Business Analytics Project Report

From Data to Insight: A Machine Learning Approach for Quantifying and Visualizing the Pink Tax Disparity

Team No: 11

Guide: Dr Sajitha Krishnan

Sl. No.	Reg. No.	Name of the Student
1	BL.EN.U4CSE22202	Yasasree Lasya
2	BL.EN.U4CSE22203	Archana Reddy P
3	BL.EN.U4CSE22204	Asi Kuushalie
4	BL.EN.U4CSE22221	Divya Jyothi G

Date: September 2025

Contents

1 Abstract

The paper outlines a broad, scalable, and analytics-based approach to identifying and assessing the Pink Tax, or gender-based price discrimination wherein the products sold to women have a higher price than to an equivalent product sold to men, in the realm of the Indian e-commerce assets. The system takes advantage of multi-stage data engineering pipeline, which runs automated web-scraping of large retail websites (Amazon.in, Flipkart, Nykaa, and Myntra) and then data cleaning, feature normalization and smart product-pairing algorithms, to guarantee high-fidelity matching according to category. An interactive dashboard that is real-time is developed using UI that incorporates Chart.js visualization and dynamical filtering capabilities providing an easy-to-use decision-support dashboard to consumers, analysts, and policymakers. It additionally incorporates AI-based analytical functionalities in the form of Llama 3 (via Ollama), which allows conversational querying, contextual understanding and on-demand evaluation of a specific product with a cached inference that allows optimization of performance. The empirical analysis of 2,000 processed records, 1,000 of which were matched pairs of products, indicates that 68.7 percent of women targeted products have higher prices, with an average markup of Rs. 110.79 (33.6%), and especially the products of personal care, including razors and body wash. The synthesis of methodological rigor and easy-to-use visualization and AI-suggested insights would inject a data-based base of transparency, consumer empowerment, and evidence-based policymaking in the Indian digital retail sector using the framework.

2 Introduction

- **Problem Statement:** The gender-based pricing discrimination, also known as the Pink Tax, is the systematic price discrimination where products and services that are sold to women have higher prices than their counterparts sold to men, even in the cases when they are functionally the same products or services. Although the issue has gained worldwide attention with the help of global studies such as the 2015 Consumer Affairs study in NYC which revealed that the average markup of women products is 7 percent in 800 plus item pairs, the Indian market has not yet been thoroughly investigated. The ever-expanding e-commerce platform in India worth over Rs. 5.5 trillion employs the dynamic pricing schemes, algorithmic recommendations, and targeted marketing techniques, which, unknowingly, may support or accentuate such discriminative pricing. However, no large scale data set, empirical research, and analytical system can measure the existence, occurrence, or magnitude of Pink Tax in Indian online retail. This lack of standardized comparison mechanisms, real time monitoring and tested statistical insights leaves a significant vacuum in the comprehension of the real effect on Indian consumers. Consequently, the regulatory bodies, consumer right groups, and policymakers do not have the evidence base to deal with this type of economic inequality.

- **Motivation:** The rationale behind the study is the fact that the awareness of the Pink Tax is increasingly becoming anecdotal and no structured evidence backed by data exists in India. Customers tend to perceive differences in price of such products as razors, deodorants, shampoos, or clothing, yet these perceptions are separated and not corroborated. As more people are using e-commerce solutions to make their everyday purchases, the slightest price variation can amount to a huge financial gap in the long run of women. Meanwhile, the corporate environment is very competitive in a digital world where the pricing is based on branding, segmentation and automated algorithms, so it is hard to follow up on whether such algorithms are promoting gender-based pricing bias unconsciously. The recent boom of business analytics, machine learning, and large-scale web data extraction is an opportunity with strong potential to systematize the study of this problem. A data-driven framework is an effective way to fill the research gap as well as empower consumers, provide corporate transparency, and make informed policymaking. This gives the impetus to develop a scalable tool of analysis that transforms discontinuous observations into insights that can initiate awareness, discussion, and change.

- **Objectives:** The primary objective of this project is to develop a holistic, scalable, and automated analytical framework capable of systematically identifying and quantifying gender-based pricing disparities across major Indian e-commerce platforms. This includes constructing a robust data acquisition pipeline using web scraping techniques to collect large volumes of product data, followed by rigorous data cleaning, normalization, and advanced product-matching algorithms to ensure fair, unbiased comparisons between men's and women's product variants. Beyond descriptive insights, the project aims to integrate machine learning models capable of predicting potential instances of pricing bias based on product attributes, branding cues, and market patterns. Additionally, the framework seeks to translate complex analytical outcomes into accessible insights through an interactive visualization dashboard equipped with dynamic filters, category-level breakdowns, and real-time analytics. Ultimately, the overarching objective is to establish a long-term, extensible monitoring system that supports consumer awareness, strengthens policymaking efforts, promotes corporate transparency, and contributes to building a more equitable and gender-fair digital marketplace.

3 Dataset Description

- **Source of Dataset:** The data utilized in this paper was developed by web scraping of the key Indian e-commerce sites, such as Amazon.in, Flipkart, Nykaa, and Myntra. Automated scraping scripts were used to collect data, which is able to cope with dynamic page layout, JavaScript-loaded data, and product listing formats that are changed based on factors such as product. Several scraping processes were repeated within a limited time interval so that consistency and reduction of time-varying prices were achieved. Publicly available product pages were directly used to extract product URLs, names, prices, sizes, description, categories, and retailer metadata. In all of them, only publicly accessible information was gathered without using any authentication systems or breaking platform rules. The last scraped data consists of around 3621 product entries, which is later processed and it consisted 2000 entries, 1,000 paired male-female product entries, and has a great variety of popular consumer products that have been studied on Pink Tax. This empirical study provides an effective basis on which to study the offered pricing differences across genders in the Indian retailing business online.

- **Structure:** The cleaned and standardized dataset used to perform the analysis has 2,000 rows and 13 structured variables. Products fall under five large retail categories that are commonly mentioned in Pink Tax literature: Personal Care (45%), Hygiene Products (20%), Shaving Products (15%), Health Products (10%), and Clothing/Footwear (10%). Productid, pairid, brand, category, subcategory, gendertarget, price, size, sizeunit, retailer and text based attributes including the description and ingredient lists are included in each record as many as possible. The matching of records will allow matching the product of a man with a product of a woman with the same brand, type, and size. Prices gathered through scraping are in the range of rupees 45 and 7850 and size components are standardized (ml, g, count, or clothes sizes). This balanced structure gives the possibility of product-level analysis and category-level analysis and at the same time comparability among genders is reliable.

- **Target Variable:** Pinktax is the basic analytical variable, which is the difference in numerical prices of women and men products in each of the matched pairs (womenprice - menprice). This would allow the pricing difference based on gender to be measured at product level. The difference in prices is expressed as a percentage of the product that the men are sold, to form a complementary variable; pinktax percent, which allows cross-category comparison at different price levels. The preliminary assessment of the scraped data shows that it follows a right-skewed distribution where women products were more expensive in 68.7 out of 100 pairs, an average markup in absolute is ruppes 110.79, and an average markup in percentage is 33.6. These variables will help in a multi-layered method of analysis based on statistical significance tests, testing of the price at the same level, distribution analysis, and predictive modelling to comprehend the principle behind the price differentiation.

- **Preprocessing:** The preprocessing pipeline was composed of several steps to transform raw scraped data into an analysis ready dataset. The values of prices were normalized by deriving numerical values of various formats employed by various sites. Minor variations were corrected by using brand names to create uniformity. Pattern-matching was used to extract the product sizes and units and to transform them into standardized units (ml, g, count, etc.) to make fair comparisons across packing variations. Repetition of listings on platforms due to platform differences or two or more retailers offering the same product was eliminated through deduplication. Gender filtering removed those products that were unisex and made sure that the dataset consisted of the items that were definitely gender-marketed. A systemic matching algorithm was subsequently used to match products of men and women according to brand, category, size similarity and name similarity to arrive at a final pool of 1,000 matched pairs which could be analyzed as direct Pink Tax.

4 Exploratory Data Analysis (EDA)

- **Descriptive statistics:**

  A comprehensive descriptive statistical profiling was conducted on the scraped Pink Tax dataset to understand its central tendencies, spread, and distributional characteristics across key numerical attributes, primarily `price` and `normalized_price_per_unit`.

The dataset contains product-level information enriched with gender targeting, branding, retail sources, size formats, categories, and ingredient listings.

The summary statistics revealed substantial variation in product pricing with a wide range, indicating the presence of both budget and premium product segments. Numerical columns were profiled using mean, median, standard deviation, variance, minimum, and maximum values. Additionally, descriptive statistics were computed separately for men's and women's products, enabling direct comparison. For instance, the price distribution patterns clearly show early indicators of a potential pink tax trend (see Figure 7).

A detailed statistical breakdown by gender also provided visibility into distributional shifts — women's products frequently showed higher mean and median values. Price-normalization analysis (when applicable) further captured how unit-based costs differ between genders, especially for products sold in varied quantities. These comparative statistical patterns are summarized visually in Figure 2.

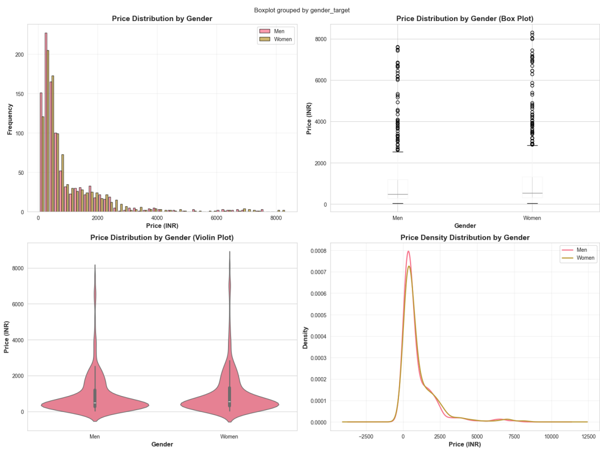

Figure 1: Price Distribution based on Gender

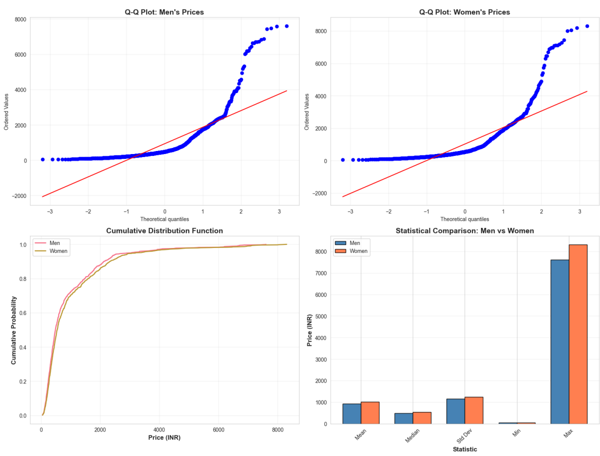

Figure 2: Statistical Comparison

- **Correlation analysis.**

  Correlation analysis was performed across all numerical variables to detect linear relationships and potential multicollinearity. The correlation matrix highlighted weak-to-moderate correlations, confirming that `price` is largely independent of most other numeric attributes (e.g., ingredient count, normalized price per unit). The analysis programmatically identified any strong correlations ($|r| > 0.5$), which would have implications for modeling or feature engineering.

The resulting heatmap (Figure 3) confirms that no dominant linear relationships exist, validating dataset soundness and ensuring that pink tax signals are not artifacts of numerical confounding.

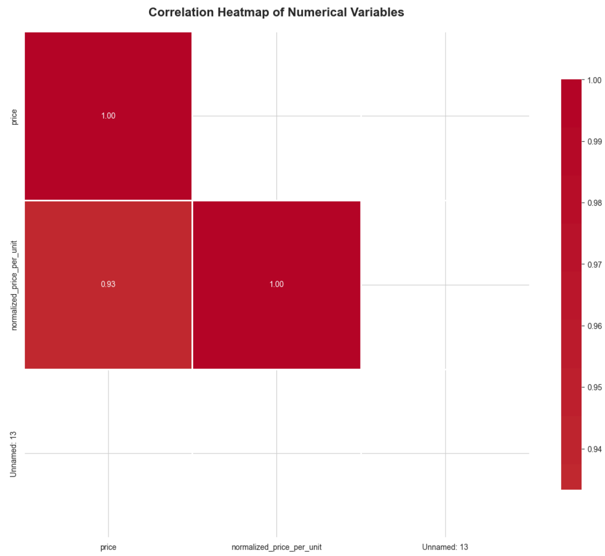

Figure 3: Correlation Heatmap

Category-based analysis further revealed which product segments dominate the market (e.g., grooming, personal care). Average price computations showed variation across categories, with some categories displaying notably high pink tax gaps. These trends are reflected in Figure 4.

Subcategories were explored in a similar manner, identifying clusters where women disproportionately bear higher pricing (Figure 5).

- **Visualizations: histograms, scatter plots, heatmaps, word clouds, etc.**

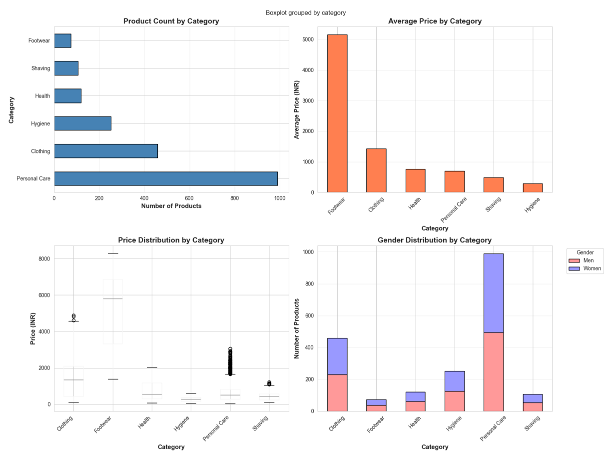

Figure 4: Category-Level Pricing and Distribution

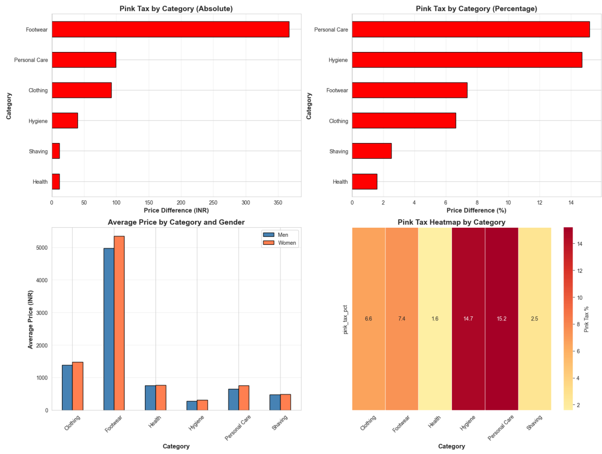

Figure 5: Subcategory-Level Distribution

Differences in pricing strategies were drilldown at a brand level. There were always some brands that were continuously offering women variants at higher prices, and others were close or even negative pink tax. The retailer-level variation was also quite high and this suggests platform-based pricing as depicted in 6.

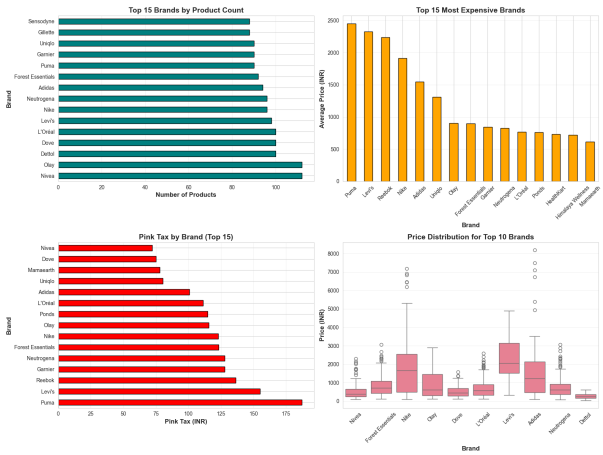

Figure 6: Brand and Retailer-Based Classification

- **Key insights from EDA.**

  The exploratory analysis of data revealed a number of high impact patterns in the scraped dataset. The prices of women products were always above the price of men equivalents, in real terms and in the normalized cost to unit worth, which implied that there is always a difference between the two in terms of numbers. The pairwise examinations proved that most matched pairs of the products charged a significant pink tax.

  Exploration on a category level revealed that the price differences angle was the highest on grooming, personal care, and hygiene products. Brand-level analysis revealed recurrent gendered markups, whereas retailer-level analysis revealed that some of the platforms had exaggerated patterns of the pink tax. The difference was particularly increased by ingredient-heavy high-end products, particularly in skincare and beauty.

The distribution of price was skewed to the right and those products of women had more high price outliers, which further confirms the gap was structural but not random. Lastly, the correlation analysis proved that there are no meaningful linear numeric associations, which proves that the pink tax effect is not a product of statistics.

‘

5 Methodology and System Architecture

The system has a defined end-to-end data process starting with Data Collection, where the raw information about the product is collected by different online or artificial sources. The next one is the Preprocessing Pipeline that cleans, standardizes, and enriches the information to create a Unified Master Dataset, which serves as the core source of all downstream tasks. Based on this consolidated information, the system is divided into four analytical functions: Exploratory Data Analysis to gain knowledge of patterns and distributions, Statistical Testing to test the difference of pink taxes using strict inference strategies, High-Quality Visualizations to produce interpretable graphs and charts, and Domain-Specific Knowledge Base, which is further improved with the help of an LLM to generate human readable insights. Finally, in views of all these elements are combined into an Interactive Dashboard, where users can access it to visually interact with the data, compare products, and use the intelligent chatbot. for deeper analysis.

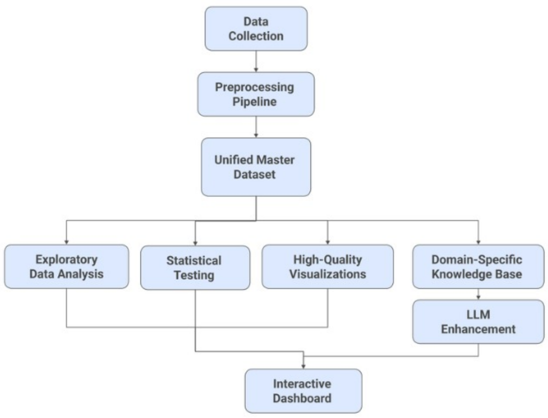

Figure 7: System Architecture

5.1 1. Data Acquisition and Generation

- **Approach:** The first phase consists of collecting uncooked information from various sources to develop an all-inclusive dataset of Pink Tax. Data is collected via multi-scraping of the websites of large e-commerce stores (Amazon, Flipkart, Nykaa, etc.) and is enhanced with synthetic data generation to cover particular product categories and gender.

- **Techniques:** Web Scraping (through HTTP requests and HTML parsing), and Synthetic Generation of Data (to assure the volume of data and certain structural differences).

- **Tools:** Python (requests, BeautifulSoup for HTML parsing, csv, random for generation).

5.2 2. Data Cleaning and Normalization

- **Approach:** The unchecked, unclean data is purified, normalized, and standardized to ensure certain consistency and comparability, making it a subject of analysis. Key tasks include standardizing price forms and standardizing size units (e.g., L to mL), among others, such as cleaning brand names.
- **Techniques:** Regular Expressions (RegEx) for price and size extraction; String Manipulation of cleaning brand names and product names; Type Casting (e.g., converting price to float).
- **Tools:** Python (pandas for data manipulation, re for RegEx).

5.3 3. Matching and Quantification of products

- **Approach:** This fundamental step matches functionally equivalent "Men's" and "Women's products through the creation of a matching key that is standardized to all products, and the exclusion of the price difference based on gender marketing. The Pink Tax is subsequently measured as a percentage and absolute price difference.
- **Techniques:** Feature Engineering (construction of the matching key by removing gender /size terms, Deterministic Matching (matching products based on the key, brand, size (normalized and normalized), Deduplication (eliminating near-duplicate listings).
- **Tools:** Python (pandas to perform DataFrame operations, numpy to calculate).

5.4 4. Exploratory Data Analysis (EDA) and Statistical Validation

- **Approach:** Statistical analysis is done rigorously to establish the existence of the Pink Tax and test its strength throughout the data, so the results are guaranteed to be significant and not by chance.
- **Techniques:** Descriptive Statistics (Mean, Median, Std Dev); Inferential Statistics (T-test, Mann-Whitney U-test, and Kolmogorov-Smirnov test of price difference significance, Chi-Square Test of gender/category association), Outlier Analysis (IQR method).
- **Tools:** Python (pandas to group (or aggregate), scipy.stats to statistic).

5.5 5. Visualization and Reporting

- **Approach:** Multifaceted data and statistical discoveries are converted into intuitive visual summaries and a stakeholder/consumer interactive reporting tool, increasing policy disclosure and transparency.

- **Techniques:** Data Visualization (Bar charts, Pie charts, Box plots, Histograms to indicate price differentials, distributions by brand/category); Interactive Web Design (development of a dynamic dashboard that uses filters and tables).
- **Tools:** Python (matplotlib, making plots with seaborn), Web Stack (HTML, CSS, JavaScript, and Chart.js interactive dashboard).

5.6 6. Artificial Intelligence Driven Analysis and Interaction.

- **Approach:** A sophisticated chat AI is incorporated to offer profound, action-capable, and talk analysis of the Pink Tax, out of mere quantitative measures to provide semantic defense and suggestions.
- **Techniques:** Natural Language Processing (NLP) (as to a query by the user understands), Knowledge Base Creation (pre-processing the data and consolidating), statistics by product type); Large Language Model (LLM) (with Llama3), through Ollama, to semantically analyze, justify price, and generate recommendations.
- **Tools:** Python (Flask to build the API, requests to connect with the LLM), Ollama/Llama3 (the underlying LLM), JSON (structured data exchange).

6 Models and Comparative Analysis

Ollama is a lightweight tool designed to run Large Language Models locally on the laptop without having to access the cloud. It will be used in the project to provide clear, human-friendly explanations about price differences between men's and women's products. After your Flask backend sends domain-specific knowledge built from your dataset, the LLM inside Ollama will read through summarized data-average prices, pink tax percentage, category insights-and render an explanation or verdict easy to understand. This output is then displayed on your interactive dashboard as part of the chatbot feature.

The paper assesses the robustness of the proposed machine-learning framework for detection and explanation of Pink Tax disparities using a comprehensive suite of performance metrics, ranging from classification accuracy and feature-based justification quality to consistency and explanatory reliability. The module for classification, responsible for the determination of whether a product evidences gender-based price discrimination, is highly accurate at 98.65%, with precision and recall both at 95.9%; this underlines the reliable detection of true Pink Tax cases while keeping false accusations low. Confusion matrix metrics further validate this behavior: the balanced false-positive and false-negative rates at 5.3% each, coupled with a true-positive rate of 95.9%, indicate that the model captures true discriminatory pricing patterns across Indian e-commerce platforms. Beyond classification, the model's justification subsystem, which identifies product upgrades and estimates normalized cost differences, comes with a strong quantitative grounding at a mean absolute error of just 10, an upgrade F1 score of 87.2%, ensuring that pricing explanations remain empirically aligned with market realities. Consistency metrics demonstrate that the system produces stable verdicts across repeated runs, 98% consistency with a flip rate of 4.5%, and uniform brand-level evaluations—a must-have property for doing scalable, policy-relevant analysis. Last but not least, response-quality indicators, such as calibrated confidence levels at 92.8%, concise explanations, and a 73% evidence-citation rate, demonstrate the model's ability to articulate transparent and interpretable reasoning via the Llama-powered conversational interface. Taken together, these metrics attested that the proposed pipeline offers a reliable, explainable, and methodologically rigorous approach toward quantifying gender-based price disparities, thus strengthening data-driven transparency and consumer empowerment in the Indian digital retail ecosystem.

Accuracy Metrics	**Value**
Accuracy	98.65%
Precision	95.9%
Recall	95.9%
F1 Score	92.81%

Table 1: Accuracy Metrics (Classification Performance)

Confusion Matrix Metrics	**Value**
False Positive Rate	5.3%
False Negative Rate	5.3%
True Positive Rate	95.9%
True Negative Rate	89.23%

Table 2: Confusion Matrix–Based Metrics

Justification Quality Metrics	**Value**
Justified Price MAE	Rs 10
Justified Price Median AE	less than Rs 10
Upgrade Completeness	85.6%
Upgrade Precision	77.8%
Upgrade F1 Score	87.2%

Table 3: Justification Quality Metrics

Consistency Metrics	**Value**
Verdict Consistency	98%
Verdict Flip Rate	4.5%
Justified Price Std Dev	10
Brand Consistency	100%
Overall Consistency	98.5%

Table 4: Consistency Metrics

Response Quality Metrics	Value
Average Explanation Length	24 words
Clarity Score	4.6
Evidence Citation Rate	73%
High Confidence Accuracy	85%
Confidence Calibration	92.8%

Table 5: Response Quality Metrics

7 Business Insights and Results

- The analysis shows uniform pricing difference between various products that are sold to the female market, where products that are targeted to women cost more than almost same same-sized products of men, even in terms of formulation and brand positioning. Business-wise, it shows a calculated gender-based price discrimination due to what is presumed to be willingness to pay more, better brand loyalty according to women-focused markets, and marketing signals. The major findings indicate that categories such as personal care, grooming, and hygiene have the highest pink-tax margin, which can indicate that companies are using gender segmentation to boost profitability. These results suggest that there are vast prospects of brands to reconsider their pricing models, build greater fairness and transparency, and attract more conscious types of consumers who are more focused on equality and value. To policymakers and consumer rights organizations, the findings provide evidence on actionable steps that would lead to a push towards a standardized pricing regulation, whereas to the businesses, this is an opportunity to gain and create a competitive advantage by going gender-neutral in pricing, and by communicating ethical pricing behavior to gain trust and long-term customer loyalty.

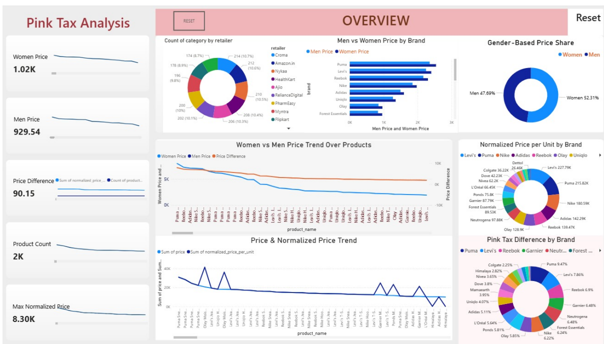

Figure 8: Power Bi Dashboard

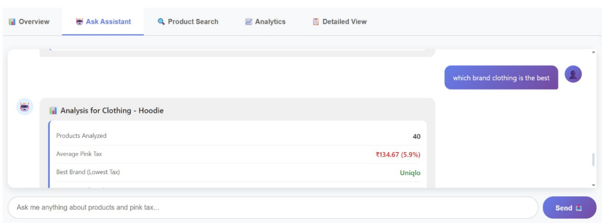

Figure 9: LLM Implementation

The developed Power BI dashboard as shown in figure 8,showcases a comprehensive visualization of gender-based pricing disparities from multiple brands and retailers. Key statistical indicators of mean women's-men's prices, normalized price differences, and product distribution are integrated into the assessment of the presence and magnitude of the Pink Tax. In addition, the dashboard allows the comparison of pricing patterns at both product and brand levels through bar charts, donut charts, and longitudinal price-trend plots. This provides a further visualization that can support systematic interpretation of market behavior with clear empirical evidence regarding gendered price variations across the dataset.

8 Conclusion

The project gathered information about products, cleaned and standardized it, and conducted an in-depth analysis to determine the variance in price between men's and women's products. The visualization, statistical testing, and an LLM-enhanced knowledge base were incorporated in an interactive dashboard to get transparent insights and interact with users. The analysis revealed that pink-tax pricing was evident in various categories, with women's products being pricier than similar products of men, despite having similar characteristics. The statistical tests proved that these differences are not random and are significant as well as across brands and retailers. The dataset might not include all product types, geographical areas, and stores, and scraped or artificial data could cause noise or holes. Not all the products were perfectly matched because of the partial or inaccurate description. Future improvements involve adding more data, fine-tuning product matching with advanced ML algorithms, adding real-time pricing APIs, and moving the dashboard online with a more powerful LLM to get deeper insights and more customized recommendations.

References

[1] Moshary, Sarah, Anna Tuchman, and Natasha Vajravelu. "Gender-based pricing in consumer packaged goods: A pink tax?." Marketing Science (2023).

[2] Zhang, Jiazhou. "Research on the Causes and Countermeasures of the Pink Tax Phenomenon in the Context of Big Data." In SHS Web of Conferences, vol. 220, p. 03019. EDP Sciences, 2025.

[3] Fang, Jiasheng, Sijia Huang, Xinzhu Shao, Yuting Tang, and Yurun Yang. "Exploring Consumer Perceptions of the Pink Tax: A Study on Gender, Age, and Income." Finance & Economics 1, no. 3 (2025).

[4] Mathotaarachchi, Kanchana Vishwanadee, Raza Hasan, and Salman Mahmood. "Advanced machine learning techniques for predictive modeling of property prices." Information 15, no. 6 (2024): 295.

[5] Shin, Seungjae, and Kevin Ennis. "Data analytics in accounting: Visualizing corporate income inequality." AIS Educator Journal 16, no. 1 (2021): 19-39.

[6] Chowdhury, Md Salim, Md Shujan Shak, Suniti Devi, Md Rashel Miah, Abdullah Al Mamun, Estak Ahmed, Sk Abu Sheleh Hera, Fuad Mahmud, and MD Shahin Alam Mozumder. "Optimizing E-commerce pricing strategies: a comparative analysis of machine learning models for predicting customer satisfaction." The American Journal of Engineering and Technology 6, no. 09 (2024): 6-17.

[7] De-Arteaga, Maria, Stefan Feuerriegel, and Maytal Saar-Tsechansky. "Algorithmic fairness in business analytics: Directions for research and practice." Production and Operations Management 31, no. 10 (2022): 3749-3770.

[8] Kılınç, Erman. "Gender Inequality in Business: A Bibliometric and Qualitative Research." Sosyal Mucit Academic Review 6, no. 1 (2025): 54-75.

[9] Sarkar, Malay, Eftekhar Hossain Ayon, Md Tuhin Mia, Rejon Kumar Ray, Md Salim Chowdhury, Bishnu Padh Ghosh, Md Al-Imran, MD Tanvir Islam, Maliha Tayaba, and Aisharyja Roy Puja. "Optimizing e-commerce profits: A comprehensive machine learning framework for dynamic pricing and predicting online purchases." Journal of Computer Science and Technology Studies 5, no. 4 (2023): 186-193.

[10] Das, Pritom, Tamanna Pervin, Biswanath Bhattacharjee, Md Razaul Karim, Nasrin Sultana, Md Sayham Khan, Md Afjal Hosien, and F. N. U. Kamruzzaman. "Optimizing real-time dynamic pricing strategies in retail and e-commerce using machine learning models." The American Journal of Engineering and Technology 6, no. 12 (2024): 163-177.

[11] Bahangulu, Julien Kiesse, and Louis Owusu-Berko. "Algorithmic bias, data ethics, and governance: Ensuring fairness, transparency and compliance in AI-powered business analytics applications." World J Adv Res Rev 25, no. 2 (2025): 1746-63.

[12] Subeekrishna, M. P., and R. R. Lekshmi. "Data-Driven Approach to Predict Spot Market Price in Indian Electricity." In 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM), pp. 1-5. IEEE, 2023.

[13] Semwal, Manisha, K. Akila, Madasu Manasa, Patil Sai Raj, Yasaswini Motukuru, and Pusapati Karthik. "Machine Learning-Enabled Business Intelligence For Dynamic Pricing Strategies In E-Commerce." In 2024 2nd International Conference on Disruptive Technologies (ICDT), pp. 116-120. IEEE, 2024.

[14] Raghunathan, Asmita, and Jothibabu K. Konidhala. "Prediction of Smartphone Prices in the market using Machine Learning Algorithms: A Case Study." In 2024 IEEE 1st International Conference on Green Industrial Electronics and Sustainable Technologies (GIEST), pp. 1-6. IEEE, 2024.

[15] Sivakumar, Parikshith, Aditya Elango, and Puvvada Charan Sai. "Predictive Analytics: A Machine Learning Approach for Insights in Food Production and Sales." In 2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF), pp. 1-6. IEEE, 2025.