ADIDAS SALES DATA

A PROJECT REPORT

Submitted by

NISHANT REDDY

DHANUSH S GOWDA

ANKUR T S

DARSHAN B

in partial fulfillment for the award of the degree

of

B.Tech(Hons.)

in

Computer Science and Engineering

School of Computer Science and Engineering

RV University

RV Vidyaniketan, 8th Mile, Mysore Road, Bengaluru, Karnataka, India - 562112

ABSTARCT

This project delves into the intricacies of Adidas Sales in the USA, driven by the recognition of Adidas as a pivotal player in the sportswear industry. The rationale behind selecting this dataset lies in its potential to offer valuable insights into consumer behavior, market trends, and operational strategies within the sportswear sector.

Our objective encompasses deciphering patterns, identifying influential factors impacting sales, and furnishing actionable insights beneficial for strategic decision-making.

The significance of our analysis extends beyond the confines of a semester project. By comprehensively understanding Adidas sales dynamics, we aim to provide a foundation for future endeavors in market research and data analytics. The insights derived can inform Adidas's marketing strategies, inventory management, and overall business decisions. Furthermore, this analysis serves as a benchmark for other industry stakeholders seeking to enhance their understanding of market dynamics and consumer preferences.

The project required meticulous efforts from our team, spanning data preprocessing, exploratory data analysis, and the application of statistical methods. Challenges, such as handling missing data and outliers, were addressed systematically to ensure the integrity and reliability of our findings. The collaborative nature of our team, encompassing diverse skill sets in Python Programming and Statistical Analysis, has been instrumental in the success of this project.

The project's relevance lies not only in its contribution to academic learning but also in its practical applications for industry players. The effort invested in understanding and interpreting the data contributes to a holistic understanding of the complexities inherent in data analysis.

In conclusion, this project serves as a steppingstone for future investigations into market dynamics, consumer behavior, and data-driven decision-making. The insights made from our analysis present a valuable resource for academia, industry professionals, and Adidas alike, positioning this endeavor as a meaningful contribution to the field of data analysis and business intelligence

METHODOLOGY:
WORK STREAM of THE CODE

1. ***Importing necessary libraries and dataset.***

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

# Loading the Adidas sales dataset
df = pd.read_csv('Adidas US Sales Datasets.csv')
df.head(10)
```

Explanation: We begin by importing essential libraries for data manipulation and visualization. Then, we load the Adidas sales dataset into a Pandas DataFrame named 'df' and display the first 10 rows to get an initial look at the data.

2. ***Data Inspection.***

```
# Getting the shape of the dataset
df.shape

# Listing column names
df.columns

# Displaying dataset information
df.info()
```

Explanation: This section provides a quick overview of the dataset. We check its dimensions, column names, and information about data types and null values to understand its structure.

3. ***Data Cleaning (Data Preprocessing).***

```
# Data cleaning and preprocessing
df['Price_per_Unit'] = df['Price_per_Unit'].astype(str).str.replace('$',
'').str.replace(',', '').astype(float)
df['Total_Sales'] = df['Total_Sales'].astype(str).str.replace('$',
'').str.replace(',', '').astype(float)
df['Operating_Profit'] = df['Operating_Profit'].astype(str).str.replace('$',
'').str.replace(',', '').astype(float)
df['Units_Sold'] = df['Units_Sold'].astype(str).str.replace(',', '').astype(float)
df['Operating_Margin'] = df['Operating_Margin'].astype(str).str.replace('%',
'').astype(float)

df['Total_Sales'] = df['Price_per_Unit'] * df['Units_Sold']
df['Operating_Profit'] = df['Total_Sales'] * df['Operating_Margin'] / 100

for column in df.columns:
   if df[column].dtype == 'O':
      df[column] = df[column].str.lower()

df['Invoice_Date'] = pd.to_datetime(df['Invoice_Date'], format="%d-%m-
%Y")
df.drop_duplicates(inplace=True)
df_clean = df.copy()
```

Explanation: This part involves cleaning and preprocessing the data. We convert currency columns to appropriate numeric types, calculate new columns for total sales and operating profit, ensure text consistency by converting to lowercase, and handle the 'Invoice_Date' column as datetime. Duplicate rows are also removed.

4. *Exploratory Data Analysis (EDA).*

```
# Descriptive statistics of the cleaned dataset
df_clean.describe()

# Calculating the range of numerical columns
range = ['Price_per_Unit', 'Units_Sold', 'Total_Sales', 'Operating_Profit',
'Operating_Margin']
for r in range:
    print("Range of ", r , " : ", df_clean[r].max() - df_clean[r].min())
```

Explanation: We explore the dataset by obtaining descriptive statistics and calculating the range of numerical columns to understand the spread and characteristics of the data.

5. *Sales Analysis*

```
most_listed_product = df_clean['Product'].value_counts().idxmax()
product_sales = df_clean.groupby('Product')['Total_Sales'].sum()

plt.figure(figsize=(8, 5))
product_sales.plot(kind='bar', color='skyblue')
plt.xlabel('Product')
plt.ylabel('Total Sales')
plt.title('Total Sales per Product')
plt.show()

df_clean['Year'] = df_clean['Invoice_Date'].dt.year
df_clean['Month'] = df_clean['Invoice_Date'].dt.month
monthly_data = df_clean.groupby(['Year', 'Month']).agg({
    'Total_Sales': 'sum',
    'Operating_Profit': 'sum'
}).reset_index()
```

```
plt.figure(figsize=(14, 7))
plt.plot(monthly_data['Date'], monthly_data['Total_Sales'], label='Total Sales')
plt.plot(monthly_data['Date'], monthly_data['Operating_Profit'], label='Operating Profit')
plt.title("Total Sales and Operating Profit over Time")
plt.xlabel('Time')
plt.ylabel('Amount')
plt.legend()
plt.show()
```

Explanation: In this section, we identify the most listed product, visualize total sales per product using a bar chart, and analyze the trend of total sales and operating profit over time with a time series plot.

6. ***Profitability Analysis.***

```
df_clean.groupby(['Product'])['Operating_Profit'].median()

df_clean.groupby(['Product'])['Operating_Margin'].median()

monthly_data = df_clean.groupby(['Year', 'Month']).agg({
    'Operating_Margin': 'median'
}).reset_index()

plt.figure(figsize=(14, 7))
plt.plot(monthly_data['Date'], monthly_data['Operating_Margin'], label='Operating Margin')
plt.title("Median Operating Margin over Time")
plt.xlabel('Time')
plt.ylabel('Amount')
plt.legend()
plt.show()
```

Explanation: This segment involves profitability analysis. We calculate the median operating profit and median operating margin by product. Additionally, we create a time series plot to visualize how the median operating margin changes over time.

7. ***Regional Analysis.***

```
print("Region wise: ",df_clean.groupby(['Region'])['Total_Sales'].sum())
print("State wise: ",
df_clean.groupby(['State'])['Total_Sales'].sum().idxmax())
print("City wise: ", df_clean.groupby(['City'])['Total_Sales'].sum().idxmax())
```

Explanation: In this part, we delve into regional analysis. We display total sales by region and identify the state and city with the highest total sales.

8. ***Retailer Analysis.***

```
# Calculating total sales by retailer
sales_by_location = df_clean.groupby('Retailer')['Total_Sales'].sum()

# Plotting a pie chart for total sales by retailer
plt.figure(figsize=(8, 8))
sales_by_location.plot.pie(autopct='%1.1f%%', startangle=90,
colors=['skyblue', 'lightgreen', 'lightcoral'])
plt.title('Total Sales by Retailer')
plt.ylabel('')
plt.show()

# Calculating total operating profit by retailer
sales_by_location = df_clean.groupby('Retailer')['Operating_Profit'].sum()

# Plotting a pie chart for total operating profit by retailer
plt.figure(figsize=(8, 8))
sales_by_location.plot.pie(autopct='%1.1f%%', startangle=90,
colors=['skyblue', 'lightgreen', 'lightcoral'])
```

```
plt.title('Total Operating Profit by Retailer')
plt.ylabel('')
plt.show()

df_clean.groupby(['Retailer', 'Sales_Method'])['Operating_Profit'].mean()
```

Explanation: This section focuses on retailer analysis. We calculate total sales and operating profit by retailer, visualize this information with pie charts, and explore average operating profit by retailer and sales method.

9. ***Pricing Analysis.***

```
df_clean[['Price_per_Unit','Units_Sold', 'Total_Sales', 'Operating_Profit',
'Operating_Margin']].corr()

heatmap = df_clean[['Price_per_Unit','Units_Sold', 'Total_Sales',
'Operating_Profit', 'Operating_Margin']]
plt.figure(figsize=(10, 8))
sns.heatmap(heatmap.corr(), cmap='viridis', annot=True, fmt=".2f",
linewidths=.5)
plt.title('Correlation Heatmap')
plt.show()

plt.figure(figsize=(10, 8))
sns.boxplot(x='Product', y='Price_per_Unit', data=df_clean)
plt.title('Box Plot of Price Per Unit Grouped by Product')
plt.xlabel('Product')
plt.ylabel('Price Per Unit')
plt.xticks(rotation=45)
plt.show()

plt.figure(figsize=(10, 8))
sns.boxplot(x='Product', y='Operating_Margin', data=df_clean)
plt.title('Box Plot of Operating Margin Grouped by Product')
plt.xlabel('Product')
plt.ylabel('Operating Margin')
```

```
plt.xticks(rotation=45)
plt.show()
```

Explanation: Finally, we enter the domain of pricing analysis. We calculate the correlation matrix for relevant columns, visualize the correlations using a heatmap, and create box plots to understand the distribution of price per unit and operating margin grouped by product.

This sequential walkthrough guides us through the comprehensive analysis of Adidas sales data, from cleaning and exploring the data to uncovering insights about sales, profitability, regions, retailers, and pricing

CONCLUSION

In conclusion, our Adidas Sales Project has provided valuable insights into the dynamics of the brand's performance in the USA market. Beginning with data cleaning and exploration, we ensured the foundation for our analysis was robust.

The sales analysis revealed key trends, identifying popular products and showcasing the evolution of total sales and operating profit over time. This phase set the stage for a comprehensive understanding of Adidas's market presence.

Profitability analysis deepened our insight into the financial performance of each product, offering a nuanced perspective on their contributions. The regional analysis brought geographical nuances to the forefront, emphasizing the impact of location on sales patterns.

Retailer analysis illuminated the distinctive contributions of each player, contributing to a more comprehensive understanding of Adidas's market partnerships. Finally, the pricing analysis delved into the strategic aspects, decoding the relationships between pricing metrics.

In essence, this project has been a systematic exploration, moving beyond the surface to uncover the intricacies of Adidas sales. Each phase added a layer of understanding, transforming raw data into meaningful insights. This comprehensive analysis not only informs our understanding of past performance but also lays a solid foundation for future strategic considerations in the dynamic landscape of the sportswear market.