E-Commerce Data Analysis

Nityam Pareek
Mehta Family School of Data Science and Artificial Intelligence
Indian Institute of Technology Guwahati
Guwahati, India
nityam@iitg.ac.in
Github: https://github.com/NityamPareek/ecommerce_analytics

***Abstract*—In the rapidly evolving landscape of ecommerce, data-driven decision-making has become pivotal for success. This project presents the development of an advanced ecommerce analytics portal, designed to empower business users with actionable insights through comprehensive data analysis and interactive visualizations. Our solution integrates a range of analytical tools and SQL-based queries to extract meaningful patterns and trends from extensive ecommerce datasets. The portal is structured into distinct sections, each focusing on critical aspects like customer behavior, order processing, product performance, and seller dynamics. Utilizing technologies such as Python, Django, SQL, and JavaScript with Chart.js, the system not only demystifies complex ecommerce data but also presents it in an intuitive and accessible manner. The application of this portal ranges from identifying top-selling products and average order processing times to analyzing customer spending habits and product review trends. By converging these diverse insights into a unified platform, the portal stands as a testament to the power of integrating data analytics in ecommerce for strategic business growth and customer satisfaction enhancement.**

***Index Terms*—Big Data Analytics, E-Commerce, Data Visualisation, Python, Django, JavaScript**

I. Introduction

The surge of ecommerce in the global market has revolutionized the way businesses operate and interact with their customers. With an ever-increasing volume of data generated by online transactions, there is a growing need for effective tools that can not only process and analyze this data but also translate it into actionable business insights. This project introduces an innovative ecommerce analytics portal, meticulously designed to address these requirements by providing comprehensive data analysis and interactive visualizations.

The motivation for developing this portal stems from the observed gap in efficiently utilizing the vast amounts of data accumulated in ecommerce platforms. Many businesses, especially small to medium-sized enterprises, struggle to harness this data due to a lack of accessible and intuitive tools. Traditional data analysis methods often fail to keep pace with the dynamic nature of ecommerce, leading to missed opportunities and suboptimal decision-making.

The primary objective of this project is to create a user-friendly, interactive portal that simplifies complex data analysis tasks. By presenting key metrics and trends in an understandable format, the portal aims to assist businesses in making informed decisions to enhance customer satisfaction, optimize product offerings, and streamline order processing.

The portal integrates a variety of technologies, including Python for backend processing, Django as the web framework, SQL for database queries, and JavaScript with Chart.js for frontend visualizations. The system architecture is designed to be scalable, ensuring it can handle large datasets typical in ecommerce environments. The portal is divided into multiple sections, each focusing on different aspects such as customer analytics, order details, product performance, and seller activity. Advanced SQL queries underpin the data extraction process, providing insights such as sales trends, customer demographics, order processing times, and product category performance.

This project contributes to the field of ecommerce analytics by offering a solution that is not only technologically advanced but also user-centric. It bridges the gap between complex data analysis and business decision-making, thereby enhancing the overall efficiency and effectiveness of ecommerce operations.

II. Dataset

The dataset used for this project was provided by Rakamin Academy. It has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allows viewing an order from multiple dimensions and includes (but is not limited to) the following data:

- Customer Details
- Product Attributes
- Payment Methods
- Customer Reviews
- Order Details
- Product Catalogues
- Seller Details
- Location Details

The dataset consists of 8 CSV files and takes up approximately 120 MB of disk space. The CSV files are linked in the following manner:

This dataset will be stored as a relational database. For such an implementation, the Entity-Relationship (ER) Diagram and the corresponding relationships are demonstrated using Fig 1.

III. Working Methodology

A. Frontend Development

- The frontend development of the ecommerce analytics portal employed a user-centric and agile methodology,

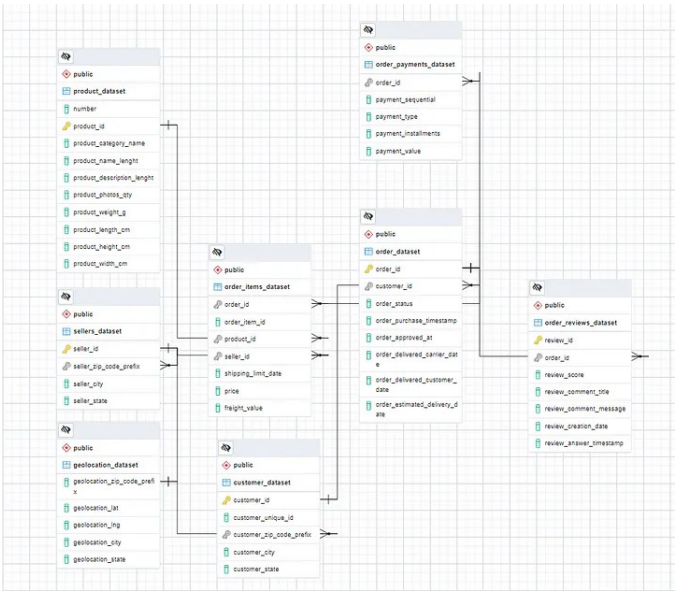

Fig. 1. Entity-Relationship Diagram.

focusing on creating an intuitive and responsive user interface (UI) that simplifies complex data interactions.

- The process began with a detailed analysis of user requirements, leading to the creation of the visual and interactive aspects of the portal.
- Utilizing HTML5 and CSS3, I crafted a clean and modern layout, ensuring cross-browser compatibility and mobile responsiveness.
- The interactive elements and data visualizations were powered by JavaScript, with Chart.js being pivotal for rendering dynamic and visually appealing charts.
- Iterative development and was integral to my methodology, allowing me to refine the UI/UX progressively. This approach ensured that the frontend not only aligns with the users' needs but also effectively presents data analytics in an engaging and comprehensible manner.

B. Backend Development

- The backend of my ecommerce analytics portal was engineered with a focus on robust and scalable data handling, leveraging Python and Django for their efficiency and ease of use.
- Python's versatility and rich library ecosystem made it an ideal choice for backend processing, allowing for effective handling of complex data operations.
- Django, a high-level Python web framework, was utilized for its rapid development capabilities and clean, pragmatic design, which significantly streamlined the backend development process.
- Data processing and analytics formed the backbone of the backend functionality. I implemented various Python modules to process and analyze the incoming data, transforming raw data into meaningful insights.
- This involved aggregating, filtering, and computing key metrics and trends, which were then relayed to the frontend for visualization.
- Security and performance optimization were also paramount in the development process, ensuring the integrity and responsiveness of the system even under heavy loads.

C. Data Storage and Retrieval

- The database was designed using SQL, a choice motivated by its proven reliability and widespread usage in handling complex queries and large datasets.
- The database schema was meticulously planned to optimize data retrieval and storage efficiency, with tables structured to store diverse datasets ranging from customer details to order histories and product information.
- I also implemented normalization techniques to eliminate data redundancy, ensuring data integrity and consistency.
- The data was stored in a relational database management system (RDBMS), which facilitated efficient data organization and retrieval through well-defined relationships between tables.
- This approach allowed me to execute complex analytical queries necessary for generating insights, such as customer behavior patterns, order processing times, and sales trends.
- The backend's architecture ensured seamless integration with the frontend, providing a robust foundation for the portal's data-driven functionalities.

IV. Implementation

A. Overview Page

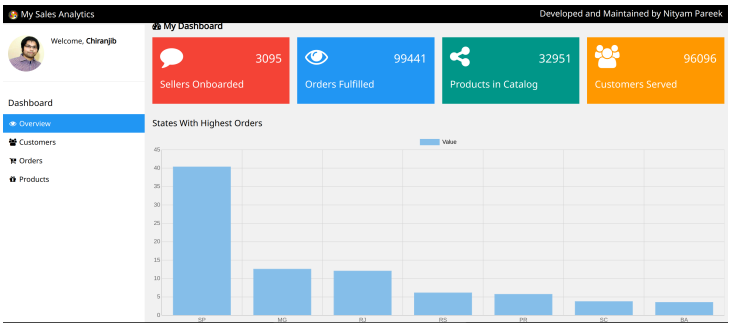

Fig. 2. Home Page

- The homepage is named overview. For simplicity and the best user insights, it starts off by highlighting the top statistics that every business should know, that are sellers onboarded, orders fulfilled, products in catalog, and the number of customers served.
- This is then followed by a bar graph, which gives geographical analysis of the regions from where the firm is getting most sales.
- Lastly, this page also contains the latest reviews which have been left by the customer.

B. Customers Page

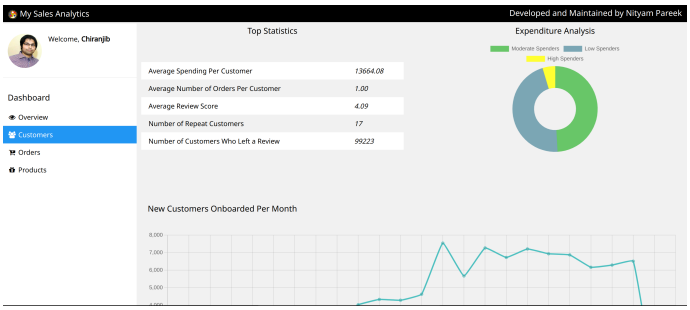

Fig. 3. Customers Page

- The customers section is designed to provide in depth analysis regarding customer behaviour, habits, and growth.
- The section starts with some must know statistics such as average spending per customer, average number of orders per customer, average review score, number of repeat customers, and the number of reviews.
- The doughnut chart on the right is significant in understanding user habits as it categorizes users as low, medium, and high based on their spending patterns.
- The last data on this page is a line chart depicting the customer acquisition statistics over time.

C. Orders Page

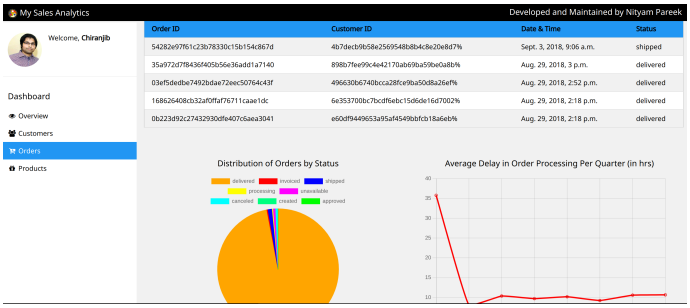

Fig. 4. Orders Page

- The orders page is important for the company to have a good idea about its logistics and delivery times to ensure that they deliver a good customer experience.
- On the top we see a table enlisting the most recent purchases
- This is followed by a pie chart depicting the number of orders that are in different stages of delivery, such as approved, delivered, canceled, etc.
- This is followed by a line chart depicting the average order fulfillment time. This is an important success metric for the company with regards to its logistical capabilities.

D. Products Page

- The products page is important for the company for its inventory analysis and also for finding out the trends in customer spending patterns.

Fig. 5. Products Page

- The page starts off by listing the inventory by product type, ie the number of subproducts present in each product category.
- This is followed by a bar graph which enlists the popularity of each subcategory by calculating the number of orders placed for each.
- Lastly, a line graph depicts the gross sales value per month for the firm.

V. Impact

This initiative makes a valuable contribution to the realm of ecommerce analytics by presenting a solution that combines technological sophistication with a user-centric approach. It effectively closes the divide between intricate data analysis and the decision-making processes within businesses, resulting in heightened efficiency and effectiveness across ecommerce operations.