Generative AI for E-Commerce

Muhammad Hashim Bilal Qureshi
SBASSE
LUMS
Lahore, Pakistan
24030022@lums.edu.pk

Zaid Siddiqui
SBASSE
LUMS
Lahore, Pakistan
24030005@lums.edu.pk

Usama Habib
SBASSE
LUMS
Lahore, Pakistan
25110235@lums.edu.pk

Saad Nadeem
SBASSE
LUMS
Lahore, Pakistan
25110012@lums.edu.pk

***Abstract*—This study explores the role of generative AI and fuzzy logic in transforming e-commerce operations using a large-scale Amazon sales dataset. By applying fuzzy inference systems, Monte Carlo simulations, and fine-tuning language models, we analyzed delivery risks and automated sales queries. The research also demonstrates how Retrieval-Augmented Generation enhances natural language interaction with structured business data.**

***Index Terms*—E-commerce, Fuzzy Logic, Generative AI, Amazon Sales Dataset, Random Forest, Monte Carlo Simulation, RAG, LangChain, Mistral-7B, Fine-tuning, Delivery Risk, Natural Language Querying, Data Mining, Transaction Utility Theory**

I. Introduction

In a world where technology touches every part of our lives, the way we shop has undergone a quiet revolution. People used to shop in person, but with the advances made in technology, now prefer to get their orders delivered to their homes. With the advent of artificial intelligence (AI), the world of e-commerce has been facing new challenges, especially the businesses. As convenience and personalization become the norm, businesses are under growing pressure to meet rising expectations, manage inherent supply chain risks, and stand out in crowded digital marketplaces.

To study the role of technology in the e-commerce space, especially Generative AI we conducted a study on an Amazon Sales dataset that was extracted from Kaggle. The dataset had records of more than 120,000 transactions along with columns that we would expect to see in a seller's records. Our analysis comprises the application and comparison of existing data mining literature for the e-commerce industry. Additionally, we improve upon the work done in the literature and showcase the true potential that Generative AI holds in the e-commerce sector.

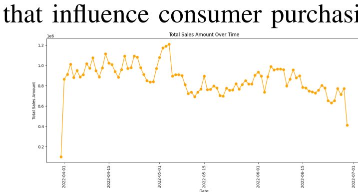
II. Exploratory Data Analysis

Our dataset contained 4 files that were collectively analyzed for this study:

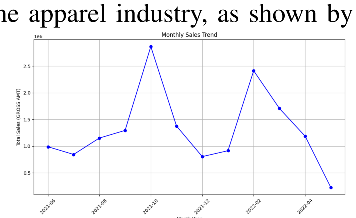
- Amazon Sales Dataset – Order-level data including status, fulfillment, product, and region
- International Sales Report – Global transaction details with style, customer, and gross amounts.
- May-22 Dataset – Local retail transactions with pricing, discounting, and quantity.
- P L March 2021 Dataset – Same structure as May but for March 2021.

A. Sales and Pricing Trends Analysis

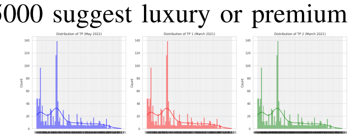
First, we looked at the amount of sales that took place over time from Amazon. The sales trends over time help us understand seasonality, revenue growth, and potential slow periods. The sales data reveals clear trends, with certain periods reflecting higher transaction volumes. This could indicate periodic promotions, festive sales, or other seasonal factors that influence consumer purchasing behavior.

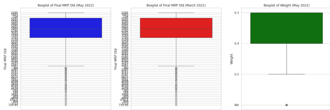


Then, we looked at the international sales report from the dataset. The total gross sales amount revealed notable fluctuations, with peak sales observed in Feb-22 and Oct-21. This shows there is a clear seasonality effect that exists within the apparel industry, as shown by the line graph.



The TP Distribution Graph shows that transaction prices (TP) in both May 2022 and March 2021 (TP1) are clustered around 500-700, suggesting consistent mid-range pricing. However, TP2 in March has lower averages, likely representing discounted or smaller items. The Price Change Histogram indicates that most MRPs remain stable between March 2021 and May 2022, with price changes centered around zero. Some items show significant hikes, possibly due to inflation, while a few reflect reductions, likely from clearance sales. The Boxplot of Final MRP Old confirms consistent pricing, with median MRPs between 2000-2500 for both months. Outliers above 5000 suggest luxury or premium items.

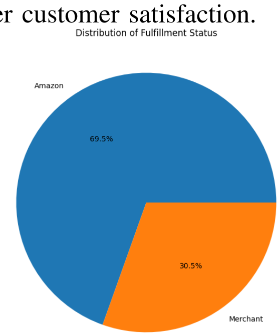




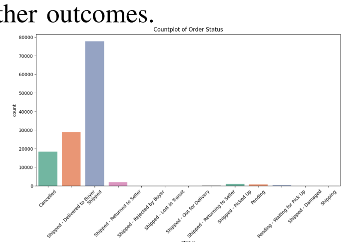
B. Fulfillment Order Status Analysis

This section examines the breakdown of who fulfills the orders—whether Amazon handles the fulfillment process or it's managed by individual merchants.

The pie chart indicates that 69.5% of orders are fulfilled by Amazon, while 30.5% are handled by merchants. Amazon's dominance in fulfillment may reflect its streamlined logistics operations, which could contribute to faster deliveries and better customer satisfaction.



The next graph describes the overall distribution of order statuses to assess fulfillment success rates, cancellations, and other outcomes.

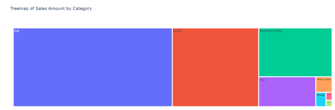


The countplot highlights the breakdown of different order statuses. It shows that a significant number of orders are successfully delivered, reflecting a smooth fulfillment process for most transactions. However, there is a smaller, though notable, percentage of canceled orders. This finding suggests that there may be areas where improvements can be made to minimize cancellations, potentially through better inventory management, clearer product descriptions, or enhanced customer communication.

C. Product Product and Customer Analysis

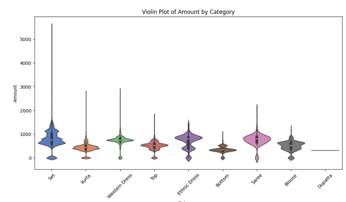
This section analyzes which product categories contribute most to sales and order volume. Understanding top-selling categories can guide inventory planning, marketing strategies, and product assortment decisions.

The treemap visualization provides a hierarchical view of sales across various product categories. Categories like kurtas, sets, and western dresses dominate the sales landscape, reflecting their popularity among customers. This insight can help businesses focus on maintaining stock and offering attractive discounts on high-demand categories.

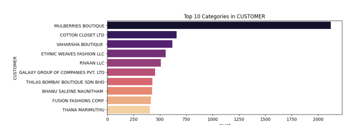


The violin plot provides further insights into the distribution and density of sales amounts within each category: Sets exhibit the widest variability, as indicated by the longest needle on the plot, which extends well beyond 5,000. This suggests that while most sales of sets occur within a mid-range price, some high-value orders contribute significantly to overall sales. Kurtas and western dresses also show notable variability, with high median values and occasional spikes in sales amount, reflecting their popularity across different price tiers. Other categories show tighter distributions, indicating more consistent pricing and lower variability in sales amounts.

By combining the breadth (from the treemap) and depth (from the violin plot), this analysis helps pinpoint which categories offer opportunities for further sales growth, inventory optimization, or targeted marketing campaigns.

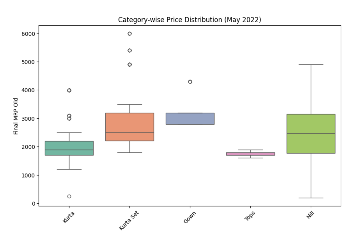


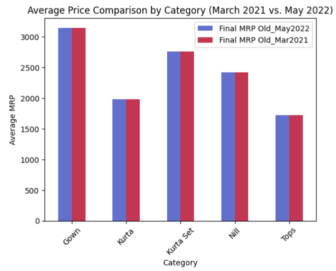
The most frequent customer was ”MULBERRIES BOUTIQUE,” which consistently placed multiple orders and emerged as the clear top customer in terms of transaction frequency. This suggests that MULBERRIES BOUTIQUE could be a key business client with a strong purchasing relationship.



In the May-22 and March 2021 datasets, the Boxplot of Category-wise Price Comparison shows that Gowns have the highest price range, with some items priced above 5000, while Tops are the cheapest. Kurta and Kurta Sets dominate the dataset, reflecting moderate price distribution around 1000-3000.

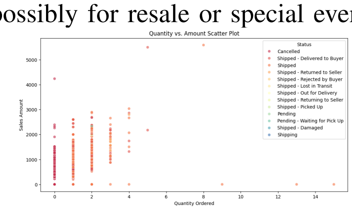
The Average Price by Category Bar Graph reveals price increases across categories between March and May. Gowns show the highest hikes, possibly due to rising material costs or increased demand, while Tops remain stable.





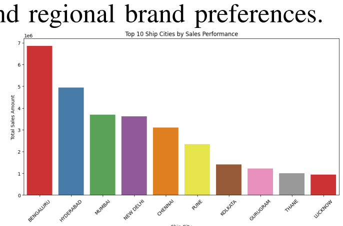
D. Quantity Sold vs. Amount Analysis

The scatter plot shows that while most orders consist of small quantities, higher quantities sometimes yield significantly higher revenue. This reflects occasional bulk purchases, possibly for resale or special events.



E. Shipping and Regional Analysis

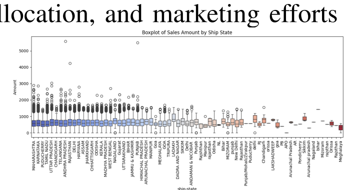
The geographic distribution of orders provides insights into regional demand patterns. Certain regions, such as Maharashtra, Karnataka, and Tamil Nadu, dominate the order volume, as shown in the bar chart. This regional concentration could be influenced by factors like population density, income levels, and regional brand preferences.



The chart highlights the leading cities contributing to sales, reflecting key regional markets where the brand is well-established or experiences higher customer demand. The boxplot reveals variations in sales amounts across different shipping states. The distribution for each state shows the median sales value, the interquartile range (IQR), and potential outliers.

States with a broader IQR demonstrate more variability in sales amounts, indicating diverse customer spending patterns or product price ranges. States with a narrower IQR show more consistent sales values. The presence of outliers in certain states may reflect high-value purchases or bulk orders, which could be investigated to uncover further trends or anomalies.

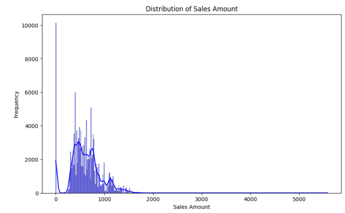
This analysis highlights regional sales dynamics, providing actionable insights to optimize shipping logistics, inventory allocation, and marketing efforts tailored to different regions.



F. Payment and Currency Trends

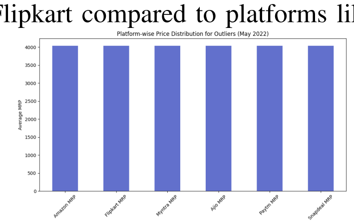
This section explores payment amounts and currency variations to detect anomalies or inconsistencies.

Most transactions fall within a standard price range, but occasional outliers reflect high-value purchases. These outliers merit further investigation to understand customer behavior or possible errors.



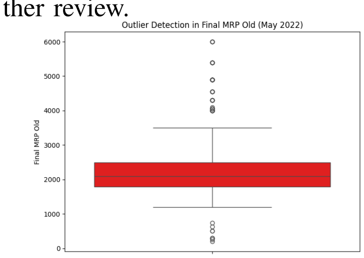
The Platform-wise Price Distribution for Outliers reveals uniform pricing across platforms for high-value items, indicating centralized pricing control. This suggests that premium products may have manufacturer-set MRPs, ensuring consistency across sales channels. A notable outlier priced at 254 may represent a clearance sale or potential pricing error.

The analysis of payment trends also emphasizes Amazon's pricing dominance, with average prices higher on Amazon and Flipkart compared to platforms like Ajio and Snapdeal.



G. Outlier Analysis

The Boxplot of Outlier Detection identifies extreme high-priced products, particularly in Kurta Sets and Gowns, with items priced around 5997 and 5395. These prices may reflect luxury items, though potential pricing anomalies warrant further review.



III. Literature Review

The intersection of artificial intelligence (AI), fuzzy logic, and e-commerce has seen a significant rise in academic and applied interest, particularly in improving personalization, decision-making, and operational efficiency across digital platforms. One of the emerging approaches in this domain is the application of fuzzy association rule mining within live-streaming e-commerce platforms. Liao [1] demonstrated that fuzzy logic provides a scalable decision support foundation by capturing imprecise consumer behaviors such as fluctuating sales counts and variable promotional effectiveness. This approach not only increases consumer engagement but also enhances adaptability and customer retention through filtered, rule-based content delivery.

Furthering the capabilities of fuzzy logic, Liao [2] introduced a Fuzzy Logic Multi-Criteria (FLMC) model for sales forecasting. Unlike linear models, the FLMC incorporates subjective variables—such as expert ratings or customer sentiment—interpreted in linguistic terms (e.g., "high," "medium," "low"), improving inventory management, pricing, and marketing optimization. Building on this, predictive modeling with Monte Carlo simulations can augment such fuzzy systems by introducing probabilistic forecasting, which remains underexplored in current literature.

In parallel, research by Chen and Chang [3] introduced Transaction Utility Theory (TUT) to analyze the dynamics of Online-to-Offline (O2O) commerce. The study identifies marketing stimuli (e.g., e-coupons, free trials) and contextual factors (e.g., mobile app usability, pickup speed) as central to consumer repurchase behavior. The findings can be extended by applying TUT-based hypothesis testing on real transaction datasets to identify drivers of successful e-commerce order deliveries.

Complementing this, Pratama et al. [4] examined customer clustering using collaborative filtering methods to optimize offline retail recommendations. Memory-based techniques, particularly k-Nearest Neighbors (k-NN), outperformed model-based approaches like SVD. This segmentation (active, semi-active, inactive) can be expanded by integrating metrics such as customer lifetime value and retention rates, and by applying principal component analysis (PCA) to uncover implicit drivers of consumer satisfaction.

On the data mining front, Fageeri et al. [5] presented an efficient Bitwise-based structure for Market Basket Analysis, surpassing traditional algorithms such as Apriori and FP-Growth in computational efficiency. This model is particularly suitable for large datasets (e.g., ¿100,000 transactions) and could evolve into a rule-based recommender or forecasting system when integrated with domain expertise and stochastic models.

The psychological aspects of e-commerce adoption were explored by Ding and Najaf [6], who investigated the role of chatbot interactivity and perceived humanness in building trust. The study, rooted in Expectation-Confirmation Theory, demonstrated that trust acts as a key mediator in adoption decisions, with perceived enjoyment amplifying this relationship. These insights highlight the necessity of refining chatbot designs through personalization, voice, and visual cues, while also recognizing the need to explore long-term user engagement and negative emotional responses.

Generative AI (GenAI) is rapidly transforming innovation management, as shown by Mariani and Dwivedi [7], who identified ten emerging research themes, including intellectual property, creativity, and ethical challenges. In the retail sector specifically, Kulkarni and Bansal [8] illustrated real-world GenAI applications—from adaptive homepages to AI-driven customer service—that improve marketing, logistics, and product development. However, issues such as biometric surveillance, emotion recognition bias, and integration with legacy IT systems remain inadequately addressed.

Kshetri [9] extended this discourse to include how GenAI equalizes competition for small merchants, citing tools like Shopify Magic and Amazon's automated creatives. Yet, despite improved operational efficiency and search facilitation, the article falls short on measuring business KPIs (e.g., conversion rates) or addressing cross-cultural AI adoption behaviors.

Anand [10] emphasized the role of foundational AI technologies—ML, NLP, CV, and recommendation systems—in enhancing customer experience. These technologies support segmentation, chatbot assistance, visual search, and personalized recommendations, yet empirical validation, algorithmic transparency, and resource constraints for SMEs remain open research challenges.

Operational efficiencies within logistics were addressed by Zhang et al. [11] via two task optimization models for warehouse order picking, including a reinforcement learning-enhanced Q-Mix model. These models outperform legacy systems in speed and efficiency but require scalability, real-world constraints, and broader adaptability to become deployment-ready.

Fake review detection, a growing challenge in digital commerce, was tackled by Geetha et al. [12] using a hybrid MBO-DeBERTa model. The integration of Monarch Butterfly Optimization enhanced feature selection and adversarial resistance, though real-time deployment feasibility and multilingual extensions remain to be investigated.

Krishna et al. [13] contributed to recommendation system enhancement by integrating sentiment analysis through the MLA-EDTCNet model with collaborative filtering. While improvements in precision and recall were observed, real-time applicability and interpretability of sentiment influence warrant further exploration.

In cross-border marketing, Lin et al. [14] proposed a hybrid MLP-GWO-CNN model combining user rating data with semantic tags. The Grey Wolf Optimization algorithm enhanced convergence, leading to high recall and accuracy. However, concerns around multilingual adaptability, interpretability, and scalability remain unaddressed.

Rana et al. [15] introduced the BERT-BiGRU-Senti-GCN architecture for e-commerce sentiment analysis, achieving over 93% accuracy. The model's graph-based semantic representation is robust, yet lacks investigation into multilingual reviews and deployment in live systems.

In the context of digital entrepreneurship, a longitudinal study on female-led MYPES in Peru by [16] showed that favorable attitudes and increased investment in e-commerce translated into higher sales. However, the study emphasizes the need for scalability research, training support, and regional comparative studies.

From a consumer experience lens, Huang and Liu [17] examined customer complaints through the framework of moral transgressions. Justice violations—whether distributive or interactional—were central to dissatisfaction, suggesting that AI systems in customer service should be designed with a deeper understanding of ethical and communicative expectations.

Pu et al. [18] applied the UTAUT model to study Chinese Gen Z engagement with AI tools. Findings indicate that performance and effort expectancy positively influence behavior, though social influence was surprisingly negative. Further research is needed on longitudinal usage, cultural comparisons, and system design tailored to Gen Z preferences.

Lastly, Chen et al. [19] introduced a Bayesian learning-based fuzzy constraint agent negotiation model to enhance doctor-patient shared decision-making. The model improves agreement efficiency and satisfaction by forecasting preferences and adapting to uncertainty. Results showed increased overall satisfaction and fewer negotiation rounds. Future work should address real-world implementation, scalability, and ethical concerns.

In sum, the current body of literature reveals promising yet fragmented advancements in applying AI, fuzzy logic, and GenAI to e-commerce. Key areas of convergence include decision support, personalization, and efficiency optimization. However, gaps persist in empirical validation, interpretability, real-world deployment, cross-cultural applicability, and ethical governance. Future work must bridge these domains through robust, scalable, and user-centric solutions.

IV. Methodology

A. Fuzzy Logic

For our analysis, we implemented a fuzzy logic system to test the findings of the literature. In real-world problems, things are rarely black or white. In traditional computer science, binary relationships are used in the form of booleans, but they can not really capture the intersections between different states. So to capture this effect, we use fuzzy logic. It is particularly useful, especially in e-commerce, because there is an inherent delivery risk associated with orders. A delivery isn't always 100

Traditional models like K-Means or hard decision trees force each case into a strict box. But fuzzy logic lets us handle uncertainty and rate the risk on a scale, rather than just label it as "Safe" or "Risky." That flexibility is what made it a great fit for our use case. Following is its implementation:

1) Selected and Cleaned the Data: We began by picking the columns that influence delivery: things like the order's fulfillment method, courier status, quantity, amount, and whether it was a B2B sale. We filled in missing values and dropped any rows where core fields were completely missing.

2) Converted Text to Numbers: Since fuzzy logic systems can't understand text like "Shipped" or "Amazon", we used label encoding to convert all categorical fields into numbers. These were used as inputs for the fuzzy system.

3) Scaled the Numbers: Quantity and amount vary widely, so we scaled both to a 0–1 range. This ensures that no one feature dominates the fuzzy decision-making just because it has bigger numbers.

4) Defined Fuzzy Variables and Rules: We told the system how to interpret low/high values:

- For qty and amount, we used automatic fuzzy categories like "poor", "average", "good."
- For courier_status, we manually created categories like "unknown," "unshipped," "cancelled," and "shipped."
- For b2b, we used a simple yes/no distinction.

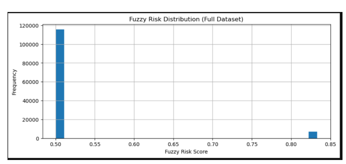
We then wrote rules like:

- If quantity is good, amount is high, and the courier is shipped, then risk is low.
- If quantity is low and status is cancelled, then risk is high.
- If courier is unshipped or unknown, risk is medium.
- If it's a B2B order and quantity is good, risk is medium.

These rules formed the basis of how fuzzy logic would assign a risk score to each order.

5) Ran the Fuzzy Inference on the Whole Dataset: Instead of testing on a sample, we processed the entire dataset (over 128,000 rows) and gave each order a fuzzy risk score between 0 and 1.

6) Distribution of Fuzzy Risk Scores: As it can be seen, most of the orders received a score close to or less than 0.5, which means they were classified as "medium risk." A small cluster landed around 0.8+, which we marked as "high risk." Interestingly, no orders were marked as "low risk." That might be a sign that our rules are conservative — they lean toward caution.



7) Risk Category Counts: Out of all the orders:

- 115,422 were labeled as Medium Risk
- 6,872 as High Risk
- 0 as Low Risk

DeliverySuccess	0	1
risk_category		
Medium	86661	28761
High	6864	8

These findings make sense in our context, given our rules. We only defined *low risk* when three positive conditions were met (good qty, good amount, shipped), which seems rare in this dataset.

8) Comparison with Actual Delivery Outcomes: From the results, we see that:

- Of the high-risk orders, only 8 were successful, while 6,864 failed. That's a 99.9% failure rate, which strongly validates our fuzzy rules.
- Among medium-risk orders, there was a mixed result. About 75% failed, and 25% succeeded.

B. Fuzzy Logic And Monte Carlo Simulations

1) Transaction Utility Theory: This theory says that customers evaluate both *price* and *perceived value* of the transaction. Our findings are in line with the literature. In our case:

- Value = fast and reliable delivery
- Risk = chance of failure (damaging the perceived value)

So, by modeling *delivery success probability* as a fuzzy score, we're directly helping the business optimize for transaction utility. For example:

- Push high-risk orders to safer fulfillment options
- Offer discounts for risky deliveries to offset perceived pain

Random Forest Comparison

We used Random Forests to see which features influence delivery success the most. We used Random Forest to understand which features in our e-commerce dataset actually influence delivery success. In simple terms, Random Forest helps us measure "what matters most" by analyzing thousands of random decision trees. Each tree is built on a slightly different version of the data and helps vote on the output. This makes it very powerful for feature importance analysis—a key step before designing our fuzzy logic system. If we simply included every column, we'd risk adding noise. But with Random Forest, we got a ranked list of variables based on how often and how effectively they helped split the data to predict success or failure. That gave us a clean starting point for our fuzzy rules.

Random Forest is a well-established method in both academic research and industry because it strikes the perfect balance between accuracy and interpretability. It's non-parametric (makes no assumption about how data is distributed), handles both numbers and categories, and is highly resistant to overfitting due to its use of bootstrapping and averaging. In risk modeling or customer analytics, it's often the first thing analysts run—not just to make predictions, but to understand the drivers behind those predictions. That's exactly what we needed: clarity on what features mattered, so our fuzzy model could be meaningful, not just mathematically complex.

2) Random Forest Models Used:
Top Features Only: Fulfilment, Service Level, City **All Features**: Used everything (Order ID, SKU, Courier Status, etc.)

3) Results:

- Top Features gave *sharper* fuzzy separation between delivered vs not delivered (see boxplot).
- Full Model included noise, diluting the signal.
- Random Forests confirmed that the key drivers that are more important compared to rest of the columns.

So, the refined fuzzy model with only important features is **more aligned with actual outcomes**.

4) Monte Carlo Simulations: We tested the robustness of the fuzzy scores by simulating delivery outcomes using noise and repeated sampling.

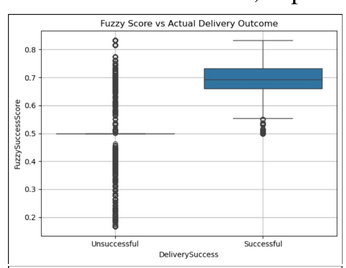
- Full model had mean simulated success 0.535
- Refined model: tighter curve, mean 0.41 (less optimistic, more realistic)

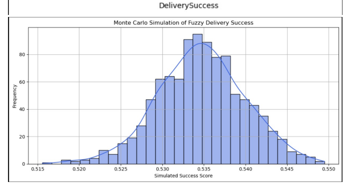
That tells us:

- Simulations confirm that fuzzy scores are meaningful.
- Refined models generalize better.

5) Working of the Model: The first model's logic was broader and captured more relationships, but possibly included some noise or weakly relevant variables. We ran **1,000 simulations** where samples of fuzzy scores were randomly drawn and slight noise was added for each model. We saw that for the first model:

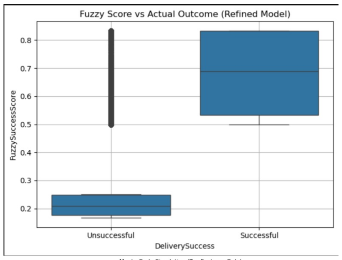
- The distribution of scores was centered around 0̃.53–0.54, indicating a moderate chance of success overall.
- - Spread was wide, which meant that there is more variance in delivery prediction for this model.
- The boxplot of Fuzzy Score vs Actual Delivery Outcome showed some overlap between successful and unsuccessful deliveries — i.e., separation was not perfect.

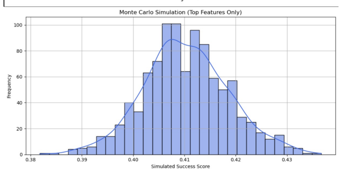




For the second model, we again ran 1000 simulations using only the importance-wise top features of the dataset. We saw that this time:

- Distribution was **tighter and centered around 0̃.41**, showing lower overall expected delivery success.
- The **boxplot showed better separation** between successful and unsuccessful outcomes.
- This suggests that while the refined model may underestimate average success slightly, it was **more consistent and reliable** in distinguishing good from bad delivery outcomes.





Why The Refined Model Is Better

The **refined fuzzy model** using top features performs better because:

- It is simpler and less noisy.
- It aligns closely with actual delivery outcomes.
- It supports better decision-making and classification.

C. RAG using LANG-CHAIN

This project aimed at implementing a Retrieval-Augmented Generation (RAG) pipeline to answer natural language queries from structured business data stored in our dataset in CSV format. The system combines information retrieval techniques with a powerful open-source model, which is Mistral-7b-Instruct, to deliver accurate and content-centered responses.

1) Objectives: Our primary objective was to create an efficient system capable of understanding and answering user queries and questions from sales and P&L reports. Traditional Querying approaches often require SQL knowledge, whereas RAG enables a natural, conversational interface over tabular data.

First of all, we converted rows from multiple CSV files into descriptive text documents. Only relevant columns such as "Date", "Status", "SKU", "Sales Channel", and "Product" were extracted. We then formatted these rows into strings that preserve their tabular meaning. After that, we chunked the data into segments using *LangChain's CharacterTextSplitter* to generate semantic embeddings using the sentence transformers model *all-MiniLM-L6-v2*. We stored these embeddings in an FAISS vector store for efficient similarity-based retrieval.

2) Local Language Model and Query Flow: For the language model component, we loaded the mistral-7b-Instruct model locally using HuggingFace's *transformers* library. As we worked in the Google Colab environment for implementing RAG using Colab's T4 GPU environment, we quantized the model with 8-bit precision using *bitsandbytes*, and we also enabled offloading to prevent runtime crashes. Our model works in the following way:

Whenever a user enters a query, the system retrieves the top-matching document chunks from the FAISS index. These chunks are then passed as context to the Mistral model, with the full prompt. The model then generates a detailed answer based on the retrieved context.

3) Results: The system successfully answered queries such as "What products were cancelled in April?" or "List down all orders shipped by Amazon". The results we got for these queries included multi-line detailed responses listing specific SKUs and statuses. The biggest challenge for us was to overcome HuggingFace model restrictions, accessing gated models, and managing GPU memory for large models. We resolved this by switching to local model loading with authentication and memory-efficient loading strategies.

D. LLM Fine-Tuning

Large Language Models like GPT-2 and LLaMA have demonstrated impressive generative capabilities. However, adapting them to specific domains often requires labeled data for supervised fine-tuning. In this work, we circumvent the lack of labels by applying self-supervised training on business data represented in CSVs, allowing the model to learn language patterns and terminology specific to the domain.

1) Dataset: We collected data from 7 distinct CSV files, each representing a different business domain:

- Sales reports
- Profit & loss sheets
- Expense logs
- Amazon and international sales
- Cloud storage comparison charts

Each row was transformed into a single, structured sentence using the format:

column_1: value_1 — column_2: value_2 — ...

This unification enables token-level modeling of relational data without requiring response fields.

2) Model and Tokenization: We used the pre-trained *gpt2* model and its associated tokenizer:

- *pad_token* set to *eos_token* to accommodate padding
- Tokenization with *max_length=128* and *truncation=True*
- No use of special response fields

3) Fine-Tuning Strategy: We applied self-supervised training using *Trainer* from Hugging Face Transformers:

- Language Modeling Objective: *Causal LM*
- Batch Size: 64
- Epochs: 3
- Tokenized data split: 90% train, 10% test
- Optimizer: *AdamW*
- Collator: *DataCollatorForLanguageModeling*

The training was run on a TPU v2-8 and later tested on GPU and CPU environments for evaluation.

4) Evaluation : Since no ground-truth exists, we employed the following techniques:

5) Perplexity Evaluation:

- Performed on the test split
- Tokenized each row and calculated average cross-entropy loss
- Converted to perplexity: ppl = exp(mean_loss)
- Score obtained was 8.91

6) Qualitative Generation: Sample prompts generated text like:

"Revenue: 100000 — Cost: 75000 — Month: March"

- Decoding used greedy search (do_sample=False) for speed and consistency
- Attention masks passed manually to avoid GPT-2 padding issues

V. Limitations and Future Implications

A. Limitations

The dataset that we worked with was for the apparel industry, so we can not generalize these findings for all industries due to internet differences that lie within sectors. The dataset was also scattered over several months. If we had access to a collective panel dataset that contained transaction records for all local and international orders, the analysis could have been integrated and made more comprehensive. There were only a few customer attributes, with a large set of customer attributes it would have been possible to understand consumer psyche even better.

B. Future Implications

The work presented in this study can be applied by industry practitioners, especially in the e-commerce space. The e-commerce players can use the methods employed in this study to not only scale their businesses but also to manage their existing businesses more efficiently. The decision-making process and interpretation of data can be simplified by using a Generative AI agent, as shown in the article.

VI. Conclusion

Hence, the study confirms that fuzzy logic provides an effective method to assess delivery risk in e-commerce, especially when coupled with data-driven rules and transaction theories. Random Forest and Monte Carlo simulations validated our risk model, while generative AI models enabled natural querying over tabular data. These techniques together offer scalable, interpretable, and intelligent tools for decision-making in online retail.

References

[1] H. Liao, "E-commerce Live-Streaming Platform and Decision Support System Based on Fuzzy Association Rule Mining," *Int. J. Comput. Intell. Syst.*, vol. 18, no. 1, Feb. 2025.

[2] W. Liao, "Uncertainty Handling and Decision Optimization of Fuzzy Logic Algorithm in E-commerce Sales Forecasting," *Int. J. High Speed Electron. Syst.*, Jan. 2025.

[3] J. Chen and Y.-W. Chang, "How retailers launch IT-based strategies to win new and loyal customers in O2O commerce," *Electron. Markets*, vol. 35, no. 1, Feb. 2025.

[4] B. Y. Pratama, I. Budi, and A. Yuliawati, "Product Recommendation in Offline Retail Industry by using Collaborative Filtering," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 9, 2020.

[5] S. O. Fageeri, M. A. Kausar, and A. Soosaimanickam, "MBA: Market Basket Analysis Using Frequent Pattern Mining Techniques," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 5s, 2023.

[6] Y. Ding and M. Najaf, "Interactivity, humanness, and trust: a psychological approach to AI chatbot adoption in e-commerce," *BMC Psychol.*, vol. 12, 2024.

[7] M. Mariani and Y. K. Dwivedi, "Generative Artificial Intelligence in Innovation Management," *J. Bus. Res.*, vol. 175, 2024.

[8] N. D. Kulkarni and S. Bansal, "Exploring Real-World Applications of GenAI in Retail," *J. Artif. Intell. Cloud Comput.*, vol. 2, 2023.

[9] N. Kshetri, "Generative Artificial Intelligence and E-Commerce," *Computer*, vol. 57, no. 2, pp. 125–128, Feb. 2024.

[10] Anand, "Leveraging Artificial Intelligence for Enhanced Personalization," *ASRJETS*, 2024.

[11] S. Zhang et al., "Real-time task planning for order picking in intelligent logistics warehousing," *Sci. Rep.*, vol. 15, no. 7331, Mar. 2025.

[12] S. Geetha et al., "High-performance fake review detection using pretrained DeBERTa optimized with Monarch Butterfly paradigm," *Sci. Rep.*, vol. 15, no. 7445, Mar. 2025.

[13] E. S. P. Krishna et al., "Enhancing E-commerce recommendations with sentiment analysis," *Sci. Rep.*, vol. 15, no. 6739, Feb. 2025.

[14] Z. Lin et al., "Optimization design of cross border intelligent marketing management model," *Sci. Rep.*, vol. 15, no. 5150, Feb. 2025.

[15] M. R. R. Rana et al., "BERT-BiGRU-Senti-GCN," *Int. J. Comput. Intell. Syst.*, vol. 18, Art. no. 21, Feb. 2025.

[16] "El comercio electrónico como estrategia para promover la competitividad en MYPES," *Scopus*, 2025.

[17] X. Huang and X. Liu, "Understanding complaints: the role of moral transgressions in e-commerce interactions," *Humanit. Soc. Sci. Commun.*, vol. 12, Jan. 2025.

[18] L. Pu et al., "The potential of AI tools in shaping digital consumers' behavior," *Asia Pac. J. Mark. Logist.*, Feb. 2025.

[19] X. Chen et al., "Bayesian learning-based agent negotiation model to support doctor-patient shared decision making," BMC Med. Inform. Decis. Mak., Feb. 2025.