Introduction About Contact

ZARA Sales Analysis

End-to-End Sales Data Analysis & Business Insights

Tools: Python (EDA), SQL (Queries & Views), Power BI (Dashboard)

01

02

Introduction

Zara is a global fast-fashion giant known for quickly bringing runway trends to stores, offering trendy, affordable clothing for women, men, and kids, with a focus on speed, flexibility, and customer-driven designs, often featuring limited runs that encourage quick purchases.

01

02

03

Objective

- This analysis aims to identify what sells, why it sells, and where decisions can be optimized.
- The focus is on transforming raw sales data into actionable business insights.
- Analyzed ZARA sales data to understand product performance, pricing, and demand trends
- Used Python for data cleaning, SQL for querying, and Power BI for visualization
- Focused on answering business-driven questions

Why This Analysis Was Needed

To Know the following:

1. Which product categories generate the highest sales volume, and what is their contribution percentage to overall sales?
2. How do promotions impact sales volume—what uplift do we see in units sold for promoted vs. non-promoted products?
3. Which product positions drive the highest sales, and how does visibility affect performance?
4. Is there a relationship between product price and sales volume—do lower-priced or premium products sell better across categories?
5. Which categories rely the most on promotions to drive sales, and where can the company reduce discount dependency?

02

03

04

Analytical Approach

Python was used to explore trends and validate data quality.

SQL was applied to perform structured, business-focused queries.

Power BI was used to visualize insights and enable interactive exploration.

Data Understanding & Preparation

```
from google.colab import files
uploaded = files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving ZARA sales COPY export 2025-12-02 09-05-07.csv to ZARA sales COPY export 2025-12-02 09-05-07.csv
```

```
import pandas as pd
df = pd.read_csv("ZARA sales COPY export 2025-12-02 09-05-07.csv")
print(df)
```

```
248      199233           Aisle       Yes        Clothing       No
249      137044           Aisle        No        Clothing      Yes
250      154736   Front of Store       Yes        Clothing      Yes
251      141434   Front of Store        No        Clothing      Yes

     Sales Volume brand                                        url  \
0            2823  Zara  https://www.zara.com/us/en/basic-puffer-jacket...
1             654  Zara  https://www.zara.com/us/en/tuxedo-jacket-p0889...
2            2220  Zara  https://www.zara.com/us/en/slim-fit-suit-jacke...
3            1568  Zara  https://www.zara.com/us/en/stretch-suit-jacket...
4            2942  Zara  https://www.zara.com/us/en/double-faced-jacket...
..            ...   ...                                               ...
247          1014  Zara  https://www.zara.com/us/en/basic-100-wool-swea...
248          2222  Zara  https://www.zara.com/us/en/colorblock-knit-cro...
249          2534  Zara  https://www.zara.com/us/en/hooded-technical-ja...
250          1466  Zara  https://www.zara.com/us/en/houndstooth-suit-ja...
251          2870  Zara  https://www.zara.com/us/en/bomber-jacket-p0534...

                sku                                        name  \
```

1. Examined the structure, data types, and completeness of the sales dataset using Google Collab.
2. Cleaned inconsistent values and validated key fields such as price, category, and sales volume.
3. Ensured the dataset was reliable before further analysis.
4. Removed all the null values using replaces and fillna
5. Copied this clean data into a csv file for further analysis in SQL.

```
[252 rows x 16 columns]
```

```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 252 entries, 0 to 251
Data columns (total 16 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Product ID        252 non-null    int64
 1   Product Position  252 non-null    object
 2   Promotion         252 non-null    object
 3   Product Category  252 non-null    object
 4   Seasonal          252 non-null    object
 5   Sales Volume      252 non-null    int64
 6   brand             252 non-null    object
 7   url               252 non-null    object
 8   sku               252 non-null    object
 9   name              251 non-null    object
 10  description       250 non-null    object
 11  price             252 non-null    float64
 12  currency          252 non-null    object
 13  scraped_at        252 non-null    object
 14  terms             252 non-null    object
 15  section           252 non-null    object
dtypes: float64(1), int64(2), object(13)
memory usage: 31.6+ KB
None
```

```
print(df.isnull().sum())
```

```
Product ID          0
Product Position    0
Promotion           0
Product Category    0
Seasonal            0
Sales Volume        0
brand               0
url                 0
sku                 0
name                1
description         2
price               0
currency            0
scraped_at          0
terms               0
section             0
dtype: int64
```

```
print(df.columns.tolist())
```

```
['Product ID', 'Product Position', 'Promotion', 'Product Category', 'Seasonal', 'Sales Volume', 'brand', 'url', 'sku', 'name', 'description', 'price', 'currency', 'scraped_at
```

```
cols_to_drop = [
    "description", "url", "scraped_at"]
df = df.drop(columns=cols_to_drop, errors="ignore")
```

```
df.columns = df.columns.str.lower().str.replace(' ', '_')
print(df.columns.tolist())
```

```
['product_id', 'product_position', 'promotion', 'product_category', 'seasonal', 'sales_volume', 'brand', 'sku', 'name', 'price', 'currency', 'terms', 'section']
```

```
print(df[df['name'].isna()])
```

```
    product_id product_position promotion product_category seasonal  \
72      173576          End-cap       Yes         Clothing       No

    sales_volume brand             sku name  price currency    terms section
72          1838  Zara  336378923-700-2  NaN  129.0      USD  jackets     MAN
```

```
df['name'] = df['name'].fillna('unknown')
```

```
print(df.isnull().sum())
```

```
product_id          0
product_position    0
promotion           0
product_category    0
seasonal            0
sales_volume        0
brand               0
sku                 0
name                0
price               0
currency            0
terms               0
```

```
import sqlite3
import pandas as pd

df.to_csv("zara_analysis.csv", index=False)

from google.colab import files
files.download("zara_analysis.csv")
```

04

05

06

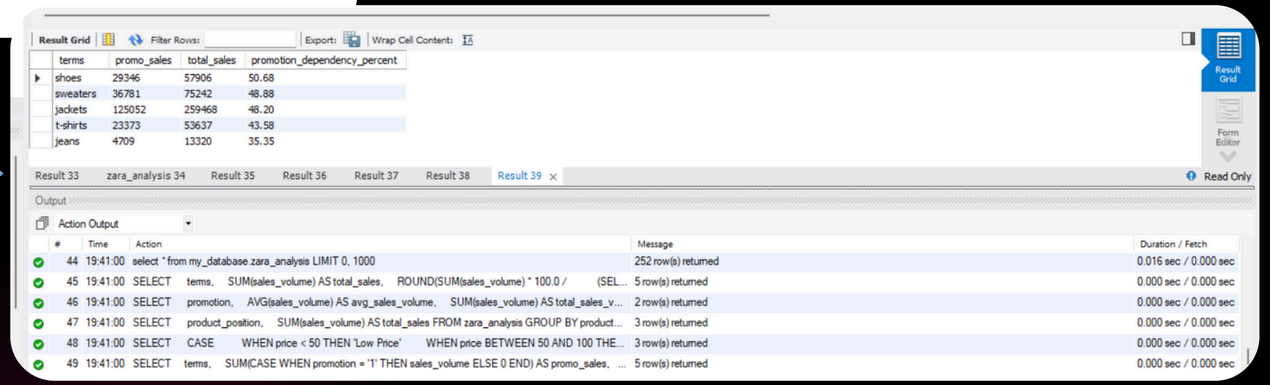
4. Removed all the null values using replaces and fillna.
5. Copied this clean data into a csv file for further analysis in SQL.

ZARA

Sales Patterns & Product Performance

```sql
SHOW DATABASES;
USE my_database;
select * from my_database.zara_analysis;
-- Which product categories generate the highest sales volume, and what is their contribution percentage to overall sales?
SELECT
    terms,
    SUM(sales_volume) AS total_sales,
    ROUND(SUM(sales_volume) * 100.0 /
        (SELECT SUM(sales_volume) FROM zara_analysis), 2) AS contribution_percent
FROM zara_analysis
GROUP BY terms
ORDER BY total_sales DESC;
```

- Imported the data into a structured dataset to efficiently address business queries.
- Used SQL functions such as SUM, GROUP BY, and CASE statements to derive meaningful insights.
- Created SQL views to simplify reuse of results and support consistent reporting.



terms	promo_sales	total_sales	promotion_dependency_percent
shoes	29346	57906	50.68
sweaters	36781	75242	48.88
jackets	125052	259468	48.20
t-shirts	23373	53637	43.58
jeans	4709	13320	35.35

Result 33 | zara_analysis 34 | Result 35 | Result 36 | Result 37 | Result 38 | Result 39 | Read Only

Output

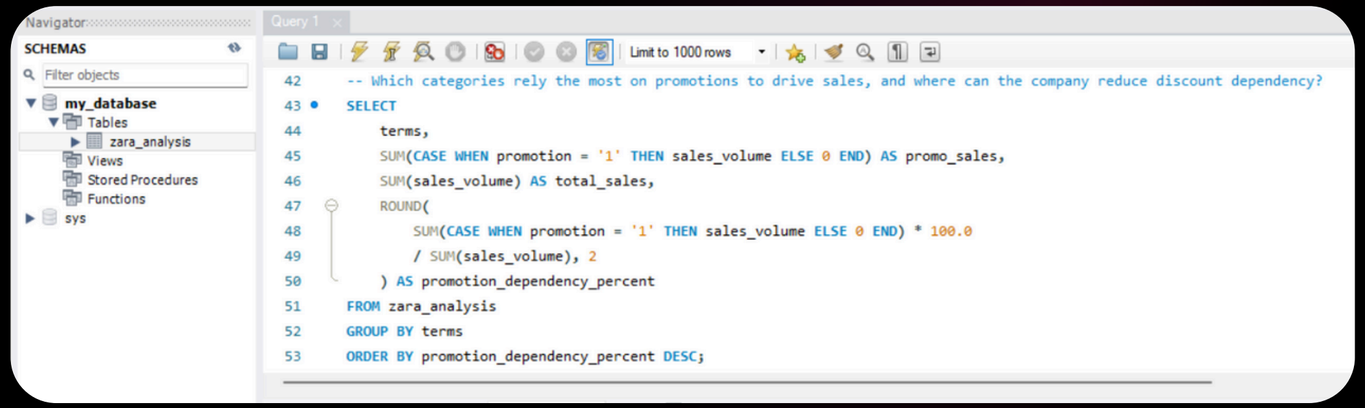
Action Output

#	Time	Action	Message	Duration / Fetch
44	19:41:00	select * from my_database.zara_analysis LIMIT 0, 1000	252 row(s) returned	0.016 sec / 0.000 sec
45	19:41:00	SELECT terms, SUM(sales_volume) AS total_sales, ROUND(SUM(sales_volume) * 100.0 / (SEL...	5 row(s) returned	0.000 sec / 0.000 sec
46	19:41:00	SELECT promotion, AVG(sales_volume) AS avg_sales_volume, SUM(sales_volume) AS total_sales_v...	2 row(s) returned	0.000 sec / 0.000 sec
47	19:41:00	SELECT product_position, SUM(sales_volume) AS total_sales FROM zara_analysis GROUP BY product...	3 row(s) returned	0.000 sec / 0.000 sec
48	19:41:00	SELECT CASE WHEN price < 50 THEN 'Low Price' WHEN price BETWEEN 50 AND 100 THE...	3 row(s) returned	0.000 sec / 0.000 sec
49	19:41:00	SELECT terms, SUM(CASE WHEN promotion = '1' THEN sales_volume ELSE 0 END) AS promo_sales, ...	5 row(s) returned	0.000 sec / 0.000 sec

05

06

07



Navigator

SCHEMAS

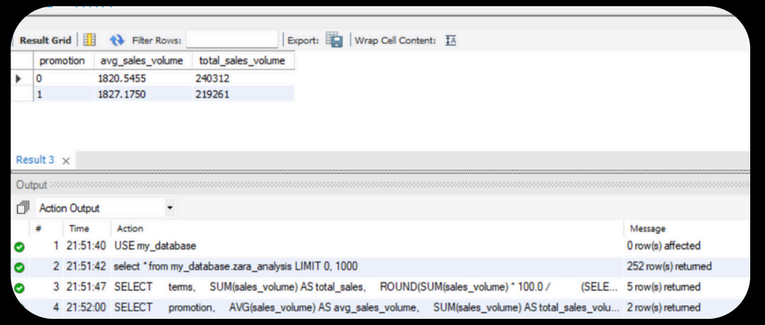
Filter objects

- my_database
  - Tables
    - zara_analysis
  - Views
  - Stored Procedures
  - Functions
- sys

Query 1

Limit to 1000 rows

```sql
-- Which categories rely the most on promotions to drive sales, and where can the company reduce discount dependency?
SELECT
    terms,
    SUM(CASE WHEN promotion = '1' THEN sales_volume ELSE 0 END) AS promo_sales,
    SUM(sales_volume) AS total_sales,
    ROUND(
        SUM(CASE WHEN promotion = '1' THEN sales_volume ELSE 0 END) * 100.0
        / SUM(sales_volume), 2
    ) AS promotion_dependency_percent
FROM zara_analysis
GROUP BY terms
ORDER BY promotion_dependency_percent DESC;
```



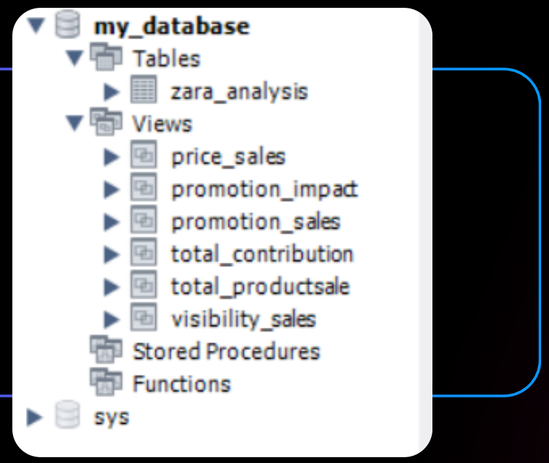
promotion	avg_sales_volume	total_sales_volume
0	1820.5455	240312
1	1827.1750	219261

Result 3

Output

Action Output

#	Time	Action	Message
1	21:51:40	USE my_database	0 row(s) affected
2	21:51:42	select * from my_database.zara_analysis LIMIT 0, 1000	252 row(s) returned
3	21:51:47	SELECT terms, SUM(sales_volume) AS total_sales, ROUND(SUM(sales_volume) * 100.0 / (SELE...	5 row(s) returned
4	21:52:00	SELECT promotion, AVG(sales_volume) AS avg_sales_volume, SUM(sales_volume) AS total_sales_volu...	2 row(s) returned



What SQL provided:

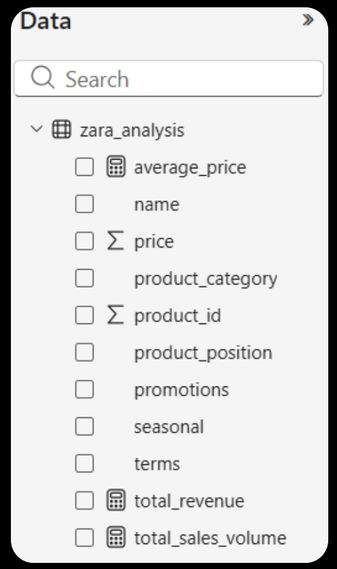
- Sales performance varies significantly across product categories.
- Certain products contribute disproportionately to overall sales.
- Product positioning and visibility play a key role in driving demand.

Impact of Business Variables on Sales

- Pricing directly influences sales volume, but the effect differs by category.
- Promotions increase sales in some segments while showing limited impact in others.
- Seasonal patterns highlight demand shifts over time

ZARA

Steps taken in powerbi



- Connected SQL query results to Power BI for reporting.
- Verified column data types and ensured numeric fields were analysis-ready.
- Created calculated measures for consistent sales reporting.
- Ensured filters and slicers work correctly across visuals.

Visualizing Insights for Decision-Makers

- Key sales metrics and trends are presented in a clear, interactive format.
- Stakeholders can quickly identify high-performing products and categories.
- Enables faster and more informed business decisions.

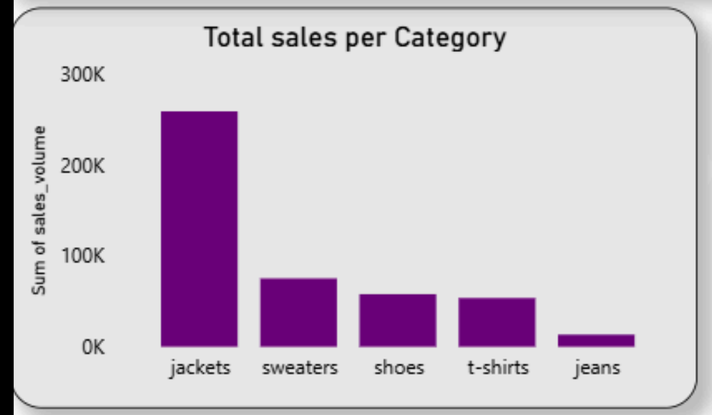
07

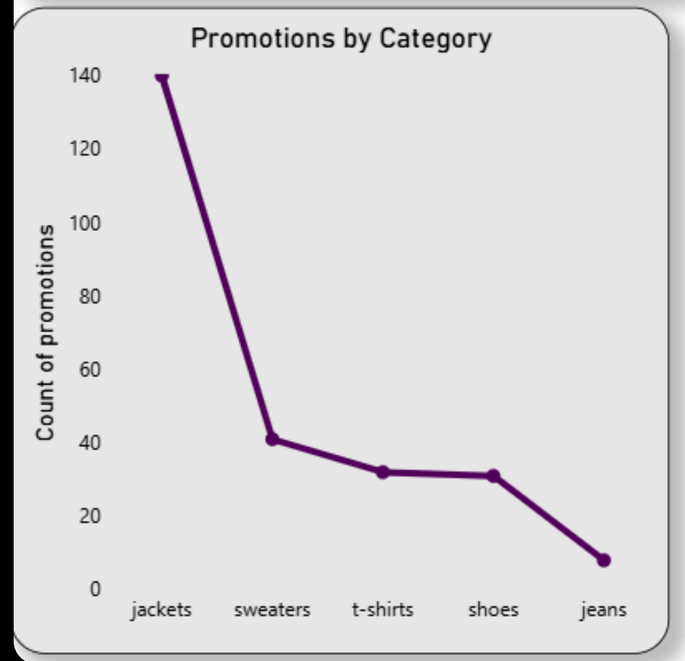
08

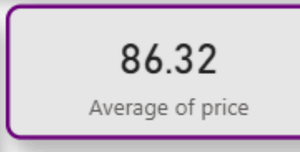
09

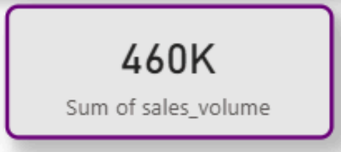
Data Modeling & Visualization

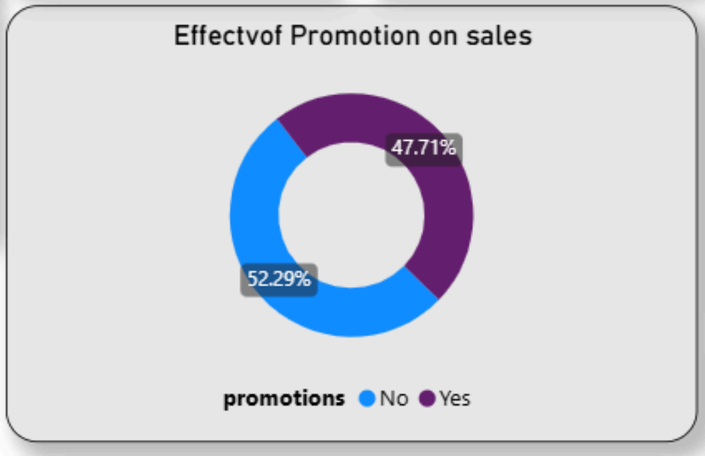
Zara Sales Analysis Dashboard

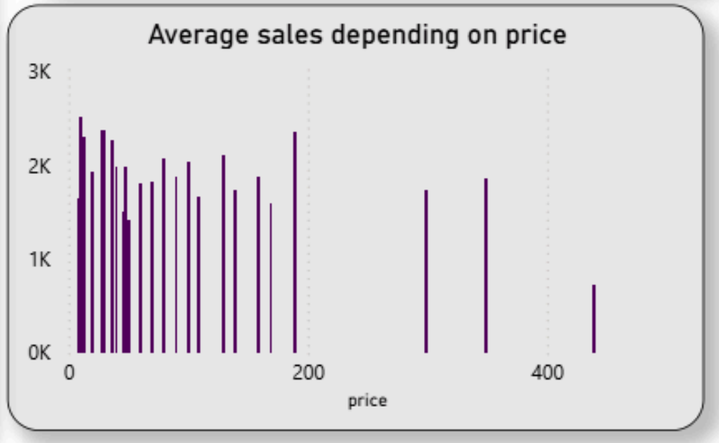


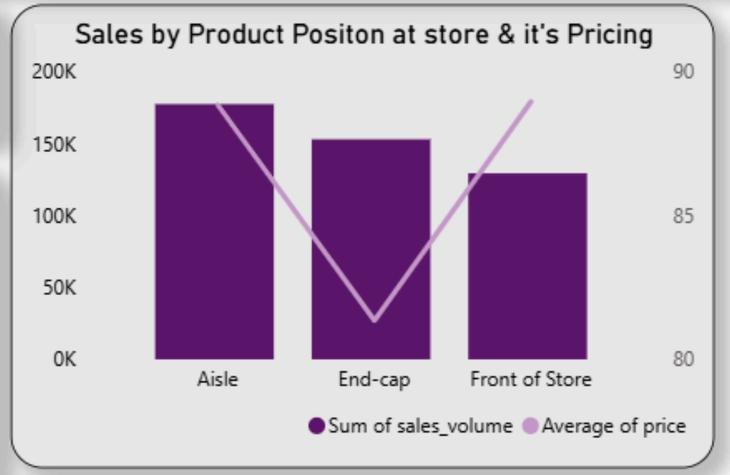


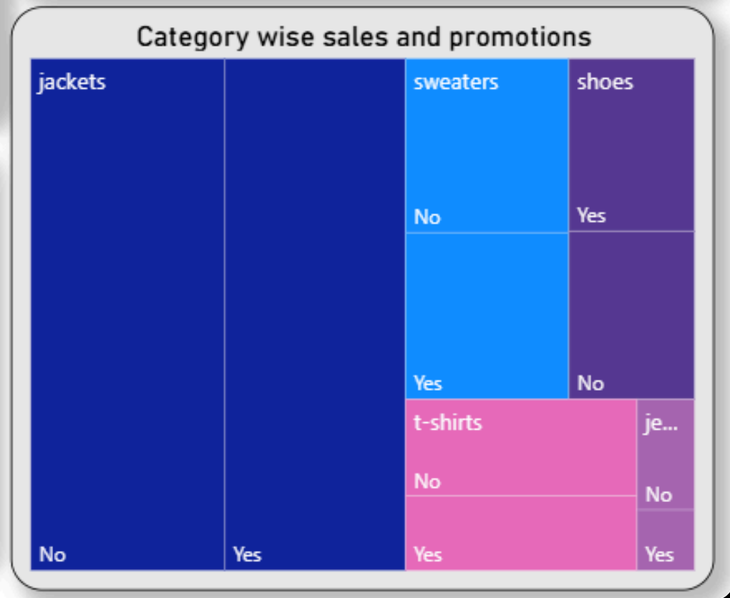












08

09

10

Key Insights

What the Data Reveals

- **Jackets contribute 56% of overall sales volume**, making them the top-performing category and a key revenue driver for the business.
- **Non-promoted items outperform** promoted products, indicating that a significant portion of **sales is driven organically rather than through discounts.**
- **Aisle product** positioning **generates the highest sales**, suggesting that greater **visibility and accessibility** strongly **influence customer** purchase behavior.
- **Price and sales volume show an inverse relationship**, where lower-priced products consistently achieve higher sales across categories.
- The **Shoes** category shows **higher dependency on promotions to drive sales**, highlighting an opportunity for the company to re-evaluate pricing and reduce discount reliance.

ZARA

Business Recommendations

- *Prioritize Jackets inventory and visibility, as they drive the majority of sales, while using Sweaters and Jeans to support seasonal demand.*
- *Reduce aggressive discounting, since T-shirts and Jeans perform strongly through organic sales, helping protect profit margins.*
- *Maximize aisle placements for high-value categories, as visibility directly increases purchase likelihood.*
- *Adopt smart pricing strategies, keeping lower-priced items competitive to drive volume while maintaining value-based pricing for premium categories.*
- *Rework promotion strategy for Shoes, focusing on product positioning and pricing to reduce discount dependency.*

ZARA

Thank You

BY KHUSHI SANHOTRA