

ADIDAS SALES DATA ANALYSIS

Project by-: Shashank Singh

1.INTRODUCTION

This project delves into the intricacies of Adidas Sales in the USA during 2020 & 2021, driven by the recognition of Adidas as a pivotal player in the clothing industry.

The rationale behind selecting this dataset lies in its potential to offer valuable insights into consumer behaviour, market trends, and operational strategies within the clothing and sportswear sector.

This project serves as a stepping stone for future investigations into market dynamics, consumer behavior and data-driven decision-making. The insights from this analysis will be valuable for Adidas professionals, and Adidas alike, positioning this endeavor as a meaningful contribution to the field of data analysis and business intelligence.

2.TOOLS AND LIBRARIES

- Pandas
- NumPy
- Matplotlib
- Google Collab
- SciPy
- Seaborn

1.DATA COLLECTION

The data collection process involves gathering measuring, and recording information or data points, typically through systematic methods such as surveys, interviews, or sensor technology. We collected Adidas Sales data from the years 2020 & 2021 in USA from various open data base sources and Kaggle.

2. DATA CLEANING

The data cleaning process involves identifying and rectifying errors in accuracies, and inconsistencies in a dataset to ensure its accuracy, reliability, and suitability for analysis.

3. DATA ANALYSIS

The data analysis process involves deriving insights from data through organizing, exploring, interpreting, and modeling, to make informed decisions and solve problems effectively. We used various data analysis tools in Python, including Pandas and Numpy to analyze the data and perform feature engineering.

4. DATA VISUALIZATION

The data visualization process involves representing and presenting data in visual form using charts, graphs, and other visual elements to facilitate understanding and analysis. We used Matplotlib and Seaborn to create visualizations that insights into key trends and patterns in the sales data.

1. Importing necessary libraries and dataset.

I begin by importing essential libraries for data manipulation and visualization. Then, I load the Adidas sales dataset into a Pandas data frame named 'sales' and display the first 20 rows to get an initial look at the data.

2. Data Inspection.

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. There were 9648 rows and 14 columns.

3.Data Cleaning (Data Preprocessing).

This part involves cleaning and preprocessing the data. Change the 'Invoice Date' column as datetime. There were no duplicate rows and columns involved in the dataset.

Also Checks For the Null Value and there is No NULL Value Present .

4.Exploratory Data Analysis (EDA).

1. Top 5 States With Highest Total Sales

This code groups the sales data by state and calculates the total sales for each. Then, it uses the `n largest()` function to retrieve the top 5 states with the highest total sales. This helps identify the highest-performing states in terms of sales.

2. Sales By Product Category

Here, the sales data is grouped by product category to calculate the total sales for each category. The result shows which product categories contribute the most to total sales, which is useful for product performance analysis.

3. Price per Unit vs. Units Sold

This scatter plot visualizes the relationship between the price per unit and the number of units sold. By plotting these two variables, we can analyze how pricing influences the quantity sold and find trends in the data.

4. Sales by Gender Type

This bar plot aggregates the total sales based on gender type, providing insight into which gender category drives more sales. It can be used to assess market targeting and identify gender-specific preferences.

5. Units Sold by Region

This bar plot aggregates the total units sold by region. It helps identify which regions have higher sales volumes and can be useful for regional sales strategy and performance comparison.

6. Monthly Sales Trend

A line plot is created to visualize the trend of units sold over time (monthly), helping to observe sales patterns and seasonality in the data. This is useful for trend analysis and future sales forecasting.

7. Correlation Heatmap for the Specified Numerical Columns

This heatmap visualizes the correlation matrix between numerical variables like total sales, price per unit, and operating profit. It helps

to quickly spot relationships between these variables for deeper analysis.

8. Monthly Sales Heatmap

A heatmap is created to visualize total sales by day and month, showing how sales vary across different days and months. This helps to identify trends or peak days in sales.

9. Sales Distribution by Product Category

This box plot shows the distribution of sales for each product category, providing a view of the spread and outliers in the data. It's useful for Analyzing variability in sales between categories.

10. Sales Density by Region

The Kernel Density Estimate (KDE) plot shows the distribution of sales for each region. This is useful for analyzing the concentration of sales and identifying high-density regions.

11. Relationship Between Total Sales and Operating Profit by Product Category

This linear regression plot helps visualize the relationship between total sales and operating profit for each product category. It can be used to assess the profitability of different product lines.

12. Which Region has the Highest Sales Revenue?

This code groups the data by region and sums the total sales for each region. The result highlights the region with the highest sales, which can inform regional sales strategies and resource allocation.

13. Which State has the Highest Total Profit?

The code groups the sales data by state and sums the operating profit for each state. It then identifies the state with the highest total profit, which can guide state-specific profit optimization strategies.

14. Which Items have the Highest and Lowest Total Sales?

This scatter plot shows the total sales for each product category, with special markers for the highest and lowest sales. This helps visually compare sales across categories and spot extremes.

15. Which Sales Channel has the Highest Average Unit Price?

This pie chart visualizes the average unit price for each product category, helping to compare how pricing differs across product categories and which category has the highest average price.

CONCLUSION

In conclusion, Adidas Sales Project provided valuable insights into the brand's performance in the USA market. Starting with data cleaning and exploration, we established a robust foundation for analysis. The sales analysis highlighted key trends, popular products, and the evolution of sales and operating profit over time, offering a comprehensive view of Adidas's market presence.

Profitability analysis offered a nuanced understanding of each product's financial contribution, while regional analysis emphasized the geographical impact on sales patterns. Retailer analysis showcased the roles of market partnerships, and pricing analysis decoded strategic pricing relationship