

# QUANTUM UNIVERSITY



## INTERNSHIP REPORT

**ANEJA SONS PLATINUM STORE  
ROORKEE**

**SUBMITTED BY:**

**KRISHNA RAJOO**

**B. TECH CSE-DS (5<sup>th</sup> Sem)**

**SUBMITTED TO:**

**MR. RAJARSHI BANERJEE**

# DECLARATION

I, KRISHNA RAJOO, a student of Bachelor of Technology in Computer Science and Engineering with specialization in Data Science at Quantum University, Roorkee, hereby declare that the internship report is my original work. I confirm that all data analysis, insights, and recommendations are based on my independent work, conducted during my internship at ANEJA SONS PLATINUM STORE.

**KRISHNA RAJOO**  
**(B. Tech CSE-DS)**  
**(5<sup>th</sup> Sem)**

# INTERNSHIP OFFER LETTER

## Aneja Sons Platinum Store

GST NO-05AANPA76608122

### INTERNSHIP OFFER LETTER

Dear Krishna Rajoo

Congratulations on being selected for the "Data Science" Internship! We Aneja Sons-Platinum Store are thrilled to have you join us. This internship will span 8 weeks from JUNE 2<sup>nd</sup> ,2025 to JULY 27<sup>th</sup> ,2025.

This internship is designed for Business experience, focusing on learning, market analysis, skill development, and gaining practical knowledge. As an intern, we expect you to:

1. Manage complete sales record efficiently.
2. Analyze the sales and profit of different brands in the store.
3. Follow any lawful and reasonable instructions provided by your supervisors.
4. Participate actively in any discussion.
5. Provide regular updates on your progress.
6. Seek feedback and apply it to improve your performance.

We trust that you will approach all tasks with diligence and enthusiasm. We are confident that this internship will be an enriching for you. We look forward to working with you and supporting you in achieving your career aspirations.

Best regards,

Mr. Chandan Aneja

Owner, Aneja Sons – Platinum Store

Contact: 09758320000

  
Aneja Sons Platinum Store  
Civil Lines, Prem Mandir Road,  
Roorkee-247 667

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my mentors and owner of the store for their valuable guidance, encouragement, and support throughout the complete internship. Their insights and expertise greatly contributed to the successful completion of this internship.

I also extend my appreciation to all those who provided data and technical assistance, enabling a practical understanding of customer behaviour through data analytics. This internship has helped me apply theoretical knowledge in a real-world scenario using Python and enhanced my skills in data preprocessing, visualization, and statistical analysis.

# OBJECTIVE

The primary objective of this internship was to gain practical experience and industry exposure in the field of Data Science. Through this internship, I aimed to bridge the gap between theoretical knowledge acquired during academic coursework and real-world applications involving data analysis, predictive modelling, and business problem-solving. The internship provided an opportunity to work on datasets, perform exploratory data analysis, apply statistical techniques, and develop machine learning models to extract meaningful insights.

Another key objective was to enhance my technical proficiency in tools and technologies commonly used in the data science domain, such as Python, Pandas, NumPy, Power BI and data visualization libraries. Throughout the internship, I focused on improving my skills in data preprocessing, handling missing values, feature engineering, and building accurate predictive models. The internship also emphasized understanding the importance of data-driven decision-making in solving business challenges.

Additionally, this internship aimed to develop my ability to work in a professional environment, collaborate within teams, and communicate analytical findings effectively. It provided exposure to real-world datasets, problem statements, and end-to-end project execution, thus preparing me for future roles in the data science and analytics industry.

# EXECUTIVE SUMMARY

This report provides a comprehensive overview of my internship experience at **ANEJA SONS PLATINUM STORE**, where I undertook a **Summer Internship of 8 weeks** from **June 2, 2025 to July 27, 2025**. The primary objective of the internship was to apply theoretical knowledge in a practical business environment, enhance my technical proficiency in data analysis tools, and contribute to ANEJA SONS Platinum Store data-driven decision-making processes.

Throughout the internship, I engaged in various activities such as **data visualization and reporting, data mining and extraction, statistical modeling, and SQL-based database management**. Using tools like **Excel, Power BI**, and I created interactive dashboards to visualize sales trends and operational efficiencies.

A significant portion of my work focused on ensuring **data quality and accuracy**, which is critical for reliable analysis and decision-making. I implemented data validation processes and performed quality checks to maintain consistency across datasets. Managing databases allowed me to optimize data retrieval processes and support the team in handling large datasets effectively.

Throughout the internship, I developed both **technical and soft skills**. My proficiency in data analysis tools improved significantly, and I gained experience in applying **statistical models** to real-world problems. The internship also enhanced my **problem-solving abilities, communication skills, and teamwork**. Collaborating with various departments provided insights into the importance of clear communication and cross-functional collaboration in achieving organizational goals.

This report is structured to reflect on these experiences, detailing the responsibilities undertaken, challenges faced, and solutions implemented. It also highlights the value I added to Aneja Sons Platinum Store through my contributions and the key takeaways from this transformative experience. Overall, the internship has been instrumental in shaping my career aspirations, providing a solid foundation for future endeavors in data analysis and beyond.

# TABLE OF CONTENT

<b>S.NO.</b>	<b>TOPICS</b>	<b>PAGE NO.</b>
<b>1.</b>	Introduction	<b>1</b>
<b>2.</b>	Key Resources Used	<b>2</b>
<b>3.</b>	Methodology	<b>3</b>
<b>4.</b>	Project Given	<b>4</b>
<b>5.</b>	Project I	<b>5-9</b>
<b>6.</b>	Project II	<b>10-13</b>
<b>7.</b>	Project III	<b>14-17</b>
<b>8.</b>	Project IV (Overview)	<b>18-20</b>
<b>9.</b>	Contribution to Organization	<b>21</b>
<b>10.</b>	Conclusion	<b>22</b>
<b>11.</b>	Key Benefits of Internship	<b>23</b>
<b>12.</b>	Internship Certificate	<b>24</b>
<b>13.</b>	Personal Growth & Challenges	<b>25</b>
<b>14.</b>	Recommendation for Future Intern	<b>26</b>

# INTRODUCTION

The internship was undertaken at **Aneja Sons Platinum Store**, a reputed retail establishment specializing in readymade garments for ladies, gents, girls, boys, and kids. The store offers a wide range of products from multiple well-known national and international brands, including **Madame, Levis, Monte Carlo, Pepe Jeans, Octave, Libas, Indian Terrain, ONLY, Camla**, and many more. With its diverse product portfolio and customer base, Aneja Sons Platinum Store has established itself as a trusted name in the apparel industry in the region.

During my two-month internship, I was assigned to the Data Analysis department, focusing specifically on evaluating the performance of various brands under the **Gents** and **Girls** wear categories. The primary aim was to analyse sales and profit data to identify the top-performing brands that contribute significantly to revenue generation and profitability. This analysis was crucial for store management to make informed decisions regarding inventory planning, brand promotion, and future procurement strategies.

Through this internship, I gained practical exposure to real-world data handling, sales performance analysis, and the application of data science techniques in a retail business environment. The project not only enhanced my analytical and technical skills but also deepened my understanding of how data-driven insights can support business growth and operational efficiency in the competitive retail sector.



# KEY RESOURCES USED

## 1. Python 3.x

Python is a powerful, high-level programming language widely used for data analysis, machine learning, and automation. Its clear syntax and extensive support for data manipulation make it ideal for working with structured datasets. In this project, Python served as the core language to implement all data processing, statistical analysis, and visualization tasks.

## 2. Jupyter Notebook

Jupyter Notebook is an open-source web-based tool that allows interactive development of Python code. It is especially useful for data science projects because it lets users write and execute code, view data outputs, and add explanations or visuals — all in a single document. It was used throughout this project to explore the dataset, write and test code, visualize results, and document the analysis.

## 3. Pandas

Pandas is a Python library essential for data manipulation and analysis. It provides data structures like DataFrames that allow easy handling of tabular data. In this project, Pandas was used to clean the dataset, handle missing values (if any), group data by categories (like region or age), and perform statistical aggregations such as mean, sum, and count.

## 4. Matplotlib

Matplotlib is a popular Python library for creating static, animated, and interactive visualizations. It was used in this project to generate bar charts, line plots, and horizontal bar charts. These visualizations played a critical role in interpreting insights and making the results of data analysis more accessible and intuitive.

## 5. Seaborn

It provides high-level functions to create attractive and informative statistical graphics with minimal code. Seaborn is widely used for visualizing distributions, relationships, and trends within datasets in the field of data science.

## 6. Power BI

Power BI is a visualization tool which is used for creating a dashboard for analysis. Here, we have used Power BI to draw a conclusion of this project so that the complete analysis can be shown on a single page.

# METHODOLOGY

The methodology of this project was structured into several systematic steps to ensure a comprehensive analysis of customer purchase behaviour:

1. **Data Collection and Loading**

The dataset was provided in CSV format and loaded into the Python environment using Pandas. This dataset included fields such as age, gender, income, product category, region, promotion usage, and satisfaction score.

2. **Data Cleaning and Preprocessing**

The dataset was examined for missing values, duplicates, and incorrect entries. Since the data was clean and consistent, it was directly used for analysis. Columns were appropriately categorized for analysis, and new age group categories were created using binning techniques.

3. **Exploratory Data Analysis (EDA)**

Using Pandas and Matplotlib, various groupings and summaries were calculated. Insights were derived about customer preferences based on age, product category, region, and promotion usage. Bar plots and line charts were used to visualize these patterns clearly.

4. **Statistical Analysis**

The scipy. stats module was used to conduct statistical analysis (e.g., correlation) to validate assumptions like the relationship between income and satisfaction or purchase amount. This helped ensure that the trends were not just visual but statistically supported.

5. **Insight Generation and Visualization**

Key insights, such as the most preferred product category and the highest-spending age group, were visualized using horizontal bar plots, grouped bar charts, and line plots. These visual tools made complex data more understandable.

# PROJECT GIVEN

## **Project I:**

- Adidas Sales Analysis
- This project enables the owner to borrow particular apparel for retailing purpose which is required most in the market

## **Project II:**

- Analysis of production and sales of **Zara**
- This project enables the owner to borrow particular apparel for retailing purpose which is most preferred by Customer

## **Project III:**

- Analysis of Top Trending Brand in the market
- This project enables the owner to get franchise of top brand which is most preferred by Customer

## **Project IV:**

- Analysis of Brand gives more profit and most preferred by Customer
- This project helps the owner to analyse their most preferred and profitable brand so that they can give priority to that Brand first

# PROJECT I

## ADIDAS SALES ANALYSIS

### Introduction:

The Adidas sales Analysis project focuses on exploring, cleaning, and analysing product-level data to derive meaningful insights into pricing patterns, discount strategies, and product distribution. The dataset contains key information such as original prices, discounted prices, product categories, and other attributes relevant to understanding Adidas' product offerings and pricing behaviour. The primary objective is to handle missing data effectively, visualize critical trends, and assist in data-driven decision-making for pricing strategies.

### Python Code I:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import seaborn as sns
```

Explanation: These are the libraries imported in the Jupyter Notebook for reading csv file, performing numerical and scientific calculation, visualization with different techniques and so on.

### Python Code II:

```
df=df.drop('url',axis=1)
df=df.drop('images',axis=1)
df=df.drop('description',axis=1)
df=df.drop('brand',axis=1)
df=df.drop('country',axis=1)
df=df.drop('language',axis=1)
df
```

Explanation: By using drop attribute, we have removed unwanted columns from the dataset which have same entry in all the tuples.

### **Python Code III:**

```
missing_values = df.isnull().sum()
print("Missing values per column:\n", missing_values)
print("\nData Types:\n", df.dtypes)
```

Explanation: By using isnull(), we have checked out the null values in the dataset which can affect the analysis and result.

### **Python Code IV:**

```
print("\nUnique values in 'color':", df['color'].unique())
print("\nUnique values in 'category':", df['category'].unique())
print("\nUnique values in 'breadcrumbs':", df['breadcrumbs'].unique())
```

Explanation: By using unique(), we have find out the unique values in the defined columns which can helps us for further analysis.

### **Python Code V:**

```
plt.figure(figsize=(8,5))
sns.histplot(df['original_price'],kde=True,color='darkblue')
plt.title('Distribution Of Original Price')
plt.xlabel('Original Price')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```

Explanation: By using histogram plot visualization, we have find out the distribution type of the given column so that we can select whether the null values should be replace by mean value or median value.

### **Python Code VI:**

```
plt.figure(figsize=(8,6))
sns.scatterplot(data=df,x='original_price',y='selling_price',alpha=0.6)
plt.title("Comparison of prices")
plt.show()
```

### **Python Code VII:**

```
plt.figure(figsize=(10,5))  
  
sns.countplot(data=df,x='category',order=df['category'].value_counts().index)  
  
plt.title("Product Count by Category")  
  
plt.xticks(rotation=45)  
  
plt.show()
```

Explanation: This visualization helps to checkout that which Category product is mostly preferred.

### **Python Code VIII:**

```
Top4=df.nlargest(4,'original_price')  
  
plt.figure(figsize=(10,5))  
  
sns.barplot(data=top10,x='name',y='original_price')  
  
plt.title('Top 4 Expensive Product')  
  
plt.xticks(rotation=45)  
  
plt.show()
```

Explanation: This visualization helps us to findout the top 4 most preferred product

### **Python Code IX:**

```
plt.figure(figsize=(8,6))  
  
sns.heatmap(df.corr(numeric_only=True),annot=True,cmap='coolwarm')  
  
plt.title('Correlation Heatmap')  
  
plt.show()
```

Explanation: This visualization helps us to create relation between the numerical columns in the given dataset.

### **Python Code X:**

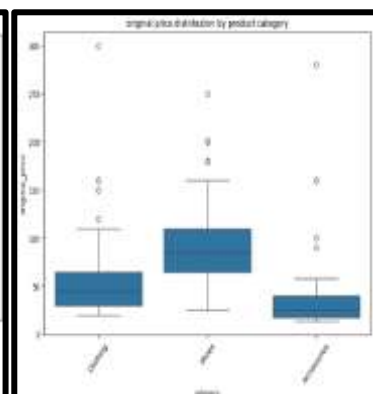
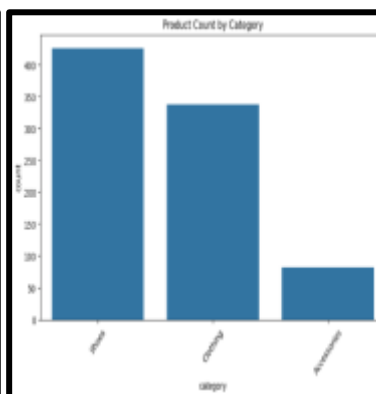
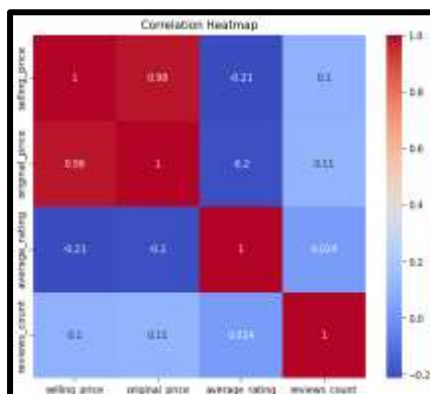
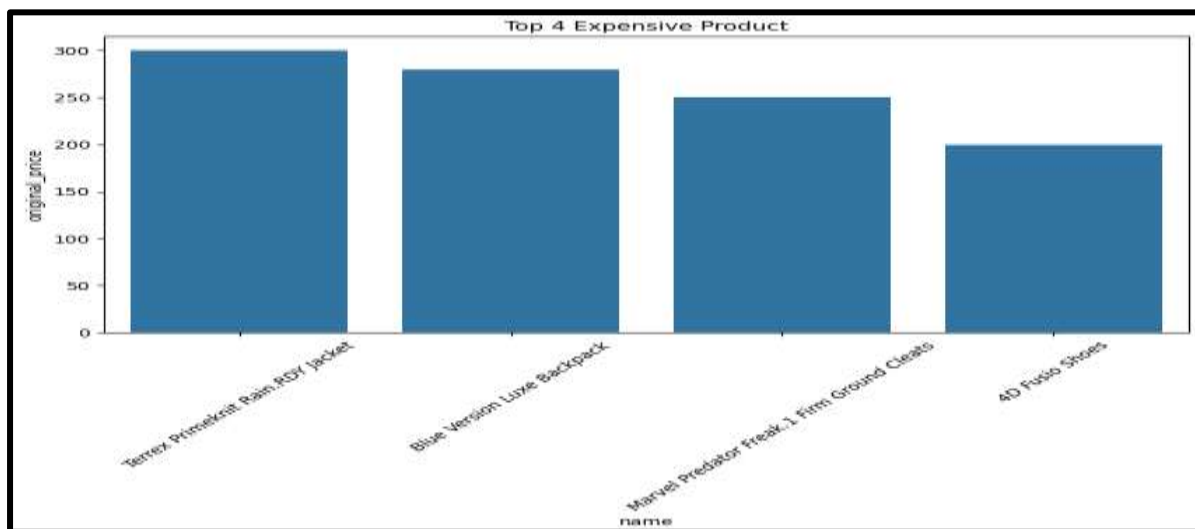
```
sns.boxplot(data=df,x='category',y='original_price')  
  
plt.title('original price distribution by product category')  
  
plt.xticks(rotation=45)  
  
plt.show()
```

## Outcome of all the codes:

```
Unique values in 'color': ['Black' 'Grey' 'White' 'Blue' 'Purple' 'Pink' 'Green' 'Yellow' 'Red'
'Multicolor' 'Gold' 'Burgundy' 'Beige' 'Orange' 'Multi' 'Turquoise'
'Silver' 'Brown']

Unique values in 'category': ['Clothing' 'Shoes' 'Accessories']

Unique values in 'breadcrumbs': ['Women/Clothing' 'Women/Shoes' 'Kids/Clothing' 'Five Ten/Shoes'
'Men/Clothing' 'Men/Accessories' 'Kids/Shoes' 'Soccer/Shoes' 'Men/Shoes'
'Essentials/Shoes' 'Originals/Shoes' 'Women/Accessories'
'Training/Accessories' 'Originals/Accessories' 'Swin/Shoes'
'Soccer/Accessories' 'Running/Shoes' 'Originals/Clothing'
'Essentials/Clothing' 'Sportswear/Clothing' 'Running/Accessories'
'Kids/Accessories']
```



## Outcome Using Power BI:



## Conclusion:

The Adidas Sales Analysis project provided valuable insights into the pricing structure, discount trends, and category-wise product distribution within the dataset. Through thorough data cleaning and visualization, we identified key patterns such as the right-skewed nature of original prices, significant differences between original and discounted prices, and varying discount percentages across product categories. Handling missing values using the median proved to be the most suitable approach due to the presence of outliers and skewness in the pricing data.

The visualizations revealed that certain product categories consistently featured higher discounts, and scatter plots highlighted correlations between pricing variables. The analysis not only enhanced understanding of Adidas' pricing and discounting strategy but also demonstrated the importance of exploratory data analysis in deriving business insights. This project lays a strong foundation for further analysis such as customer behaviour prediction, sales forecasting, or optimization of discount strategies.



# PROJECT II

## ZARA PRODUCTION & SALES ANALYSIS

### Introduction:

The Zara production & sales Analysis project focuses on exploring, cleaning, and analysing product-level data to derive meaningful insights into pricing patterns, discount strategies, and product distribution. The dataset contains key information such as original prices, discounted prices, product categories, and other attributes relevant to understanding Zara product offerings and pricing behaviour. The primary objective is to handle missing data effectively, visualize critical trends, and assist in data-driven decision-making for pricing strategies.

### Python Code I:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import seaborn as sns
```

Explanation: These are the libraries imported in the Jupyter Notebook for reading csv file, performing numerical and scientific calculation, visualization with different techniques and so on.

### Python Code II:

```
missing_values = df.isnull().sum()
print("Missing values per column:\n", missing_values)
print("\nData Types:\n", df.dtypes)
```

Explanation: By using isnull(), we have checked out the null values in the dataset which can affect the analysis and result.

### **Python Code III:**

```
print("\nUnique values in 'Product Position':", df['Product Position'].unique())  
print("\nUnique values in 'Product Category':", df['Product Category'].unique())  
print("\nUnique values in 'name:', df['name'].unique())  
print("\nUnique values in 'terms:', df['terms'].unique())  
print("\nUnique values in 'section:', df['section'].unique())  
print("\nUnique values in 'currency:', df['currency'].unique())
```

Explanation: By using unique(), we have find out the unique values in the defined columns which can helps us for further analysis.

### **Python Code IV:**

```
df=df.drop('brand',axis=1)  
df=df.drop('description',axis=1)  
df=df.drop('currency',axis=1)
```

Explanation: By using drop attribute, we have removed unwanted columns form the dataset which have same entry in all the tuples.

### **Python Code V:**

```
df.describe(include='all')
```

Explanation: The describe method helps us to get the complete details about the dataframe like mean, median, standard deviation, quartile ranges, minimum value and maximum value.

Outcome of all the codes:

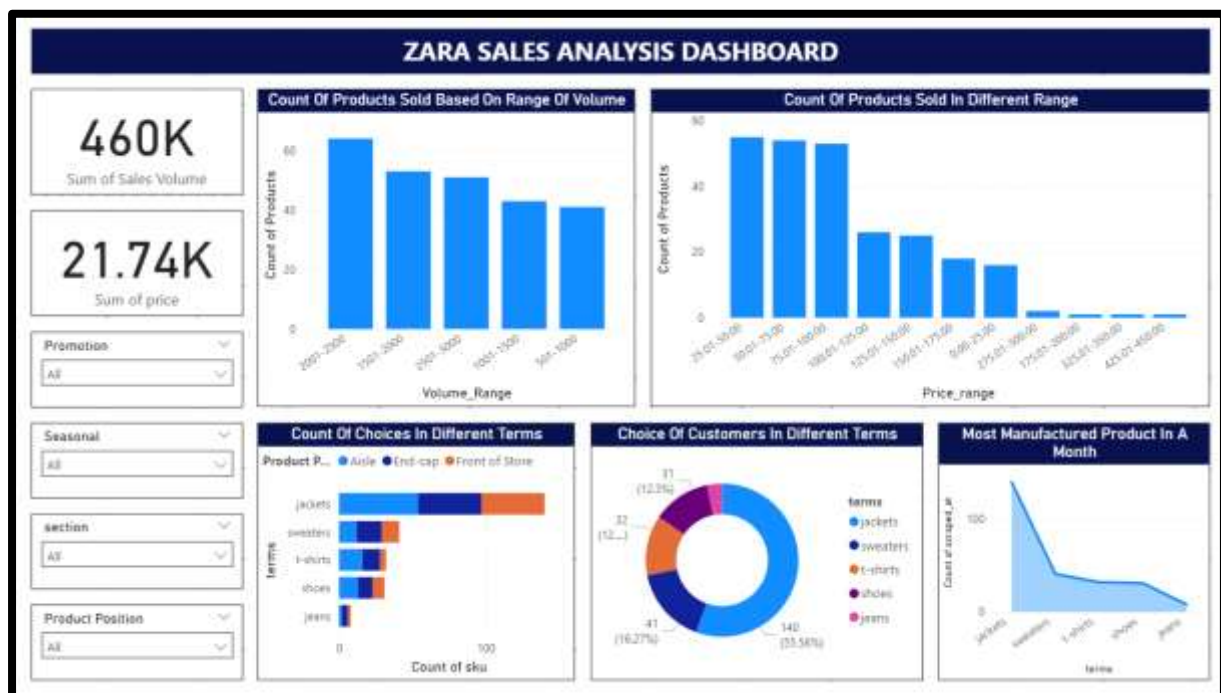
Missing values per column:		Data Types:	
Product ID	0	Product ID	int64
Product Position	0	Product Position	object
Promotion	0	Promotion	object
Product Category	0	Product Category	object
Seasonal	0	Seasonal	object
Sales Volume	0	Sales Volume	int64
brand	0	brand	object
url	0	url	object
sku	0	sku	object
name	1	name	object
description	2	description	object
price	0	price	float64
currency	0	currency	object
scraped_at	0	scraped_at	object
terms	0	terms	object
section	0	section	object
dtype: int64		dtype: object	

Unique values in 'Product Position': ['Aisle' 'End-cap' 'Front of Store']	
Unique values in 'Product Category': ['Clothing']	
Unique values in 'name': ['BASIC PUFFER JACKET' 'TUXEDO JACKET' 'SLIM FIT SUIT JACKET' 'STRETCH SUIT JACKET' 'DOUBLE FACED JACKET' 'CONTRASTING COLLAR JACKET' 'FAUX LEATHER PUFFER JACKET' 'SUIT JACKET IN 100% LINEN' '100% WOOL SUIT JACKET' '100% FEATHER FILL PUFFER JACKET' 'HERRINGBONE TEXTURED JACKET' 'OVERSIZED CROPPED JACKET LIMITED EDITION' 'LEATHER BIKER JACKET' 'CROPPED LEATHER JACKET' 'FAUX LEATHER BOXY FIT JACKET' 'FAUX LEATHER JACKET' 'FAUX SUEDE BOMBER JACKET' 'DENIM BOMBER JACKET' 'BOUCLÉ TEXTURED JACKET' 'JACQUARD DENIM JACKET' 'PADDED DENIM JACKET' 'LEATHER JACKET' 'LIGHTWEIGHT BOMBER JACKET' 'SUIT JACKET' 'FAUX LEATHER BOMBER JACKET' 'PATCH BOMBER JACKET' 'STRETCH POCKET OVERSHIRT' 'RIB COLLAR JACKET' 'FAUX LEATHER OVERSIZED JACKET LIMITED EDITION' 'CONTRASTING PATCHES BOMBER JACKET' 'CROPPED BOMBER JACKET LIMITED EDITION' 'BOMBER JACKET' 'FAUX SUEDE JACKET' 'SUEDE JACKET' 'TEXTURED JACKET' 'CROPPED TEXTURED JACKET' 'POCKET PUFFER JACKET' 'TECHNICAL JACKET WITH POCKETS' 'RIPPED DENIM JACKET' 'TEXTURED POCKET JACKET' 'FAUX SUEDE PATCH JACKET' 'PUFFER JACKET WITH POUCH POCKET' 'TEXTURED WEAVE OVERSHIRT' 'STRAIGHT SUIT JACKET' 'HOODED QUILTED JACKET' 'LIGHTWEIGHT PUFFER JACKET' 'COTTON BLEND BOMBER JACKET' 'POCKET JACKET' 'OVERSIZED BOMBER JACKET' 'EMBROIDERED PATCH JACKET' 'ACID WASH DENIM JACKET' 'VINTAGE EFFECT LEATHER BOMBER JACKET' 'TEXTURED DENIM JACKET LIMITED EDITION' 'WOOL BLEND JACKET']	
Unique values in 'terms': ['jackets' 'shoes' 'sweaters' 'jeans' 't-shirts']	
Unique values in 'section': ['MAN' 'WOMAN']	
Unique values in 'currency': ['USD']	

[7]:	Product ID	Product Position	Promotion	Product Category	Seasonal	Sales Volume	url	sku	name	price	scraped
0	185102	Aisle	No	Clothing	No	2823	https://www.zara.com/us/en/basic-puffer-jacket...	272145190-250-2	BASIC PUFFER JACKET	19.99	191085035.654
1	188771	Aisle	No	Clothing	No	654	https://www.zara.com/us/en/tuxedo-jacket-p0889...	324052738-800-46	TUXEDO JACKET	169.00	191085036.536
2	180176	End-cap	Yes	Clothing	Yes	2220	https://www.zara.com/us/en/slim-fit-suit-jacke...	335340580-800-44	SLIM FIT SUIT JACKET	129.00	191085037.321
3	112917	Aisle	Yes	Clothing	Yes	1568	https://www.zara.com/us/en/stretch-suit-jacket...	328303236-420-44	STRETCH SUIT JACKET	129.00	191085037.882
4	192936	End-cap	No	Clothing	Yes	2942	https://www.zara.com/us/en/double-faced-jacket...	312368260-800-2	DOUBLE FACED JACKET	139.00	191085038.453

	Unnamed: 0	Product ID	Product Position	Promotion	Product Category	Seasonal	Sales Volume	url	sku	name		
count	252.000000	252.000000	252	252	252	252	252.000000		252	252	251	251
unique	NaN	NaN	3	3	1	2	NaN		228	220	194	
top	NaN	NaN	Aisle	No	Clothing	Yes	NaN	https://www.zara.com/us/en/oversized-leath-swea...	318209563-800-2	PLAID	OWS48481	
freq	NaN	NaN	97	152	252	128	NaN		2	2	6	
mean	125.500000	153370.593968	NaN	NaN	NaN	NaN	1825.702307	NaN	NaN	NaN	NaN	81
std	72.890328	26168.844548	NaN	NaN	NaN	NaN	687.703748	NaN	NaN	NaN	NaN	52
min	0.000000	110075.000000	NaN	NaN	NaN	NaN	529.000000	NaN	NaN	NaN	NaN	1
25%	42.750000	131063.750000	NaN	NaN	NaN	NaN	1243.000000	NaN	NaN	NaN	NaN	41
50%	125.500000	151681.500000	NaN	NaN	NaN	NaN	1855.500000	NaN	NaN	NaN	NaN	71
75%	188.250000	179469.750000	NaN	NaN	NaN	NaN	2296.750000	NaN	NaN	NaN	NaN	108
max	251.000000	199631.000000	NaN	NaN	NaN	NaN	2989.000000	NaN	NaN	NaN	NaN	431

## Outcome using Power BI:



## Conclusion:

The Zara Sales Analysis revealed valuable insights into the brand's product manufacturing and sales distribution. Among all product categories, **jackets were the most frequently manufactured items**, indicating strong demand or a strategic focus on outerwear. In contrast, **jeans appeared to be the least manufactured**, which may reflect either lower demand, targeted inventory control, or seasonal preferences.

By analysing SKU-level data along with attributes like promotion, price, and seasonal tags, we identified Zara's emphasis on high-selling categories and the limited production of less-demanded items. These findings can help inform merchandising strategies, optimize inventory planning, and guide future decisions in product development and marketing.

# PROJECT III

## TOP TRENDING BRAND ANALYSIS

### Introduction:

This analysis focuses on the top brand preferred by customer and its sales record. This analysis focuses on exploring, cleaning, and analysing product-level data to derive meaningful insights into pricing patterns, discount strategies, and product distribution. The dataset includes brand details, product purchase price, reviews and other details for analysis. This analysis helps to find out which is the most preferred brand in the market.

### Python Code I:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Explanation: These are the libraries imported in the Jupyter Notebook for reading csv file, performing numerical and scientific calculation, visualization with different techniques and so on.

### Python Code II:

```
missing_values = df.isnull().sum()
print("Missing values per column:\n", missing_values)
print("\nData Types:\n", df.dtypes)
```

Explanation: By using isnull(), we have checked out the null values in the dataset which can affect the analysis and result.

### Python Code III:

```
print("\nUnique values in 'Gender':", df['Gender'].unique())
print("\nUnique values in 'NumImages':", df['NumImages'].unique())
print("\nUnique values in 'Color':", df['PrimaryColor'].unique())
```

Explanation: By using unique(), we have find out the unique values in the defined columns which can helps us for further analysis.

### **Python Code IV:**

```
most_common_brand=df['ProductBrand'].value_counts().idxmax()
most_common_count=df['ProductBrand'].value_counts().max()
print(f'Most entered ProductBrand:{most_common_brand} ({most_common_count})
entries")
```

Explanation: This code will helps us to find out the most common brand in the dataset whose customer count is maximum

### **Python Code V:**

```
most_common_product=new_df['ProductName'].value_counts().idxmax()
most_common_count=new_df['ProductName'].value_counts().max()
print(f'Most entered ProductBrand:{most_common_product} ({most_common_count})
entries")
```

Explanation: This code will helps us to find out the most common product in the dataset which is purchased in higher quantity

### **Python Code VI:**

```
most_occured_brand=brand_counts[brand_counts>50]
most_occured_brand
most_occured_product=product_counts[product_counts>2]
most_occured_product
```

Explanation: This code helps us to find out the most occurred brand and product in the dataset which have specify count.

Outcome of all the codes:

```

# Unique values for each column
unique_values = {}
for column in df.columns:
    unique_values[column] = df[column].unique()

# Data Types
data_types = {}
for column in df.columns:
    data_types[column] = df[column].dtypes

# Unique values for 'Gender'
unique_gender = df['Gender'].unique()

# Unique values for 'NumImages'
unique_images = df['NumImages'].unique()

# Unique values for 'PrimaryColor'
unique_color = df['PrimaryColor'].unique()

# ProductBrand
ProductBrand
Indian Terrain      971
Puma                345
Pepe Jeans          340
AURELIA             307
Flying Machine      301
...
Johnny Bravo by Kook N Keech  1
Allen Solly Sport  1
HAMMER               1
CUTE CUMBER         1
Stupid Cupid        1
Name: count, Length: 677, dtype: int64

```

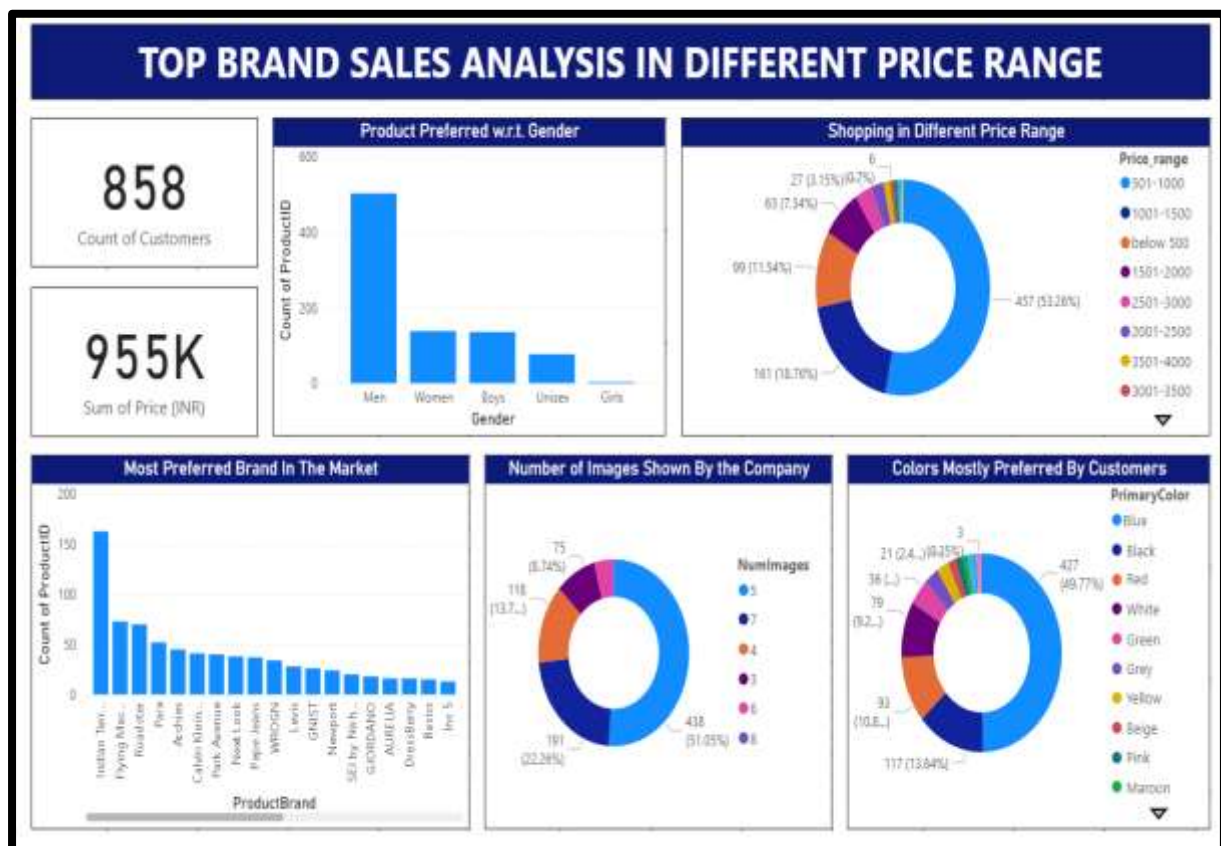
	ProductID	ProductName	ProductBrand	Gender	Price (INR)	NumImages	Description	PrimaryColor
3	10015921	Raymond Men Blue Self-Design Single-Breasted B...	Raymond	Men	5599	5	Blue self-design bandhgata suitBlue self-design...	Blue
4	10017833	Parx Men Brown & Off-White Slim Fit Printed Ca...	Parx	Men	759	5	Brown and off-white printed casual shirt, has ...	White
6	10017669	Parx Men Blue Slim Fit Checked Casual Shirt	Parx	Men	719	5	Blue checked casual shirt, has a spread collar...	Blue
8	10000571	Parx Men Brown Tapered Fit Solid Regular Trousers	Parx	Men	664	5	Brown solid regular trousers regular trousers	Red
11	10000245	Parx Men Green Printed Polo Collar T-shirt	Parx	Men	629	5	Green printed T-shirt, has a polo collar, and ...	Green
--	--	--	--	--	--	--	--	--
12481	1026753	Sonari Pack of 2 Bras afeen	Sonari	Women	750	4	Pack of two full-coverage bras in pink and mar...	Naft
12483	1027006	Sonari Pack of 2 Full-Coverage Bras calina	Sonari	Women	920	2	Pack of two full-coverage bras in whiteEach bra...	Naft
12484	10265177	Pepe Jeans Men Blue Printed Slim Pure Cotton R...	Pepe Jeans	Men	649	6	Blue and Green printed T-shirt, has a round ne...	Blue
12486	10262843	Pepe Jeans Men Black Hammock Slim Fit Low-Rise...	Pepe Jeans	Men	1299	2	Black dark wash 5-pocket low-rise jeans, cham...	Black
12490	10265199	Pepe Jeans Men Black & Gray Striped Polo Colla...	Pepe Jeans	Men	799	5	Black and gray striped T-shirt, has a polo col...	Black

```

# ProductName
ProductName
Parx Men Blue Slim Fit Checked Casual Shirt      16
Roadster Men Blue Skinny Fit Mid-Rise Clean Look Stretchable Jeans  14
Sonari Pack of 2 Full-Coverage Bras              14
Roadster Men Blue Slim Fit Mid-Rise Clean Look Stretchable Jeans    13
Next Look Men Blue Slim Fit Self Design Formal Shirt                12
..
Next Look Men Blue Regular Fit Printed Formal Shirt                 3
Indian Terrain Men Yellow Slim Fit Printed Casual Shirt             3
Archies Love Gifts White & Red Printed Coffee Mug                  3
Indian Terrain Men Brown & White Slim Fit Checked Smart Casual Shirt 3
SEJ by Nisha Gupta Set of 4 Printed Table Placemats                 3
Name: count, Length: 223, dtype: int64

```

## Outcome using Power BI:



## Conclusion:

In this project, we analysed the product catalogue to understand the distribution and frequency of various product brands present in the market. The key focus was to identify the most frequently entered brands and analyse the brand dominance within the dataset.

Upon detailed examination, we found that the dataset consists of a **large number of unique brands**, with most of them appearing only a **few times**. This indicates that Myntra's product range is **highly diverse and decentralized**, showcasing a wide variety of brands instead of relying heavily on a few popular names.



# PROJECT IV

## MOST PROFITABLE BRAND ANALYSIS

### Overview

#### Introduction:

As part of my internship at **Aneja Sons Platinum Store**, I conducted a data analysis project with the objective of identifying the most profitable and most preferred brands in the store's inventory. The store offers a diverse range of **readymade garments for gents, girls, ladies, boys, and kids**, and carries multiple well-known brands such as **Indian Terrain, Camla, Madame, Monte Carlo, Pepe Jeans, Octave, Machine, Dragon, Only**, and others.

During the analysis, I focused specifically on the **gents and girls wear categories**, as requested by the store management. By comparing total profit margins, sales frequency, and brand-wise performance, I identified clear trends in customer preferences and profitability.

This project not only helped improve business understanding through data but also allowed me to apply real-world analytical skills by integrating data collection, data transformation, and data visualization in a retail environment. It was a valuable learning experience in both technical and business domains.

#### Entries Collected:

To achieve this, I collected and compiled over **two months of sales records** from the store into an **Excel workbook**, which served as the initial data source. The dataset included key attributes such as:

1. Customer ID
2. Contact Information
3. Gender
4. Product Purchased
5. Brand Name
6. MRP (Maximum Retail Price)
7. Selling Price
8. Wholesale Price
9. Profit per Item
10. Date of Purchase

These fields provided a comprehensive view of the store's sales performance and enabled granular profit analysis.

## Analysis:

After the data collection phase, I used **Python** for cleaning and transforming the dataset. This included handling missing values, standardizing brand names, calculating profit margins, and preparing the data for visualization. Python libraries such as pandas and numpy were instrumental in processing the data efficiently.

The final phase involved **creating interactive dashboards using Power BI**. Power BI enabled me to visualize trends, such as:

1. Top-performing brands based on total profit
2. Profit comparison between gents and girls' wear
3. Brand-wise performance over time
4. Monthly and daily sales trends
5. Profitability per product and per transaction

These visual insights provided clear, actionable intelligence that could be used by the store's management team to prioritize specific brands, optimize inventory, and focus marketing strategies on high-profit segments.

## Key Points:

The key findings from the analysis are as follows:

- **Indian Terrain** and **Monte Carlo** emerged as the **most preferred and highest selling brands in the gents wear segment**. These brands not only had higher frequency of sales but also contributed significantly to overall profit margins.
- In the **girls wear category**, **Madame** stood out as the **most popular brand**, with consistently high sales and notable profit generation.
- Other brands such as **Pepe Jeans**, **Camla**, and **Only** showed moderate performance, while **Machine and Dragon** had lower sales and profit contributions during the observed period.

By leveraging **Power BI**, I created dynamic dashboards that allowed the store's team to interact with brand-wise and category-wise visual insights, filter by date range, and compare sales trends over time.

## Conclusion:

This project offered meaningful insights into the **sales behavior and brand profitability** at Aneja Sons Platinum Store. The data revealed that customer preferences are strongly aligned with specific brands in different segments—**Indian Terrain and Monte Carlo in gents wear**, and **Madame in girls wear**—which helps the store identify where to focus inventory, marketing, and sales strategies.

From a learning perspective, this project helped me apply real-world skills in **data collection, cleaning, analysis, and visualization**, and demonstrated how data can directly support business decision-making. The outcome of this project can assist the store's management in enhancing profitability, optimizing brand partnerships, and better understanding customer demand patterns.

# CONTRIBUTION TO ORGANIZATION

## Value Added Through Data Insights and Analysis

1. During my internship, I contributed to the organization by providing actionable insights derived from data analysis. By analyzing large datasets, I helped identify trends and patterns that informed key business decisions. For example, I analyzed sales data to uncover seasonal patterns, enabling the marketing team to optimize their campaigns and improve conversion rates.
2. Additionally, my analysis of customer behavior allowed the company to tailor its product offerings to better meet customer needs, directly impacting customer satisfaction and retention rates.

## Improvements Made to Data Processes or Workflows

1. I identified inefficiencies in the data cleaning and processing workflows. By automating repetitive tasks using Python scripts, I helped reduce the time spent on data preprocessing, allowing the team to focus more on analysis and strategic decision-making.
2. I also suggested and implemented a more streamlined reporting system using Power BI dashboards, which improved the speed and accuracy of data visualization, enabling stakeholders to access real-time insights and make quicker decisions.

## Feedback and Recognition from Supervisors

1. My supervisors provided positive feedback on my ability to quickly learn new tools and adapt to the project requirements. They particularly appreciated my contributions to improving data processing workflows and my proactive approach in suggesting new methods for data analysis.
2. I received recognition for my initiative in automating data preparation tasks, which resulted in significant time savings for the team. Additionally, my clear and concise presentations of complex data findings were commended, as they made it easier for non-technical team members to understand the insights.

# CONCLUSION

## Summary of the Internship Journey

1. My internship at ANEJA SONS PLATINUM STORE has been a transformative experience that allowed me to apply the skills I developed in my academic studies to real-world challenges. Over the course of the internship, I gained hands-on experience with data analysis tools, learned how to handle complex datasets, and improved my problem-solving, communication, and teamwork skills. I had the opportunity to work on meaningful projects that added value to the organization, such as automating data workflows and providing actionable insights to support decision-making.
2. Additionally, the mentorship and guidance I received from my supervisors and colleagues helped me grow both professionally and personally, preparing me for future roles in the data science and analytics field. The internship not only allowed me to contribute to the company but also gave me a deeper understanding of the corporate environment and its demands.

## Future Plans Influenced by the Experience

1. This internship has solidified my passion for data science and analytics, and I plan to further develop my technical skills in areas like machine learning and advanced data visualization. I am now more focused on pursuing a career where I can leverage data to drive business outcomes and solve real-world problems.
2. Moving forward, I aim to build upon the knowledge I gained during the internship by continuing to expand my expertise in data analysis tools and techniques. I also plan to seek out additional opportunities to collaborate with cross-functional teams, as I've seen how important effective communication and teamwork are in delivering impactful results.
3. Ultimately, the internship has influenced my future career plans by shaping my aspirations toward becoming a data analyst or data scientist, with a strong focus on utilizing data to inform strategy and improve business performance.

# KEY BENEFITS OF INTERNSHIP

The internship at **Aneja Sons Platinum Store** provided me with a valuable opportunity to apply academic knowledge in a real-world retail environment. Some of the key benefits I gained from this internship experience are:

## 1. **Practical Exposure to Data Analytics**

I learned how to work with real sales data, clean and prepare it using Python, and extract meaningful insights. This hands-on experience strengthened my understanding of data analysis concepts and their practical application in business decision-making.

## 2. **Enhanced Proficiency in Tools**

Through this project, I became more proficient in using tools like **Microsoft Excel** for data entry, **Python** for data transformation, and **Power BI** for interactive data visualization. This multi-tool approach helped me understand the complete data workflow from collection to dashboard creation.

## 3. **Business Understanding of Retail Sector**

By analysing sales and profitability, I gained a deeper understanding of how retail businesses operate, how brand performance is evaluated, and how data can support strategic decisions like stock planning and brand promotion.

## 4. **Improved Communication and Reporting Skills**

I developed the ability to explain complex data findings in simple terms to stakeholders. Creating visual reports and summaries enhanced my ability to communicate insights effectively.

## 5. **Time Management and Responsibility**

Managing this project independently within a limited time frame helped me improve my time management skills and taught me how to work efficiently with deadlines and deliverables.

## 6. **Confidence in Problem-Solving**

Handling real data challenges such as missing values, inconsistent brand entries, and formatting issues taught me how to solve problems logically and confidently.

# INTERNSHIP CERTIFICATE

## Aneja Sons Platinum Store

GST NO-05AANPA7660B1Z2

## CERTIFICATE

### OF INTERNSHIP EXPERIENCE

This certificate is proudly presented to

**KRISHNA RAJOO**

successfully completed the Internship at

**ANEJA SONS – PLATINUM STORE** in Data Science active participant from **June 2<sup>nd</sup>, 2025 to July 27<sup>th</sup>, 2025** with unwavering dedication.

Mr. Chandan Aneja

Owner, Aneja Sons – Platinum Store

Contact: 09758320000

  
Aneja Sons Platinum Store  
Civil Lines, Prem Mandir Road,  
Roorkee-247 667

[www.anejasonsplatinumstore.in](http://www.anejasonsplatinumstore.in)

DATE OF ISSUE:-July 29<sup>th</sup> 2025

# PERSONAL GROWTH

## 1. Professional Development and Lessons Learned

- a. This internship enhanced my technical skills in data analysis and gave me practical experience working with real-world data. I learned how to effectively collaborate with different teams, manage projects, and apply my academic knowledge in a business context.
- b. The experience also taught me how to approach problems from a solution-oriented perspective, allowing me to think critically and develop strategies to overcome challenges.

## 2. Impact on Career Aspirations

- a. My time at ANEJA SONS PLATINUM STORE has solidified my aspiration to pursue a career in data science and analytics. The hands-on experience with tools like Python, SQL, and Power BI has given me a clearer sense of the technical expertise required in the industry.
- b. Additionally, I now recognize the importance of communication and collaboration in achieving business objectives, which will influence my future career approach.

# CHALLENGES

## 1. Technical Challenges

- a. **Complex Datasets:** One of the primary technical challenges was working with large and complex datasets that required significant cleaning and preprocessing. These datasets often contained missing values, outliers, and inconsistencies, making it difficult to derive meaningful insights.
- b. **Software Limitations:** While working with tools like Power BI and Excel, I encountered limitations in handling extremely large datasets, which slowed down data processing and analysis. Some of the advanced functionalities I needed were not available in the versions of the software used at the company.

## 2. Adapting to the Corporate Environment

- a. Transitioning from an academic setting to a corporate environment required adjusting to a faster pace and different expectations. I had to quickly learn to prioritize tasks, meet tight deadlines, and adapt to formal communication styles within the team and with clients.
- b. Additionally, understanding the business context of the data was a challenge, as I initially lacked deep knowledge of the industry, which made it harder to interpret the data in a way that was valuable for decision-making.



# RECOMMENDATION FOR FUTURE INTERN

## Tips for Maximizing the Internship Experience

1. **Be Proactive and Take Initiative:** Don't wait for instructions on every task—look for opportunities where you can contribute, ask questions, and offer suggestions. Being proactive not only helps you gain more experience but also shows your enthusiasm and commitment to the role.
2. **Ask for Feedback Regularly:** Seek constructive feedback from your mentors and supervisors. This will help you understand your strengths and areas for improvement, ensuring continuous growth throughout the internship.
3. **Build Strong Relationships:** Networking within the organization is key. Take the time to build relationships with your colleagues, mentors, and team members. This will help you learn from their experiences, get advice, and create connections that may benefit your future career.
4. **Stay Organized and Manage Your Time Effectively:** Internships often come with multiple tasks and deadlines. Keeping track of your responsibilities and staying organized will help you deliver quality work on time. Use tools like task lists, calendars, or project management apps to keep everything in check.
5. **Engage with the Business Side:** While technical skills are important, understanding the business context and how your work impacts the organization is essential. Take the time to learn about the company's objectives and the industry it operates in. This will help you contribute more effectively to the team's goals.

## Areas of Focus and Improvement Suggestions

1. **Technical Skills:** For future interns, it would be beneficial to deepen knowledge in specific tools or software commonly used within the industry. While learning data analysis tools like Python, SQL, and Excel is important, focusing on specialized software (such as advanced visualization tools or machine learning platforms) would be an advantage.
2. **Communication Skills:** Although technical expertise is essential, the ability to communicate findings clearly to non-technical stakeholders is equally crucial. Future interns should focus on enhancing their presentation and communication skills to make their insights easily understandable and actionable.
3. **Handling Complex Data:** Interns should focus on improving their ability to work with large, messy datasets. Understanding how to clean and preprocess data effectively will be an important skill for any data-related role.
4. **Industry Knowledge:** Future interns could benefit from spending time understanding the business processes and industry-specific challenges the organization faces. This deeper knowledge will help interns make their work more relevant and impactful.