

Contents

1 Executive Summary	3
2 Domain Name	4
3 Title of the work	4
4 Background	4
5 Motivation Behind the Study	5
6 Research Questions	5
7 Proposed Solutions	6
7.1 Sales volume by product category	6
7.2 Impact of placement of product	6
7.3 Comparison of promotion vs non-promotion	7
7.4 Revenue by Price Range	8
7.5 Price vs Sales	8
7.6 Performance by section	9
7.7 Top Products (SKUs) by Sales Volume	10
7.8 Seasonal vs Non-seasonal Performance	10
7.9 Under performers	11
7.10 Product features among the top selling candidates	12
8 Data Collection & preprocessing	13
8.1 Data Source	13
8.2 Dataset fields - Raw	13
8.3 Data Preprocessing	13
9 Results and Discussion	14
10 Key findings and Recommendations	15
10.1 Key findings	15
10.2 Recommendations	15
10.2.1 Operational Recommendations for Zara	15
10.2.2 Analytics Recommendations	15
11 AI Contribution	15
11.1 Generative AI	15
11.1.1 Julius	15
11.1.2 Chat GPT	16
11.2 Programming with AI libraries	16
11.2.1 Python	16
11.3 Data Visualization Tools	16
12 Link for the documents	16

1 Executive Summary

This report analyzes Zara's sales performance at the product level by performing data analysis using advanced AI tools & techniques. The main objective of this study is to identify and evaluate all those key factors and drivers using advanced AI tools that impact the pricing, promotions, and seasonal factors of Zara fashion. This was further analyzed to derive some actionable insights to improve their inventory, merchandise and their profits at their corporate competition level.

In this study, Zara fashion sales dataset was downloaded from Kaggle platform. The dataset contains various SKUs like Product ID, Product Position, Promotion, Product Category, Seasonal, Sales Volume, Brand, Url, sku, name, description, price, currency, scraped_at, terms, section (MAN, WOMAN, KIDS). New calculated fields like revenue, price bins, monthly trends, and performance flags were introduced after data cleaning and data pre-processing. This was first made on the Julius AI and Chat GPT platforms, and subsequently dashboards and storytelling were done on Power BI and Tableau, respectively. Python was used to verify the details and accuracy of the AI predictions.

Based on the analysis of JULius AI, Chat GPT, Power BI, and Tableau, it was found that Zara's overall sales volume was dominated by clothing section. This creates a need to assort the strategy within this section. Placement of the clothes impacted the performance, Products placed on the front of the store demonstrated high performance and sales than those placed in aisles and End-caps. This makes the placement a strategy to improve the visibility and significantly the sales. Apart from this, promotions also demonstrated exceptional performance by influencing the sales volume positively. However few margin effects need to be evaluated further to derive key financial insights.

Regarding the pricing strategy, INR 75 to 150 (mid-range) generated the highest revenue. This shows that customers more likely spend in the mid-range band in Zara. Additionally, sales volume and price showed a weak correlation indicating that there are many other drivers which influence the sales. Seasonality effect did not show any strong influence on the sales in this data set. This can imply that more specific analysis of seasonal effect on various categories may be needed.

This analysis showed disproportionalities with the top performing SKUs that drive the Zara's revenue. This shows that targeted promotion is prioritized. Whereas, under performing SKUS represented repositioning of certain categories.

Cloth features were analyzed using text mining and was found that customers were more interested in products having pockets, collar, button, and zip, making them top-performing products at Zara.

In conclusion, this report shows that Top-performing SKU's inventory must be prioritized and by optimizing the placement (front-store, aisle, end-cap), and the promotions will significantly impact and increase the sales performance of Zara. This report will be helpful for optimizing and improving the operations and merchandise inventory of Zara to maximise their profits and overall revenue.

2 Domain Name

Retail industry - Fashion

3 Title of the work

From Raw Data to Strategic Business Insights - A Zara Fashion Sales Case Study

4 Background

Fashion retail is the most evolving industry in today's fast-growing world. In recent decades, many changes have been made to the fashion retail from e-commerce to omnichannel platforms. Zara fashion stands out to be one of the top-performing fashion retail brand among others because of its unique design, strong and adaptive supply chain, and their ability to understand their customers well. Despite being the top performer, Zara still faces few challenges in terms of managing their large inventory, SKUs, seasonal demand, and the placement of products in their store which significantly impacts their profits and revenue.

In today's increasing competition among various brands, traditional practices of maintaining their merchandise alone would not help much. Brands should constantly keep analyzing their store performance by compiling everything in the form of a data set and take strategic decisions accordingly based on the results. Data analysis uncovers all the hidden patterns affecting the performance of the store.

This study explains how the data analysis was done and how the data visualizations helped to derive the insights of the store performance of Zara fashion by calculating the effectiveness of promotions, revenue, profit by category, placement of the products.

Today's market needs speed and the precision of the analysis. Some decisions were often misjudged and this impacts the sales performance. For example, how the placement of products correlates with the revenue. This study further reveals the information about the best performing products, how pricing and the placement should be done and how are they correlated, seasonal demand and its effects, and which features are most important for the products.



Dashboard of Sales Analysis

5 Motivation Behind the Study

The motivation comes from influencing the retail operation directly by converting the raw data into strategic business insights. In fashion retail industry, every small placement to the cost of the product to the stock influences and impacts the overall revenue of the store. Hence, supporting these decisions from the data rather than from simple observation assumptions help run the business smoothly.

Further, several objectives played a role in the motivation. They are:

- Improving the efficiency of the assortment: Zara fashion has various types of products. Identifying and categorizing these products as top-performing and under-performing will help keep the right inventory of the products available during different seasons, depending on the demand for the product.
- Optimizing the placement of the products in the store: Product placement in the store significantly impacts the performance and the profit of the store. Analyzing which placement of the products (front of the store, aisle or end-cap) helps us understand the performance of the product based on placement and take significant steps to allocate the space for the products effectively.
- Evaluating the effectiveness of the promotion: Promoting certain products increases the sales. However, sometimes reduces the margins. Analyzing of promotions really increases the profits or it just shifts the demand of the product is very crucial.
- Pricing: Segmenting the prices into mid, low, high helps understand which pricing band most customers are willing to choose. This helps in fine tuning the pricing strategy of Zara fashion.
- Product features: Text mining reveals if customers are particularly interested in certain product features like zip, collar, pockets, etc. This would further help in designing the product in future and in redesigning their merchandise.
- Business decisions: Dashboards helps in visualizing the data analysis and derive certain key decisions in maximizing profits.

6 Research Questions

Based on the dataset obtained from Kaggle, few business questions were framed to derive actionable insights.

- Which product categories generated highest sales?
- Did placement of products (aisle, store front, end cap) show any impact on sales volume?
- How are promotions effective in increasing the sales?
- At what price range most customers are willing to buy and how is it impacting the sales?
- Are seasonal products sold more than other products?
- Which products are top performing, significantly increasing the sales?
- Derive any correlation if seen between pricing of the product and sales?

- How are various sections (Man, Woman, Kids) impacting the sales performance?
- Are there any slow moving stock that need repositioning?
- Which features in the product description appears the most in the top-selling products?

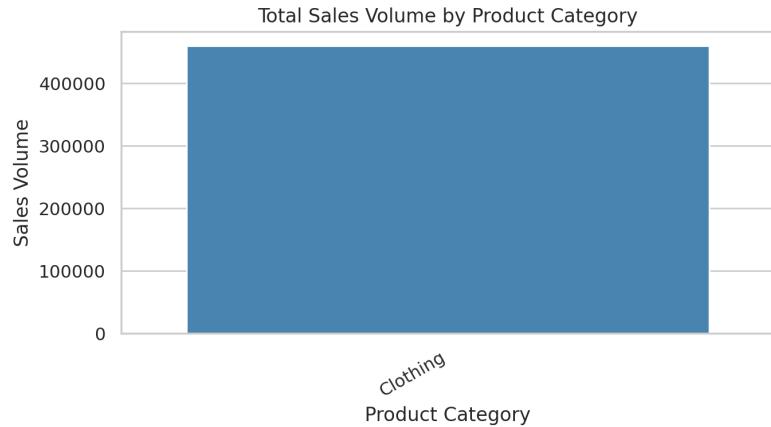
7 Proposed Solutions

This section includes the visuals of the data analysis - Graphs, dashboards, and story telling. These results were generated using Julius AI platform and was verified by Python code.

7.1 Sales volume by product category

	Product Category	Sales Volume
1	Clothing	459573

- Graph: Bar Chart



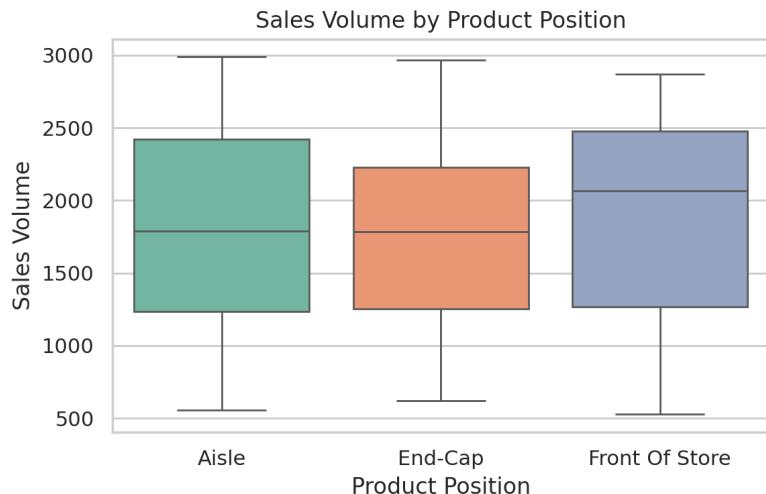
Total Sales Volume by Product Category

Clothing dominates overall sales volume.

7.2 Impact of placement of product

	Product Position	Sales Volume
1	Aisle	1791.0
2	End-Cap	1785.0
3	Front Of Store	2067.0

- Graph: Box Plot



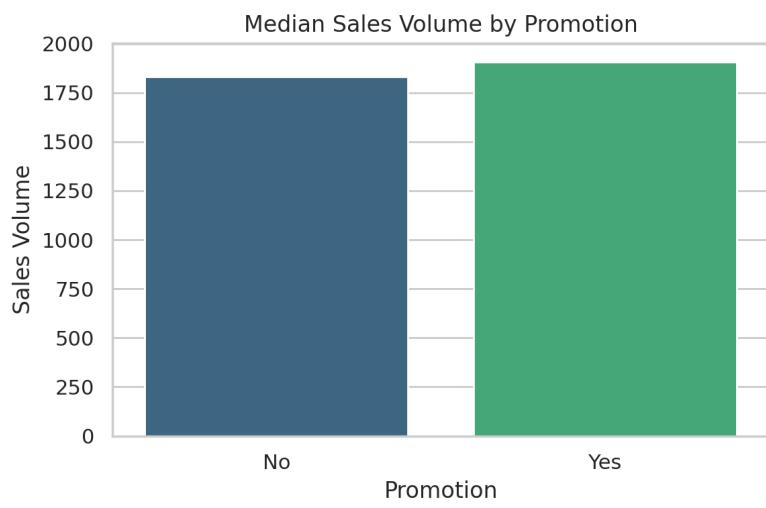
Sales volume by product position

Front of the store placed products generated high sales.

7.3 Comparison of promotion vs non-promotion

	["Promotion",""]	["Sales Volume","median"]	["Sales Volume","mean"]	["revenue","mean"]
1	No	1831.5	1820.5454545454545	143894.61803030304
2	Yes	1907.5	1827.175	166619.89083333333

- Garph: Bar Chart



Median Sales by Promotion

Products that are promoted showed a high median.

7.4 Revenue by Price Range

	price_bin	revenue
5	> 100-150	11220134.0
4	> 75-100	9139765.6
3	> 50-75	6552954.78
6	> 150-200	5467195.0
2	> 25-50	4095231.21

- Graph: Bar Chart



Revenue by Price Range

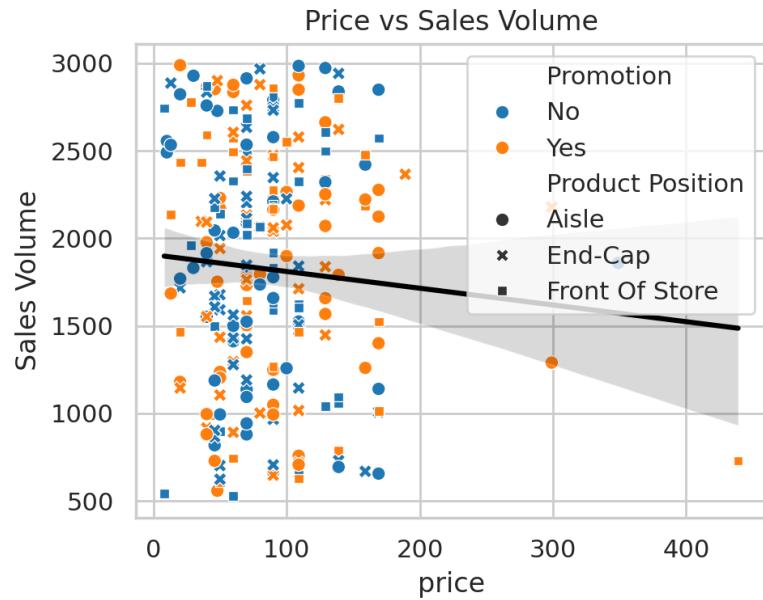
Customers most likely buy in the 75-100 and 100-150 - medium price band. This price range is a driving factor in increasing the revenue.

7.5 Price vs Sales

A correlation analysis

	sku	name	Sales Volume	revenue	price
126	281593208-015-2	PLAID TIE DYE OVERSHIRT	2989	59750.10999999999	19.99
238	313027279-800-2	SUIT JACKET IN 100% LINEN	2985	325365.0	109.0
247	320278659-251-2	RIB COLLAR JACKET	2973	383517.0	129.0
6	320298385-807-2	CONTRASTING COLLAR JACKET	2968	237143.2	79.9
5	312368260-800-2	DOUBLE FACED JACKET	2942	408938.0	139.0
35	317782474-800-2	FAUX LEATHER BOMBER JACKET	2931	319479.0	109.0
67	276213846-500-2	COTTON JACKET	2929	87840.70999999999	29.99
205	311303003-407-38	RIPPED STRAIGHT FIT JEANS	2914	203688.6	69.9
182	322451460-800-1	GATHERED WAIST KNIT SWEATER	2901	138957.9	47.9
207	267195405-500-3	PURL KNIT SWEATER	2887	37502.13	12.99

- Graph: Scatter Plot



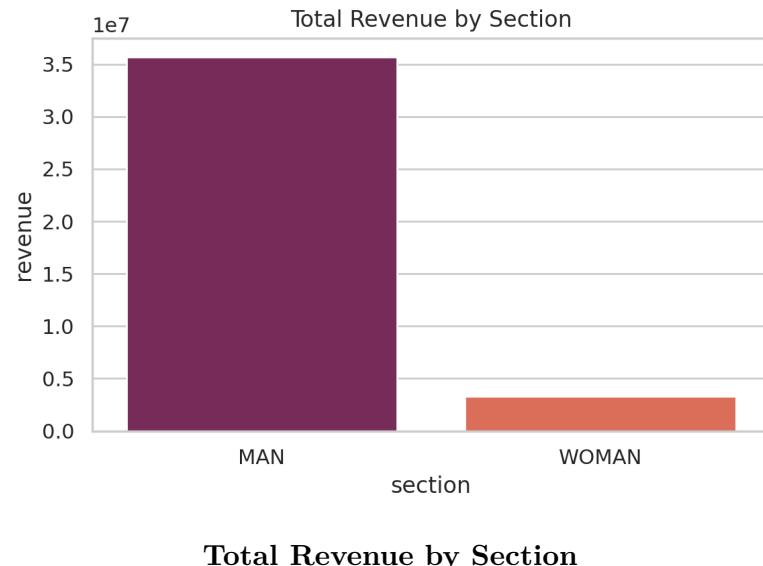
Price vs Sales

Correlation is weak to moderate and it varies from product to product.

7.6 Performance by section

	section	revenue
1	> MAN	35712663.03
2	> WOMAN	3275813.45

- Graph: Bar Chart

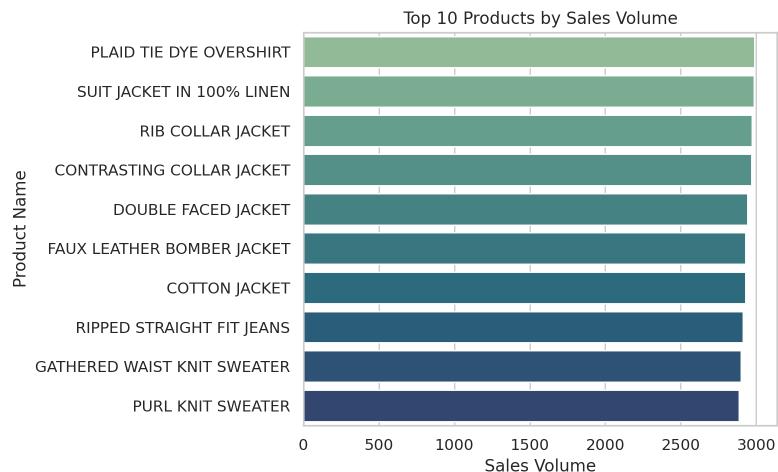


MAN section is the most selling section of Zara, increasing the revenue.

7.7 Top Products (SKUs) by Sales Volume

□	sku	name	price	Sales Volume	revenue
210	317889153-700-3	ABSTRACT JACQUARD SWEATER	59.9	529	31687.1
163	287414911-803-2	HIGH COLLAR KNIT SWEATER	7.99	542	4330.58
188	319473750-809-1	V-NECK KNIT SWEATER	47.9	558	26728.2
193	328207551-500-2	KNIT OPEN BACK PEARLY SWEATER	49.9	606	30239.399999999998
181	313890688-330-1	KNIT SWEATER WITH RIPS	49.9	622	31037.8
203	312393140-406-2	PLEATED WIDE FIT JEANS	49.9	624	31137.6
102	323212374-510-2	MIXED COLLAR JACKET	109.0	628	68452.0
94	312596416-800-2	HOODED TECHNICAL JACKET	109.0	647	70523.0
48	312363708-800-2	TEXTURED POCKET JACKET	89.9	647	58165.3
2	324052738-800-46	TUXEDO JACKET	169.0	654	110526.0
10	312372602-800-2	100% FEATHER FILL PUFFER JACKET	169.0	656	110864.0
46	311292194-731-2	FAUX SUEDE JACKET	89.9	665	59783.50000000001
20	322972473-052-46	SUIT JACKET IN 100% LINEN	159.0	669	106371.0
245	281593208-015-2	PATCH BOMBER JACKET	139.0	694	96466.0
217	317230892-251-3	RIBBED KNIT T-SHIRT	49.9	702	35029.799999999996

- Graph: Bar Chart



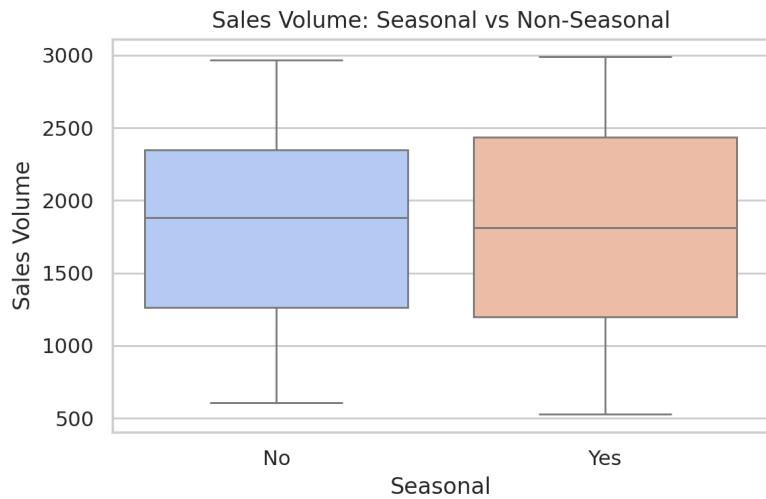
Top 10 products by sales volume

Knit/jacket styles are leading

7.8 Seasonal vs Non-seasonal Performance

□	Seasonal	Sales Volume
1	No	1882.5
2	Yes	1813.5

- Graph: Box Plot



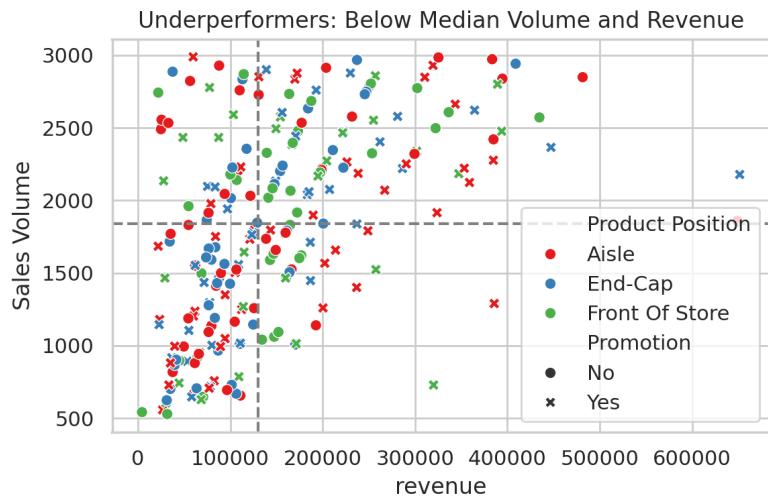
Sales Volume: Seasonal vs Non-seasonal

Non-seasonal is showing a little higher median.

7.9 Under performers

□	sku	name	price	Sales Volume	revenue
210	317889153-700-3	ABSTRACT JACQUARD SWEATER	59.9	529	31687.1
163	287414911-803-2	HIGH COLLAR KNIT SWEATER	7.99	542	4330.58
188	319473750-809-1	V-NECK KNIT SWEATER	47.9	558	26728.2
193	328207551-500-2	KNIT OPEN BACK PEARLY SWEATER	49.9	606	30239.399999999998
181	313890688-330-1	KNIT SWEATER WITH RIPS	49.9	622	31037.8
203	312393140-406-2	PLEATED WIDE FIT JEANS	49.9	624	31137.6
102	323212374-510-2	MIXED COLLAR JACKET	109.0	628	68452.0
94	312596416-800-2	HOODED TECHNICAL JACKET	109.0	647	70523.0
48	312363708-800-2	TEXTURED POCKET JACKET	89.9	647	58165.3
2	324052738-800-46	TUXEDO JACKET	169.0	654	110526.0
10	312372602-800-2	100% FEATHER FILL PUFFER JACKET	169.0	656	110864.0
46	311292194-731-2	FAUX SUEDE JACKET	89.9	665	59783.5000000001
20	322972473-052-46	SUIT JACKET IN 100% LINEN	159.0	669	106371.0
245	281593208-015-2	PATCH BOMBER JACKET	139.0	694	96466.0
217	317230892-251-3	RIBBED KNIT T-SHIRT	49.9	702	35029.79999999996

- Graph: Scatter Plot



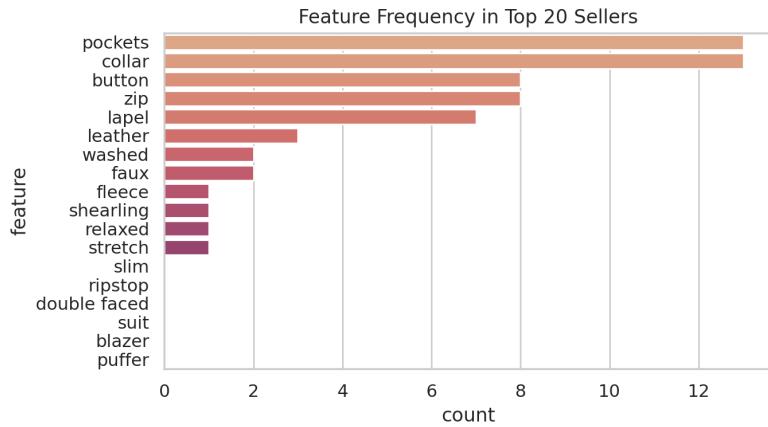
Under performers: Sales Volume vs Revenue

Above shown are the likely under performing candidates.

7.10 Product features among the top selling candidates

□	feature	count
10	pockets	13
11	collar	13
8	button	8
7	zip	8
9	lapel	7
4	leather	3
17	washed	2
18	faux	2
6	fleece	1
5	shearling	1
16	relaxed	1
3	stretch	1
2	slim	0
12	ripstop	0
13	double faced	0

- Graph: Bar Chart



Frequency of feature in top 20 selling candidates

Pockets, collars, zip, button appeared most frequently in the top selling products.

8 Data Collection & preprocessing

8.1 Data Source

Data was collected from Kaggle. Dataset contains 253 rows and 16 columns.

8.2 Dataset fields - Raw

- Product ID
- Product Position
- Promotion
- Product Category
- Seasonal
- Sales Volume
- Name
- Description
- Price
- Terms
- Section

8.3 Data Preprocessing

Data Cleaning

- Handled missing values.
- Standardized all the text fields.
- Made calculated fields.

Feature Engineering

- Computed Revenue = Sales * Price
- Created price bins (0-25, 25-50, 50-70, etc)
- Created seasonal flag, and promo flag
- Extracted month from date.

Aggregation

- Created SKU-level summary table
- Computed top selling products and under performing SKUs.

Tokenization

- Extracted certain words from the description of products.
- Counted the number of top selling products.

9 Results and Discussion

Dominance of Category

- Unit sales are dominated by clothing. Hence, its inventory requires utmost primary focus.

Placement of products

- Products placed in the front of the store showed the highest sales. This shows that the location of the product in the store matters in customer buying decisions.

Sales increased by promotion

- The products that are promoted showed increase in the sales of that product. This means, promotion uplifted the sales.

Price bands

- Most customers purchased the mid range priced products. Mid range band contributed to the highest revenue.

Seasonal Products

- Seasonal products increased the sales revenue. However, it also varies by category.

Top 10 SKUs

- Top 10 SKUs contributed to the largest share of the sales by revenue.

Relationship between price and Sales

- A weak negative correlation of -0.07 was seen between price and sales.

Identification of Under performers

- Quadrant chart of under performers showed that they exhibited low sales and low revenue (bottom-left quadrant)

Specific product features

- Top selling products has some common features in them like pockets, zip, button,

collar, etc.

10 Key findings and Recommendations

10.1 Key findings

- Zara's most sales were driven by clothing.
- Products placed in the front of the store were sold higher than others.
- Products that are promoted were sold higher than those not promoted.
- Most customers bought products in the mid range price category.
- Seasonality effect on products are prevalent but are limited.
- Top 10 SKUs contributed to the maximum sales hence they must be kept in the priority in inventory stock.
- Existence of weak negative correlation between price and sales.
- Under performing products must be considered for repositioning in the store.
- All top selling products showed some common product features like zip. collar, pockets, etc.

10.2 Recommendations

10.2.1 *Operational Recommendations for Zara*

- Increase the inventory of top selling products.
- Under performing products must be repositioned in the store to increase their sales.
- Set up discounts for other high price range products to increase their sales.
- A/B testing is recommended before going to promotions.

10.2.2 *Analytics Recommendations*

- Building weekly dashboards helps in analyzing data more efficiently.
- Use ML to do the model prediction on price elasticity
- Semantic clustering can be used to expand NLP analysis.

11 AI Contribution

11.1 Generative AI

11.1.1 *Julius*

- Julius was used for the initial data exploration, and descriptive analytics.
- Then, it was used for data visualization of the trends.
- It helped me create the visualization graphs of the descriptive analytics.

11.1.2 Chat GPT

- Julius results were compared with Chat GPT descriptive analytics.
- Unfortunately, few results were aligned with the Julius and few did not.
- Chat GPT was used to create a report on the analysis of Julius.
- It also helped in generating Python codes to verify the output of the Julius descriptive statistics.

11.2 Programming with AI libraries

11.2.1 Python

- Python was used to clean the data and perform the data preprocessing.
- Descriptive analytics were verified using Python.

11.3 Data Visualization Tools

Power BI and Tableau were used to create an interactive dashboard and story telling respectively.



Dashboard

12 Link for the documents

Google Drive

https://drive.google.com/drive/folders/1LNu4RwUbK_PW_gdJNgQbWQUW-nZBWVpf?usp=drive_link