



Order Flow Insights: Next-Gen Ecommerce Analytics

Project-Based Learning (PBL) Review 3

Team Members:

- P. Anirudh — 2410030128
- S. Rohit — 2410030030
- B. Santosh — 2410030254
- M. Maruthi Reddy — 2410030442



Novelty of Developed Methodology

Our methodology introduces an **innovative data pipeline** that automates the tracking and analysis of e-commerce orders, integrating advanced preprocessing at every stage. Unlike basic order loggers, our system brings together real-time data cleaning, transformation, and enrichment using Python and Pandas, ensuring every dataset analyzed is robust and free from inconsistencies.

Uniquely, this workflow prepares the data for **dynamic visualization**, allowing business stakeholders to gain actionable insights into order trends, customer behaviors, and supply chain bottlenecks. No manual intervention is required, significantly reducing errors and improving response time to business challenges.

Adaptability of Methodology

Modular Design

The designed system is modular and adaptable to a vast array of e-commerce business models and datasets, making it an optimal solution for small startups as well as large enterprises.

Configurable Modules

By abstracting the preprocessing logic and providing configurable modules for data import, cleaning, and visualization, it enables organizations to plug in their specific data sources while leveraging the robust analytics engine.

Future-Ready

This adaptability ensures that as business requirements evolve, new data types or reporting demands can be integrated with minimal rework.

Moreover, the entire solution is built using **open-source technologies**, making it accessible and cost-effective for broader adoption.

Methodology Evaluation

Our approach to evaluation is **holistic and data-driven**, starting from detailed preprocessing that includes null value imputation, duplicate record removal, and intelligent outlier detection. Each step is verified by measuring downstream analytical accuracy and visual clarity, ensuring that cleaned data produces reliable business intelligence.

Test cases are constructed for performance benchmarking, validating the system's speed and effectiveness on real-world order datasets. Through this rigorous procedure, the methodology demonstrates **measurable improvements in data quality**, leading to enhanced trust in generated visualizations and more informed decision-making for stakeholders.





Dissemination and Accessibility

To ensure the widest reach and impact, results are disseminated via **interactive dashboards built within Jupyter Notebooks**, where users can engage with visualizations, filter relevant data, and export insights for reporting. These dashboards are tailored for both technical analysts and non-technical users, making sophisticated analytics comprehensible without any programming expertise.

By choosing open Python libraries and emphasizing clarity in graphical outputs, the workflow enhances accessibility and ensures that **data-driven strategies are within reach for all departments**, regardless of size or expertise.

Data Preprocessing

Our project places **strong emphasis on rigorous data preprocessing** as the foundation for reliable analytics. The process starts with a thorough scan for null or missing values, using automated scripts to either impute or remove incomplete entries.

01

Null Value Handling

Automated scripts identify and handle missing values through imputation or removal of incomplete entries.

02

Duplicate Removal

Multi-criteria checks identify and remove duplicate orders to prevent analytical distortions.

03

Data Standardization

Column names are standardized for consistency, and date entries are converted to uniform formats for accurate time-based analysis.

04

Outlier Detection

Advanced outlier detection techniques filter out abnormal order values or shipment delays to present a true picture of business operations.

Most crucially, the methodology incorporates advanced outlier detection techniques, filtering out abnormal order values or shipment delays to present a true picture of business operations. Comparisons of before-and-after datasets demonstrate **substantial quality improvements**, reflecting the effectiveness and necessity of the preprocessing phase.