

Markets products Analysis

Youssef Abbas, Abdelrahman Ahmed Medhat

Artificial Intelligence Department, Nile University University
Address

Yo.abbas@nu.edu.eg

A.ahmedmedhat@nu.edu.eg

The analysis of product pricing and market trends is an important aspect of modern business. In this paper, we present a data mining project focused on analyzing product pricing and discount trends across three different markets. The project's main goal is to determine the range of product prices, count the products with discounts in each market, and compare the prices of products across the three markets to identify the best-priced supermarket.

Keywords: Market analysis, Product pricing, Discount analysis, Market comparison, Market trends, Data visualization

I. INTRODUCTION

The document addresses the need for businesses to understand pricing and discount trends in different markets to make informed decisions. By analyzing the products' pricing and discount trends across three markets, we can identify patterns and insights that can help businesses improve their pricing strategies and increase their competitiveness. Our project also demonstrates the effectiveness of data mining techniques and tools such as Power BI in analyzing complex data and identifying hidden patterns and insights.

II. WEB SCRAPPING AND COLLECTING DATA

To collect our data, it was the best choice to scrap our target data from the supermarket's websites. While scrapping we were concerned about the product barcode, product name, price before discount, price after discount, and the discount percentage. That was applied to all markets and extracted to a CSV file. Furthermore, we were concerned about the origin of the products that weren't so easy to get the product's origin from the markets websites. The best way was to understand the barcode sequence and that each value has a meaning. After some research, we found that the first three digits in the barcode are concerned about the origin of the product. In addition, we scrapped the list of country codes to merge the product's first three digits and the country codes list.

III. DATA CLEANING AND ANALYSIS

To visualize our data and merge between them we were in need to use PowerBi one of the best tools for data analysis. For the market data, we work on data cleaning by removing duplicates, and damaged data. Also, it was important to

extract the first three digits from the barcodes to be as the same type as the country codes. In addition, we added another column to make a Boolean expression for the products with discounts, where to put zero beside the product without discounts and one for the discounted product.

IV. DATA MERGING AND VISUALIZATION

By using Power Bi, it was easy to merge between markets and country codes. First, it was important to merge the extracted three digits with the country codes to get the origin for each product. Second, the merger of the A market with C market and B market to get the common products between them. In addition, the merge for B market with C market. To visualize the data for each supermarket, the use of pie charts, clustered column charts, donut chart, and multi-row chart was very specific in visualizing products. To add, we visualize the common products in the three supermarkets to get the best price for the products.

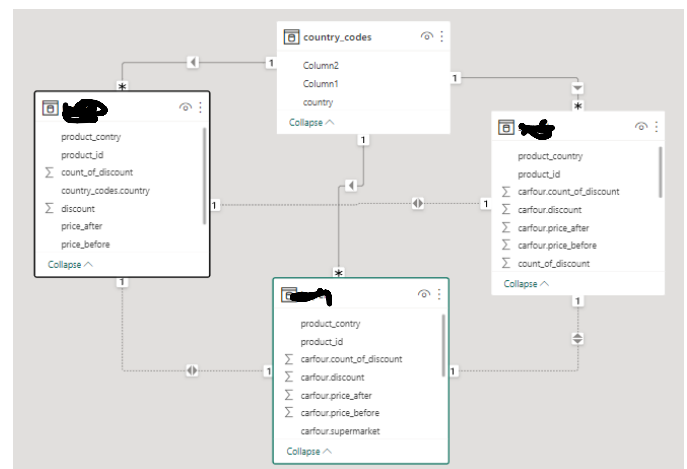


Fig. (1)

V. RESULT

As a result of each visualization and analysis, it was shown that:

- A market data set consists of 3622 products where they were from 57 countries. 1436 products were from Egypt, while 247 were from Italy, San Marino, and Vatican City. Also, it consists of 219 products with

discounts. The range of the prices of products starts at 0.3 and ends at 4631.

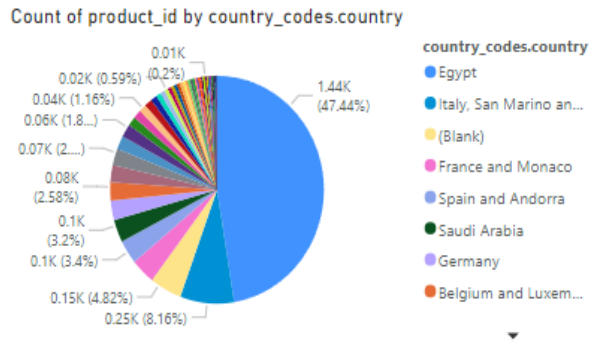


Fig. (2)

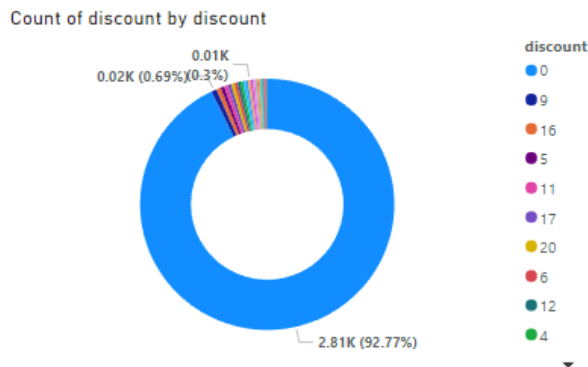


Fig. (3)

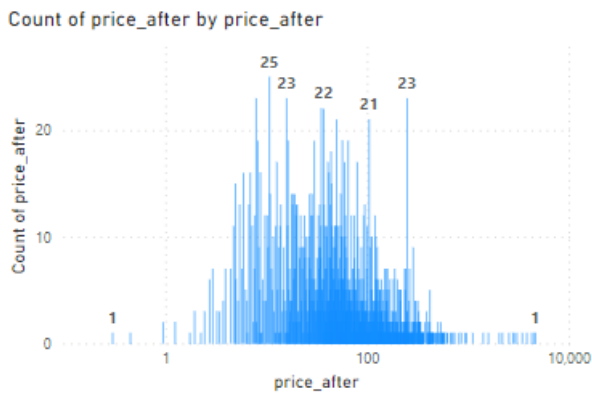


Fig. (4)

Fig. (2) Count of products was highest for Egypt at 614, followed by Belgium and Luxembourg and France and Monaco Count of products was highest for Brazil at 5. Egypt accounted for 80.16% of Count of products. Fig. (3) The discount was the highest for 0 at 642, followed by 8 and 19. 0 accounted for 83.81% of Count of discount. Fig. (4) 23.75 had the highest Count of price after at 9, followed by 16.50 and 34.95, which tied for second at 8. 23.75 accounted for 1.17% of Count of price after. Across all 413 price after, Count of price after ranged from 1 to 9.

b. B market data set consists of 1715 products, where they were from 29 countries. 614 products were from Egypt, while 18 were from Italy, San Marino, and Vatican City. Also, it consists of 124 products with discounts. The range of the prices of products starts at 0.75 and ends at 678.

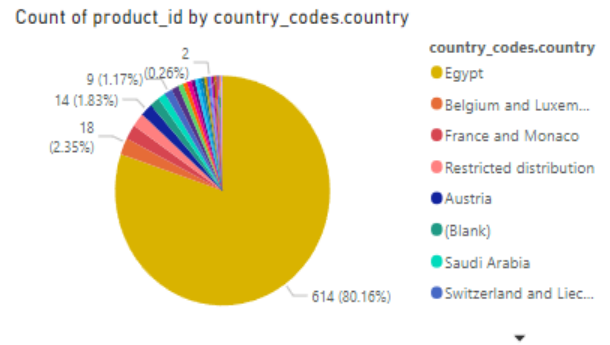


Fig. (5)

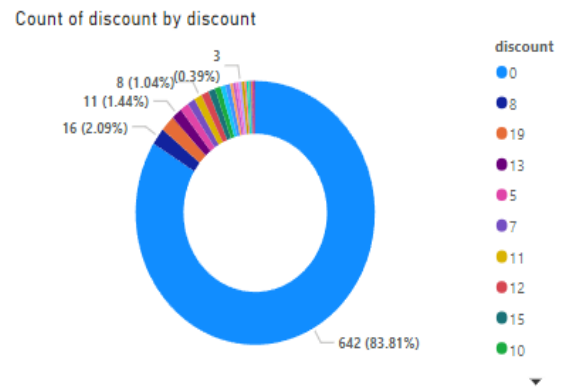


Fig. (6)

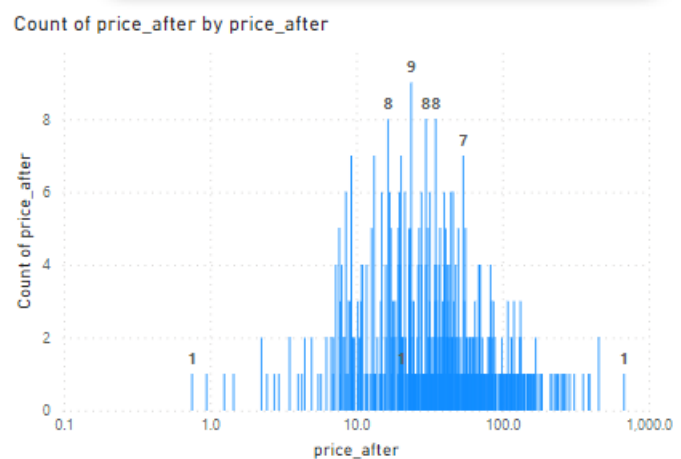


Fig. (7)

Fig. (5) Count of products was highest for Egypt at 614, followed by Belgium and Luxembourg and France and Monaco Count of products was highest for Brazil at 5. Egypt accounted for 80.16% of Count of products. Fig. (5) The discount was the highest for 0 at 642, followed by 8 and 19. 0 accounted for 83.81% of the Count of discount. Fig. (7) 23.75 had the highest Count of price after at 9, followed by 16.50 and 34.95, which tied for second at 8. 23.75 accounted for 1.17% of Count of price after. Across all 413 price after, Count of price after ranged from 1 to 9.

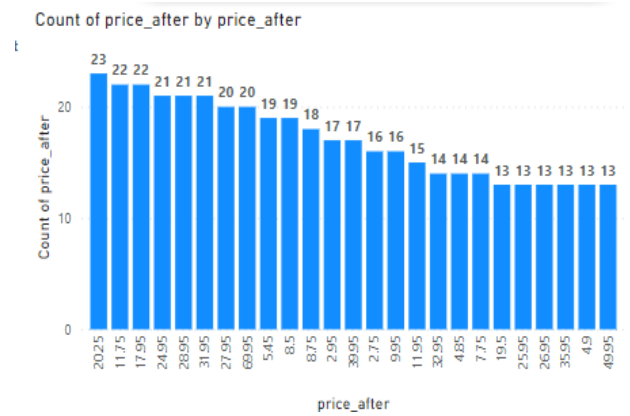


Fig. (10)

- c. C market data set consists of 2033 products, where they were from 33 countries. 1472 products were from Egypt, while 131 were from Restricted distribution. Also, it consists of 301 products with discounts. The range of the prices of products starts at 0.65 and ends at 99.95.

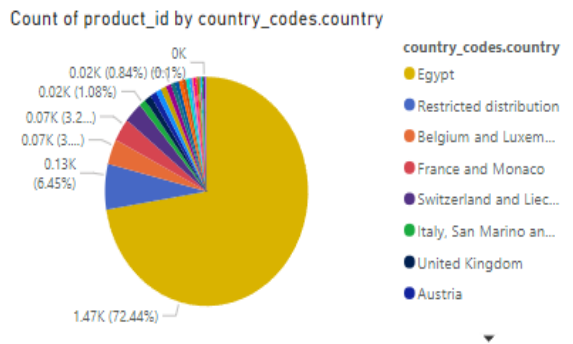


Fig. (8)

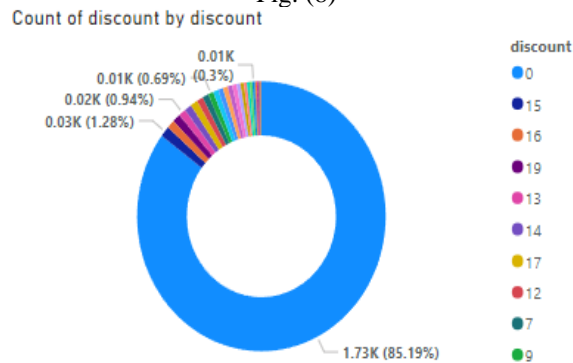


Fig. (9)

Fig. (8) Count of products was highest for Egypt at 1,472, followed by Restricted distribution and Belgium and Luxembourg. Egypt accounted for 72.44% of Count of products. Fig. (9) 0 had the highest Count of discount at 1,731, followed by 15 and 19. 30 had the lowest Count of discount at 1. 0 accounted for 85.19% of Count of discount. Fig. (10) 20.25 had the highest Count of price after at 23, followed by 17.95 and 11.75, which tied for second at 22. 20.25 accounted for 1.13% of Count of price after. Across all 650 price after, Count of price after ranged from 1 to 23.

VI. CONCLUSION

This paper presents a data mining project that analyses product pricing and discount trends across three different markets. The project's primary objectives are to determine the range of product prices, count the products with discounts in each market, and identify the best-priced supermarket. The project utilizes Power BI, a powerful data analysis and visualization tool, to import and clean data from three different Excel sheets. Create visualizations that display the range of product prices, the number of products with discounts, and the relationship between the three markets, and calculate the average price, discount rate, and other relevant metrics for each market and product. Additionally, the project integrates information about the country of manufacture for each product extracted from the barcode. The results provide valuable insights into the three markets' products, their pricing and discount trends, and their relationships. The project demonstrates the effectiveness of data mining techniques and tools in analyzing complex data and identifying hidden patterns and insights.

REFERENCES

1. <https://www.barcodetalk.com/learn-about-barcodes/resources/barcode-country-codes>
2. <https://www.simplilearn.com/tutorials/power-bi-tutorial/what-is-power-bi>

3. Thomas, D. M., & Mathur, S. (2019, June). Data analysis by web scraping using python. In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 450-454). IEEE.
4. Osman, A. S. (2019). Data mining techniques.
5. Montgomery, D. B. (1975). New product distribution: An analysis of supermarket buyer decisions. *Journal of Marketing Research*, 12(3), 255-264.
6. Chevalier, J. A. (1995). Do LBO supermarkets charge more? An empirical analysis of the effects of LBOs on supermarket pricing. *The Journal of Finance*, 50(4), 1095-1112.
7. Anand, J. V. (2020). A methodology of atmospheric deterioration forecasting and evaluation through data mining and business intelligence. *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, 2(02), 79-87.
8. Fatehnia, M., & Deaton, S. Geotechnical Data Analytics and Data Mining Using Microsoft Power BI. In *Geo-Congress 2022* (pp. 553-560).
9. Bhardwaj, B. K., & Pal, S. (2012). Data Mining: A prediction for performance improvement using classification. *arXiv preprint arXiv:1201.3418*.
10. Liu, Y., Zhao, T., Ju, W., & Shi, S. (2017). Materials discovery and design using machine learning. *Journal of Materiomics*, 3(3), 159-177.