# E-commerce app business analysis

Presented by Malika Temurova

2024

Contents

[Introduction 3](#_Toc162389193)

[Problem overview 3](#_Toc162389194)

[OLAP cube building 5](#_Toc162389195)

[Analysis 5](#_Toc162389196)

[Reference List 9](#_Toc162389197)

### Introduction

The current report describes the dataset selected for the business case of E-commerce orders of the Spanish company (Jain, 2019) and provides an analysis to assist the management in decision-making. The report is based on the data warehouse created as part of the first coursework. The dataset contains the details of transactions in the form of orders from the web application. The dimensions include product, customer, review, seller, and date. The dataset can provide insight to the top management to answer strategic questions of the business in logistics and supply-chain management, customer relationships and marketing strategies.

The source files of the SSAS project, the source files and the database creation scripts can be found in the zip file submitted along with report.

### Problem overview

A screenshot of a computer

Description automatically generated

Figure 1. The schema of the data warehouse created for the dataset.

As per the Figure 1, the structure of the data warehouse is a Star schema with a central table FactOrder and a set of dimension tables joined to the Fact table via foreign key attributes. The Star schema is known to contribute to the improved performance, albeit causing redundancy.

In the current warehouse, the key focus areas are customer, seller, payment, product, and the review, left by customers on the products they have purchased. The data warehouse ensures that all the areas of interest of the business are covered.

The key areas of interest for the businesses operating in the e-commerce sector include the marketing strategies directed at creating customer churn and customer loyalty (Smith and Brynjolfsson, 2001), supply chain management and logistics (Stock and Lambert, 2001), customer reviews and the brand image of the business that is acquired by managing the reputation of the company (Chevalier and Mayzlin, 2006).

To ensure the effective management of the business the managers and stakeholders should focus on finding data-driven answers to the business questions posed by the market. The following questions are deemed to be effective in providing insight into the current state of the company across the categories:

Marketing: Which categories of products have brought the highest income on the platform?

The graph can be displayed in the form of a pie chart to ensure there is a detailed information about the revenue brought by each category.

Supply chain and logistics: Which cities or locations have had the highest number of sales or are associated with the highest revenue both from the sellers and customers point of view?

The graph is best presented by the histograms where the Y-axis represents the revenue gained by each city. This graph is very logical in representing the number of shares brought by the top cities.

Reviews: Which categories of products have had the highest revenues based on the rating score that they have received? Which products do the customers prefer?

The graph can be represented in linear form along with the trends in the overall revenues. The display makes the most frequent category evident and intuitive.

Marketing & Goods: Is there a relationship between the description length of the product on the platform and the number of orders?

The linear chart represents the description length that has had the highest revenue.

It is argued that having answered these questions, the management can gain insight into the market and the customers, as well as define the strategies that the business needs to follow to ensure it remains liquid.

### OLAP cube building

The OLAP cube contains the set of dimensions based on the data warehouse schema. It includes Date, Product, Customer, Seller, Review dimensions under the respective section of the SSAS project.

There is a hierarchy for dates named “Simple Calendar” consisting of the Year, Month and Key attributes. The Year and Month attributes are the key pair that cannot be repeated.

The cube contains the measures of the Payment value and the Order count.

The calculated measure “Total Revenue” is used to display the total revenue received from the orders created on the platform.

The KPI for the project compares the revenue generated to the fixed value of 500 to ensure that a certain quota is met to measure the effectiveness of sales.

The KPI could be changed to fir the business case had there been more information about the order or sales. The limitation of the current dataset is the lack of details about the customers and sellers due to the omission of the sensitive data. It is argued that extra information could have helped the business track the behavior of the customer to check if there is a repeated behavior patters that could be used to create more effective marketing campaigns that could enhance the customer experience. Since limited data on customers has been found, no in-depth analysis of the customer behavior was created. Similarly, there was a limited set of information on the products and there were ambiguities in identifying if the product is unique or is it the same category of products stored in stock.

### Analysis

The data warehouse has been analyzed to retrieve answers to the questions identified.

A colorful circle with text and numbers

Description automatically generated

Figure 2. The overview of the key categories with the share of revenue.

The figure 2 represents the division of categories with the revenue amount brought by each category. The category with the highest revenue is “Beleza saude” with near 17% contribution to the total revenue, with “Cama besa banho” and “Esporte lazer” being the second and the third biggest categories. Since the figure provides insights on the categories with the highest returns, the marketing campaigns can focus on targeting and popularizing the top categories while reducing the costs associated with those categories that provide lower returns. This approach is argued to bring more revenue while reducing the extra costs.

A screen shot of a computer

Description automatically generated

Figure 3. The display of the cities associated with the highest payments made.

The figure 3, presents the biggest cities with the most payments. San Paulo, Rio de Janeiro, and Belo Horizonte represent the top-3 cities for the number of orders and the value of payments. Based on the demographics for the payments, the management can undertake measures related to the supply chain and the logistics of the goods, ensuring the establishment of the warehouses that could store the bigger amounts of goods to ensure the demand is met in bigger cities.

A graph with a line

Description automatically generated with medium confidence

Figure 4. The overview of the product categories with the highest ratings and the highest returns.

The figure above displays the categories rated positively and those on which the customers are willing to spend more money. As per diagram, the “Beleza saude” is associated with the highest rating and has had the highest return. The diagram could help the management focus on the quality of the products in the categories with the lower ratings since the quality of the supplied goods might be a cause of the lower ratings.

A graph on a white background

Description automatically generated

Figure 5. The diagram displays the most optimal description length according to the orders made.

The figure 5 has identified that the most orders were made for products with the description length of the 1893 characters. Although there might be other reasons for the high number of reasons, it can be concluded that the more detailed description of the product can prompt the customers to order the item.

A graph with blue and orange lines

Description automatically generated

Figure 6. Provides the total revenue across the date time categorized by month and year.

The revenue across the period in the dataset has totaled 1 trillion units. The trend represents that the revenues have gradually increased over the course of the years, meaning that the platform has been effective in placing orders. It is recommended that more elaborate hierarchies are created to be able to display the trends over the course of quarters or fiscal calendar.

Overall, the trends identified show that the platform is achieving popularity among the customers and more sellers are entering the platform to sell the goods. The analysis shows good geographical coverage that can help the business make informed decisions about logistics and the transportation. It is advised that the company provides more details on the sales or orders made as well as the customer information to enhance the customer experience and customer traction.

### Reference List

Chevalier, J. A., Mayzlin, D. (2006). *The Effect of Word of Mouth on Sales: Online Book Reviews*. Journal of Marketing Research, 43(3), 345-354

Jain, Aashish. (2019). Ecommerce\_orders. *Kaggle.* Available from [https://www.kaggle.com/datasets/jainaashish/orders-merged](https://www.kaggle.com/datasets/jainaashish/orders-merged%20Accessed%20last%20%5b%20)  [Accessed 1 March 2024]

Smith, M. D., Brynjolfsson, E. (2001). *Consumer Decision-Making at an Internet Shopbot: Brand Still Matters*. SSRN Electronic Journal

Stock, J. R., & Lambert, D. M. (2001). *Strategic Logistics Management (4th ed.).* New York: McGraw-Hill/Irwin