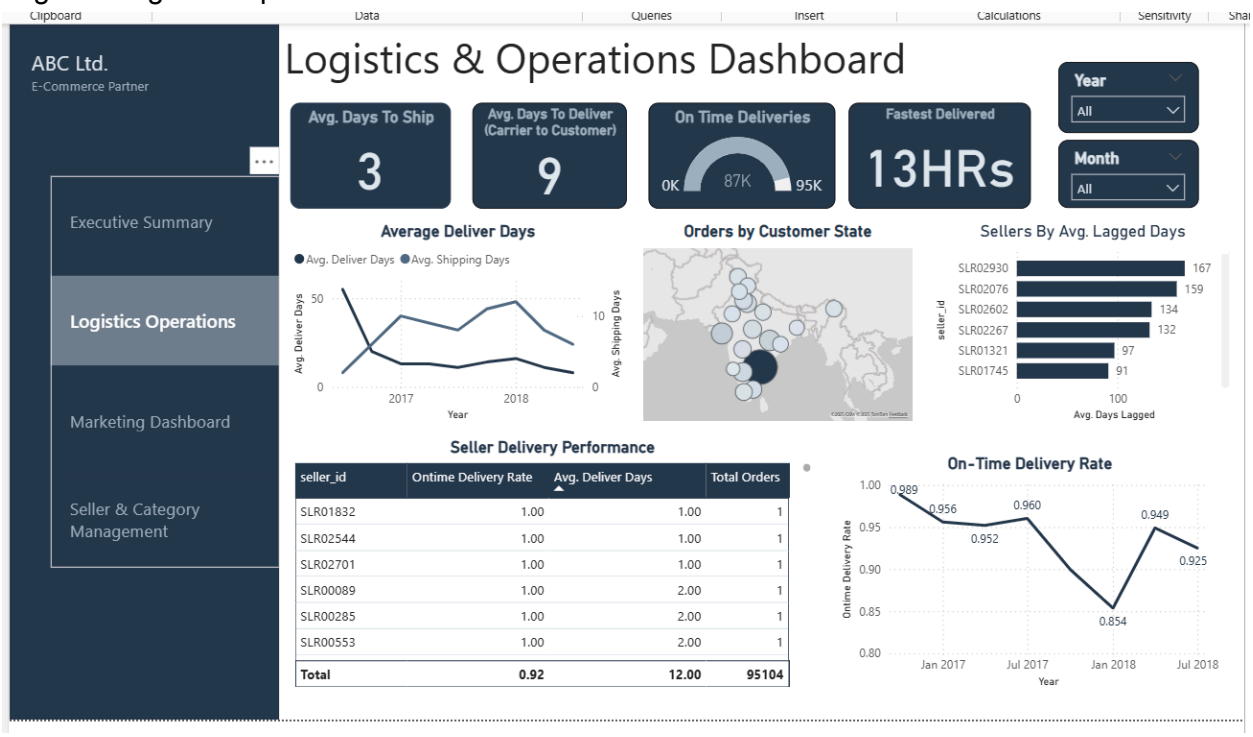
- **FOR LOGISTICS:** Make The Shipment Partners to be better to perform Fast shipments. Make a meeting with shipment partners to improve shipment speed.
- **FOR MARKETING:** Immediately reschedule all digital marketing efforts (email campaigns, social media ads) to target the user orders more getting time from the seasonality chart.
- Make Offers On Debit Cards to get more debit card usage.
- Recommend Products according to the product combination chart.
- **FOR SELLER MANAGEMENT :** Use the "Seller Performance Quadrant" and Seller Segment slicer to emerge sellers with low revenue.

8. Dashboard Overview

Page 1 : Executive Summary

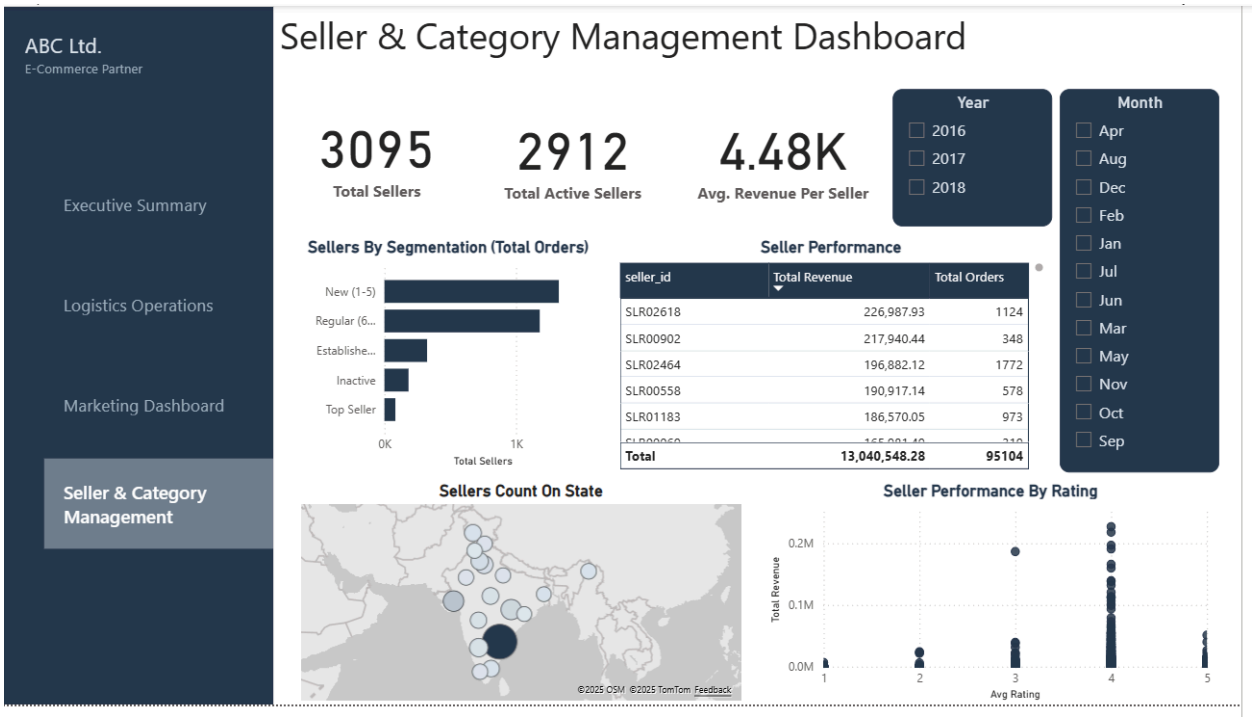


Page 2 : Logistics Operations



Page 3 : Marketing Dashboard

Page 4 : Seller & Category Management



9. Notes / Limitations

- **Data Limitation:** All product analysis is strictly at the category level. The lack of product_name prevents true Market Basket Analysis (which would show "product A is bought with product B").
- **Constraints:** The analysis is 100% based on the 8 provided CSV files. No external data (e.g., marketing ad spend, warehouse costs) was included.