CSST 104

Final Project:

Machine Learning Implementation

Topic:

Adidas Sales

Submitted by:

Garcia, Marcus Henson L.

Viray, Joyce Anne M.

BSCS 3A

TABLE OF CONTENTS

TABLE OF CONTENTS

Google Collab Link:
https://colab.research.google.com/drive/1v8RkrJZYaXVkkY7btJ3e8bHcR78VtVhT?usp=sharing

Github Website Link:
https://github.com/hensonnnn/CSST104-Garcia-Viray-3A

Data Analysis and Machine Learning Implementation

Project Documentation Template

I. Project Overview

The project aims to analyze Adidas sales data to uncover trends, patterns, and insights that can inform business decisions. The main goals are to understand sales performance, identify key factors influencing sales, and recommend strategies to boost sales and optimize inventory. Insights expected to be derived include seasonal trends in sales, geographical sales patterns, and factors driving sales variation.

Here's how each attribute contributes to understanding sales data:

1. **Retailer ID:** A special number given to every store that sells products from Adidas. It is useful for analyzing retailer-specific trends and performance indicators, pinpointing the stores that are generating the most sales, and tracking sales performance at a detailed level.
2. **Invoice Date:** The date when the sales transaction was recorded. This characteristic is crucial for understanding seasonal patterns, assessing the effect of particular marketing campaigns or events on sales, and assessing sales trends over time.
3. **Region:** Geographic region in which the sale took place. Allocating resources, developing strategic plans, and adjusting marketing campaigns to suit local trends and tastes all benefit from an understanding of sales by region.
4. **State**: Specific state within a country where the sale was made. This level of detail provides more localized insights into sales performance, helping to identify high-performing states and address areas with lower sales.
5. **City:** The city where the sale took place. This granularity is crucial for urban market analysis, understanding city-specific demand, and optimizing distribution channels and retail locations within cities.

6. **Gender Type:** Indicates the gender for which the product is intended. This attribute is vital for demographic analysis, helping to tailor product offerings and marketing strategies to different gender segments.
7. **Product Category:** The classification of the product sold (e.g., footwear, apparel, accessories). Analyzing sales by product category helps in understanding which categories are most popular, managing inventory, and planning product development.
8. **Price per Unit:** The selling price of a single unit of the product. This data is crucial for revenue calculations, pricing strategy analysis, and understanding the price sensitivity of different products and markets.
9. **Units Sold:** Number of units of a product sold in a transaction. Tracking units sold helps in volume analysis, inventory management, and understanding product demand.
10. **Total Sales:** The total revenue generated from the sale. This is a key metric for assessing overall sales performance and revenue generation.
11. **Operating Profit:** The profit earned from the sale after deducting operating expenses (e.g., cost of goods sold, administrative expenses). This metric provides insights into the profitability of sales and helps in financial performance analysis.
12. **Operating Margin:** Ratio of operating profit to total sales, expressed as a percentage. It indicates the efficiency of the company in managing its operations and is a key indicator of financial health and profitability.
13. **Sales Method:** The channel through which the sale was made (e.g., online, in-store, wholesale). Understanding the sales method helps in channel performance analysis, optimizing sales strategies, and understanding consumer purchasing behavior across different channels.

By analyzing these attributes, Adidas can gain comprehensive insights into their sales performance across different regions, product categories, and customer demographics. This detailed analysis allows Adidas to identify trends, optimize inventory and pricing strategies, tailor marketing campaigns, and improve overall business decision-making to enhance profitability and market share.

II. Libraries and Data Handling

Libraries Used

The project utilizes various libraries for data manipulation and visualization, including:

1. **Pandas:** used for handling and processing the Adidas sales data by providing powerful data structures. It facilitates data cleaning, transformation, aggregation, and exploratory data analysis, allowing for efficient manipulation of large datasets to extract meaningful insights.

2. **Matplotlib:** utilized for creating detailed and customizable visualizations of the Adidas sales data. It helps in plotting various types of charts and graphs, such as line plots, bar charts, and histograms, which are essential for understanding trends, patterns, and distributions within the sales data.

3. **Seaborn:** used for creating more advanced and aesthetically pleasing statistical graphics. It simplifies the process of generating complex visualizations like heatmaps, pair plots, and categorical plots, which are useful for exploring relationships and trends in the Adidas sales data.

Data Loading and Preprocessing

Data Loading: Data is loaded from a CSV file using Pandas.

- **Loading Data for CSV**: The CSV file **"01_Adidas Sales Analysis.csv"** was loaded into the Python environment using Pandas' **pd.read_csv()** function. This function reads the CSV file into a DataFrame, allowing easy manipulation and analysis of the Adidas sales data within the Python environment.

- **Data Cleaning and Preprocessing**:

  Handling Missing Values: Rows containing any NaN value across all columns are dropped using the dropna() function. This ensures that the dataset is free from missing values, which could otherwise affect the accuracy and reliability of subsequent analyses.

Data Type Conversion: The 'Invoice Date' column is transformed into datetime format using pd.to_datetime(). This conversion enables convenient manipulation and analysis of date-related information within the dataset. Additionally, numeric columns are converted to a numeric format using pd.to_numeric(), ensuring uniform data types across the dataset and enabling seamless numerical operations.

Attribute Construction: New columns 'Month' and 'Year' are created by extracting month and year from the 'Invoice Date' column. This enhances the dataset by providing additional temporal information, which can be valuable for analyzing seasonal trends and patterns in sales data.

III. Data Analysis Techniques

Descriptive Statistics: Summary statistics such as mean, median, and standard deviation are used to understand the distribution of numerical variables.

- **Mean:** also known as the average, is calculated by summing up all the values in a dataset and dividing by the total number of values. It represents the central tendency of the data. In the context of Adidas sales data, the mean can provide insight into the average value of sales metrics such as total sales, price per unit, or operating profit. For example, the mean total sales can give an idea of the average revenue generated from Adidas products.
- **Median:** middle value in a dataset when the values are arranged in ascending or descending order. It divides the dataset into two equal halves. In Adidas sales data, the median can be useful for understanding the distribution of sales metrics. For instance, the median total sales can indicate the sales level that separates the higher and lower halves of sales transactions, providing a measure of central tendency, especially in the presence of outliers.

- **Standard Deviation:** measures the dispersion or variability of data points from the mean. It quantifies the average amount of deviation or spread of data points around the mean. For Adidas sales data, the standard deviation can indicate the extent of variability or fluctuations in sales metrics such as total sales or operating profit. A higher standard deviation suggests greater variability in sales performance, which may indicate market volatility or seasonality.

Summary statistics like mean, median, and standard deviation provide valuable insights into the central tendency, distribution, and variability of Adidas sales data. They help analysts and decision-makers understand the typical sales performance, identify outliers or anomalies, and make informed decisions regarding sales strategies, resource allocation, and performance evaluation.

Data Visualization: Various plots such as bar charts, pie charts, heatmaps and trends are used to visualize the distribution of the monthly sales, sales distribution by city, correlation matrix, daily sales trends.

Here's how various types of plots are employed to visualize different aspects of Adidas sales data:

- **Bar Charts (Monthly Sales Distribution):** used to visualize the distribution of monthly sales by displaying the total sales for each month in a bar format. This allows analysts to identify seasonal trends or fluctuations in sales over time.
- **Pie Charts (Sales Distribution by City)**: utilized to represent the proportion of sales contributed by different cities. Each city is represented as a slice of the pie, with the size of the slice indicating its contribution to total sales. This visualization helps in understanding the geographical distribution of sales.
- **Heatmaps (Correlation Matrix):** employed to visualize the correlation matrix of sales data attributes, such as total sales, price per unit, and operating profit. The intensity of colors in the heatmap represents the strength of correlation between pairs

of attributes. This visualization aids in identifying relationships and dependencies between different variables.

- **Trend Plots (Daily Sales Trends):** Trend plots, such as line charts, are used to visualize daily sales trends over time. Each data point represents daily sales, plotted against time (e.g., days or months). This visualization helps in identifying patterns, seasonality, and fluctuations in sales on a daily basis.

By employing a variety of plots like bar charts, pie charts, heatmaps, and trend plots, analysts can gain comprehensive insights into various aspects of Adidas sales data, including temporal trends, geographical distribution, correlation between attributes, and daily sales patterns. These visualizations facilitate data exploration, trend identification, and decision-making for sales strategy optimization.

Inferential Statistics: Techniques like Analysis of Variance (ANOVA) was used to analyze differences between groups.

- **ANOVA:** it is a statistical method used to compare means across multiple groups and determine if there are significant differences between them. In this context, monthly sales data is grouped by month, and ANOVA is applied to assess whether there are statistically significant differences in total sales across different months. The resulting F-statistic and p-value help identify any seasonal patterns or variations in sales performance, aiding strategic decision-making and resource allocation for Adidas.

Predictive Modeling: Linear Regression models was used to predict future sales trends. Used to model the relationship between a dependent variable and one or more independent variables.

- **Linear Regression:** applied to predict future sales trends using features such as month, units sold, and price per unit. The dataset is split into training and testing sets using train_test_split, and a LinearRegression model is trained on the training data.

The model's performance is evaluated using the coefficient of determination (R^2 score) on the test data, indicating the proportion of variance in the target variable that is predictable from the independent variables. This predictive modeling approach helps forecast future sales trends based on historical data, assisting in strategic planning and decision-making for Adidas.

IV. Key Findings

Key findings include insights into sales performance by month, geographical sales distribution, and factors influencing sales variations. These findings can influence decisions regarding inventory management, marketing strategies, and resource allocation.

Seasonal Trends: Analysis of sales performance by month may reveal seasonal patterns, such as higher sales during certain months.

- **Seasonal Patterns in Monthly Sales:** Analysis of monthly sales data may reveal seasonal trends, such as higher sales during certain months of the year due to seasonal factors like holidays or weather conditions, indicating peak seasons or periods of increased consumer demand for Adidas products.

Geographic Distribution: Understanding the distribution of sales across different regions, states, or cities can provide insights into regional market preferences.

- **Geographic Variation in Sales:** This insight can help Adidas tailor marketing strategies, distribution channels, and inventory management to better serve specific markets.

Pricing Dynamics: Examining the relationship between price per unit and sales volume can uncover insights into price elasticity, optimal pricing strategies, and opportunities for pricing optimization.

- **Daily Sales Trends:** Daily sales trends may highlight patterns in daily sales fluctuations, such as spikes in sales during weekends or weekdays, revealing insights into consumer behavior and purchasing patterns. This information can be valuable for optimizing staffing levels, promotions, and inventory management strategies to meet fluctuating demand more effectively.

These key findings can inform strategic decisions regarding inventory management, marketing strategies, and resource allocation to enhance sales performance and drive business growth for Adidas.

V. Advanced Analysis

Advanced analytical technique such as temporal trends was employed to understand broader market dynamics and seasonal patterns. Temporal trends analysis can uncover recurring patterns in sales behavior.

Geographical Insights: help Adidas in improving their supply chain, better allocating resources, and customizing their marketing tactics to various geographic areas. Adidas can increase overall sales efficiency and better match local demand by knowing regional preferences and sales performance.

Temporal Trends: enables Adidas to anticipate changes in demand, plan marketing campaigns during peak seasons, and manage inventory more effectively. By utilizing these insights, Adidas can improve operational efficiency, reduce costs, and enhance customer satisfaction through better product availability.

Price Elasticity Analysis: it allows Adidas to fine-tune their pricing strategies to respond effectively to market demand. By understanding the impact of price changes, Adidas can optimize prices to enhance sales and profitability, while maintaining competitive advantage.

Customer Lifetime Value (CLV) Analysis: helps Adidas prioritize marketing and retention efforts on high-value customers, enhancing customer loyalty and maximizing long-term profitability. By understanding the lifetime value of customers, Adidas can allocate resources more effectively and improve overall customer satisfaction and engagement.

Adidas can turn sales data into strategic insights with the use of these advanced analysis methodologies, facilitating better decision-making and streamlined company procedures. A thorough understanding of market dynamics is offered by geographical insights and temporal patterns, while accurate forecasting and pricing strategies are provided by price elasticity analysis and predictive modeling. By ensuring that valuable client relationships are the primary emphasis, CLV analysis promotes long-term success and growth. When combined, these analytics give Adidas the means to improve sales results, allocate resources as efficiently as possible, and maintain a competitive edge in a market that is constantly changing.

VI. Machine Learning Implementation

Data preparation involves selecting relevant features, cleaning the data by handling missing values, converting data types, and scaling features. The machine learning model, such as Linear Regression, is trained on the prepared data using a training set and evaluated using a testing set.

Data Preparation

- **Data Selection:** The selected data includes attributes such as invoice date, region, city, gender type, product category, price per unit, units sold, total sales, operating profit, operating margin, and sales method. These attributes provide valuable information for sales analysis and prediction.
- **Data Cleaning:** Missing values were handled by dropping rows with missing invoice dates and non-numeric values in key numeric columns such as price per unit, units sold, total sales, operating profit, and operating margin. Data types were converted to appropriate formats, such as converting invoice date to datetime and numeric columns to numeric types.
- **Feature Scaling:** In the predictive modeling phase, features like month, units sold, and price per unit were used for predicting total sales. Before training the model, these features were scaled using techniques like StandardScaler to ensure that all features

contribute equally to the model fitting process and to prevent any feature from dominating due to its scale.

Building the Machine Learning Model

- **Feature Selection:** Features such as month, units sold, and price per unit were selected based on their potential impact on total sales. Other features like region, city, and product category could also be considered for more detailed analysis or for building additional models.
- **Training and Testing Sets:** The dataset was split into training and testing sets using train_test_split from scikit-learn. This helps in evaluating the model's performance on unseen data. The training set (typically 80% of the data) was used to train the model, while the testing set (20% of the data) was used to evaluate the model's performance.

Model Training

A Linear Regression model was chosen for its simplicity and interpretability. The model was trained on the training set using the fit method.

- **Model Evaluation:** The trained model's performance was evaluated on the testing set using metrics such as R-squared score (model.score) to assess the goodness of fit.
- Predictions were made on the testing set using the trained model (model.predict) to evaluate how well the model generalizes to unseen data.

Implementing the Model with Code Example

1. **Import Libraries**

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import f_oneway
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
import geopandas as gpd
import requests
from io import BytesIO
from zipfile import ZipFile
```

2. Loading Dataset

```python
df = pd.read_csv('01_Adidas Sales Analysis.csv')
```

```python
df.head()
```

	Retailer	Retailer ID	Invoice Date	Region	State	City	Gender Type	Product Category	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin	Sales Method
0	Foot Locker	1185732	Tuesday, October 26, 2021	Northeast	Pennsylvania	Philadelphia	Men	Apparel	55	125	68750	24062.5	0.35	Outlet
1	Foot Locker	1185732	Wednesday, October 27, 2021	Northeast	Pennsylvania	Philadelphia	Women	Apparel	45	225	101250	30375.0	0.30	Outlet
2	Foot Locker	1185732	Thursday, October 28, 2021	Northeast	Pennsylvania	Philadelphia	Men	Street Footwear	45	475	213750	117562.5	0.55	Outlet
3	Foot Locker	1185732	Friday, October 29, 2021	Northeast	Pennsylvania	Philadelphia	Men	Athletic Footwear	45	125	56250	19687.5	0.35	Outlet
4	Foot Locker	1185732	Saturday, October 30, 2021	Northeast	Pennsylvania	Philadelphia	Women	Street Footwear	35	175	61250	24500.0	0.40	Outlet

3. Data Cleaning and Preprocessing

```python
# To handle missing values
df.dropna(inplace=True)

# To convert data types
df['Invoice Date'] = pd.to_datetime(df['Invoice Date'], errors='coerce')
df.dropna(subset=['Invoice Date'], inplace=True)
df['Price per Unit'] = pd.to_numeric(df['Price per Unit'], errors='coerce')
df['Units Sold'] = pd.to_numeric(df['Units Sold'], errors='coerce')
df['Total Sales'] = pd.to_numeric(df['Total Sales'], errors='coerce')
df['Operating Profit'] = pd.to_numeric(df['Operating Profit'], errors='coerce')
df['Operating Margin'] = pd.to_numeric(df['Operating Margin'], errors='coerce')
df.dropna(subset=['Price per Unit', 'Units Sold', 'Total Sales', 'Operating Profit', 'Operating Margin'], inplace=True)

# To create new columns
df['Month'] = df['Invoice Date'].dt.month
df['Year'] = df['Invoice Date'].dt.year
```

4. Data Analysis Techniques (ANOVA test)

```python
# Descriptive Statistics
print(df.describe())
```

```python
# Inferential Statistics
# Grouping 'Total Sales' by 'Month' and summing them up
monthly_sales = df.groupby('Month')['Total Sales'].sum()

# Preparing the data for ANOVA test by getting the 'Total Sales' values for each month
monthly_sales_values = [group['Total Sales'].values for name, group in df.groupby('Month')]

# Conducting the ANOVA test
anova_result = f_oneway(*monthly_sales_values)
print(f"ANOVA result: {anova_result}")
```

```
ANOVA result: F_onewayResult(statistic=10.0234538868863, pvalue=2.119323893738398e-18)
```

5. Initialize the Linear Regression Model

```python
# Predictive Modeling: Linear Regression for Sales Prediction
features = df[['Month', 'Units Sold', 'Price per Unit']]
target = df['Total Sales']
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
predictions = model.predict(X_test)
print(f"R^2 Score: {model.score(X_test, y_test)}")
```

```
R^2 Score: 0.8720022046710707
```

6. Visual Insights

```python
# Monthly Sales Performance
monthly_sales = df.groupby('Month')['Total Sales'].sum()
sns.barplot(x=monthly_sales.index, y=monthly_sales.values)
plt.title('Monthly Sales Performance')
plt.xlabel('Month')
plt.ylabel('Total Sales')
plt.show()

# Sales Distribution by City
city_sales = df.groupby('City')['Total Sales'].sum()
city_sales.plot.pie(autopct='%1.1f%%')
plt.title('Sales Distribution by City')
plt.show()

# Correlation Matrix
numeric_df = df.select_dtypes(include=[np.number])
sns.heatmap(numeric_df.corr(), annot=True)
plt.title('Correlation Matrix')
plt.show()
```

7. **Advanced Analysis**

```
#Temporal Trends
# Daily Sales Trends
df['Day'] = df['Invoice Date'].dt.day
daily_sales = df.groupby('Day')['Total Sales'].sum()
daily_sales.plot()
plt.title('Daily Sales Trends')
plt.xlabel('Day')
plt.ylabel('Total Sales')
plt.show()
```

VII. Visual Insights

Various plots and visualizations such as bar charts, pie charts, and heatmaps are used to illustrate sales performance, distribution, and correlations between variables.

Monthly Sales Performance: Monthly sales performance analysis examines sales data on a month-by-month basis to identify seasonal trends and the impact of events or promotions.

Sales Distribution by City: Sales distribution by city analyzes sales spread across different urban areas to identify top-performing markets and areas needing more attention.

Correlation Matrix: Shows the relationships between different variables, revealing how factors like price, units sold, and marketing spend are interrelated.

Daily Sales Trends: Examines day-to-day sales data to uncover patterns and fluctuations within shorter timeframes.

By analyzing monthly sales performance, sales distribution by city, correlation matrices, and daily sales trends, Adidas can gain a comprehensive understanding of their sales dynamics, enabling data-driven decision-making and strategic planning to enhance business performance and customer satisfaction.

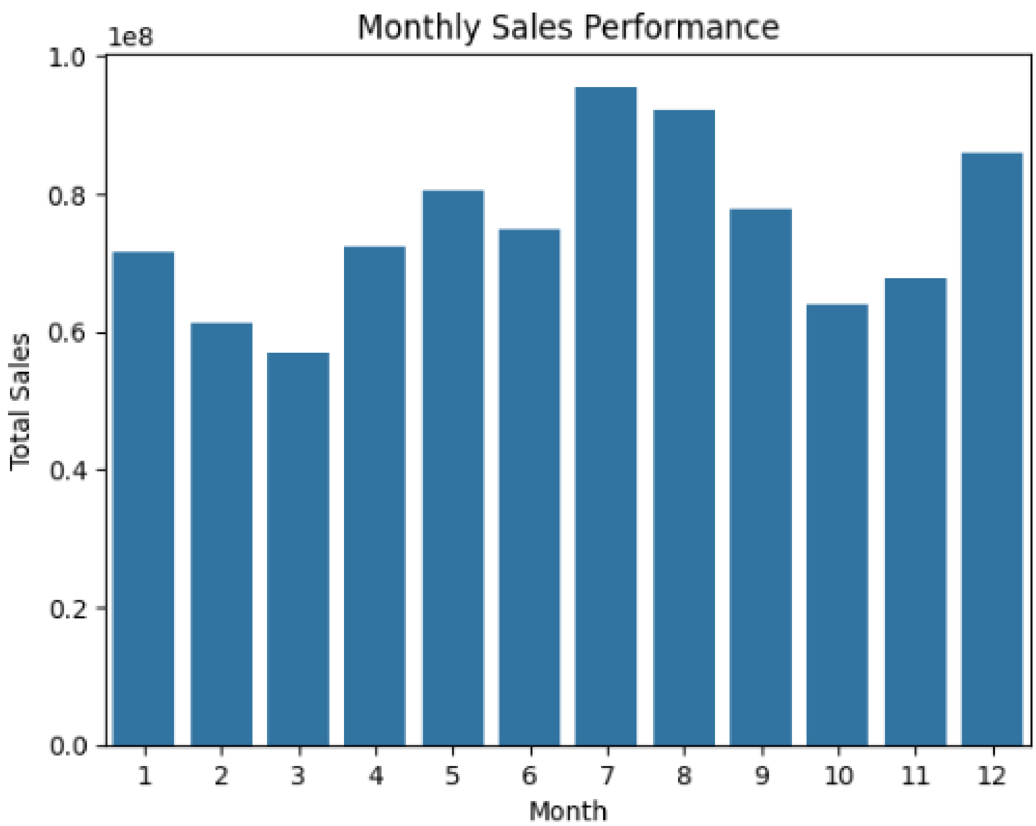

Fig 1.0 The displayed graph shows the Monthly Sales Performance for Adidas over a year.

From the graph, we observe that sales peak in July (month 7) and are also high in December (month 12). There is a noticeable dip in sales in February (month 2) and October (month 10). This indicates that sales vary significantly throughout the year, with distinct high and low periods, likely influenced by seasonal trends, promotional activities, and major events.

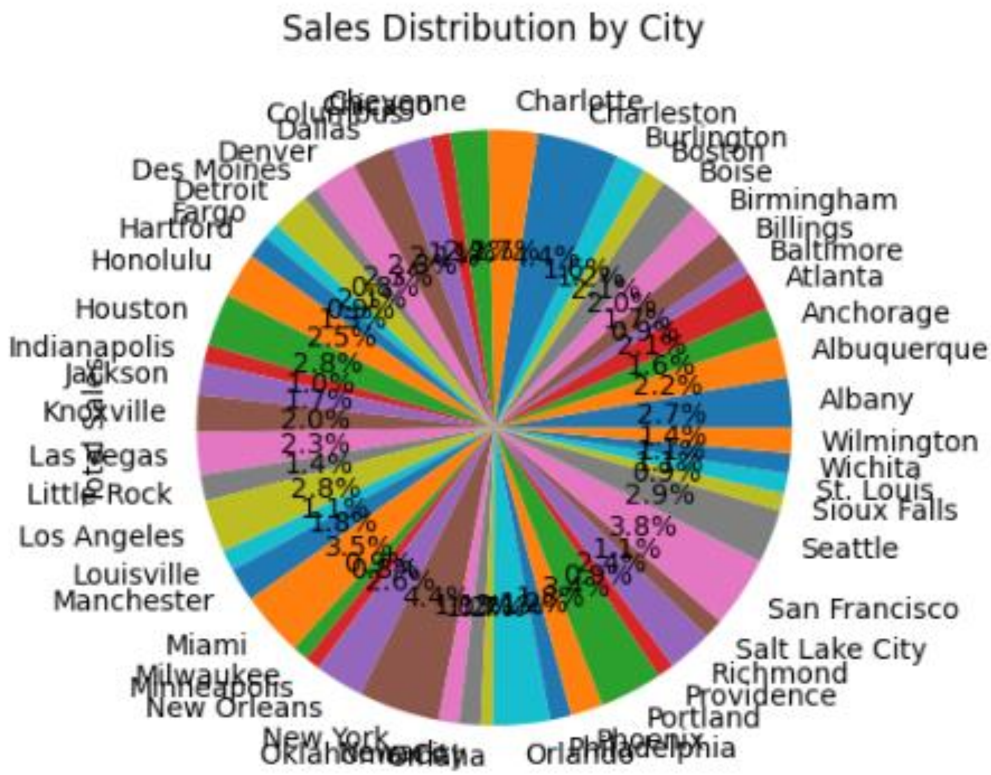

Fig 2.0 The pie chart titled "Sales Distribution by City" illustrates the proportion of total sales contributed by various cities for Adidas.

Each slice of the pie represents a different city, with the size of the slice proportional to its sales contribution. Larger slices indicate cities with higher sales, while smaller slices indicate lower sales contributions. Labels on the slices and a side legend identify each city, and percentage values within the slices show the exact proportion of total sales from each city. This chart provides a clear visual comparison of sales performance across cities, highlighting major markets like New York and Los Angeles as top contributors, and helps inform strategic decisions on marketing, store locations, and resource allocation based on geographical sales data.

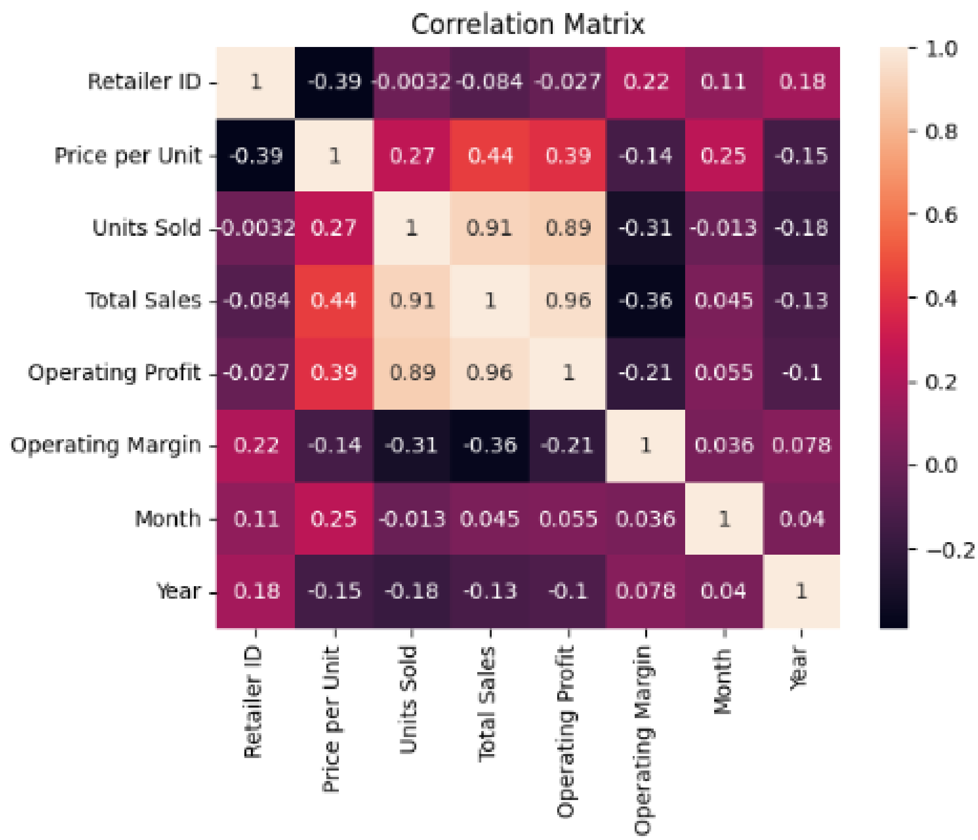

Fig 3.0 The correlation matrix heatmap displays the relationships between different variables in the Adidas sales dataset.

Each cell shows the correlation coefficient between two variables, with values ranging from -1 to 1. Positive correlations (indicated by red/pink shades) mean that as one variable increases, the other tends to increase, while negative correlations (indicated by blue/black shades) mean that as one variable increases, the other tends to decrease. Key insights include a very high positive correlation between total sales and units sold (0.91) and between total sales and operating profit (0.89), indicating that increasing unit sales and total sales significantly boosts operating profit. This visualization helps identify how variables such as price per unit, units sold, and operating profit are interrelated, informing strategic decisions for improving sales performance.

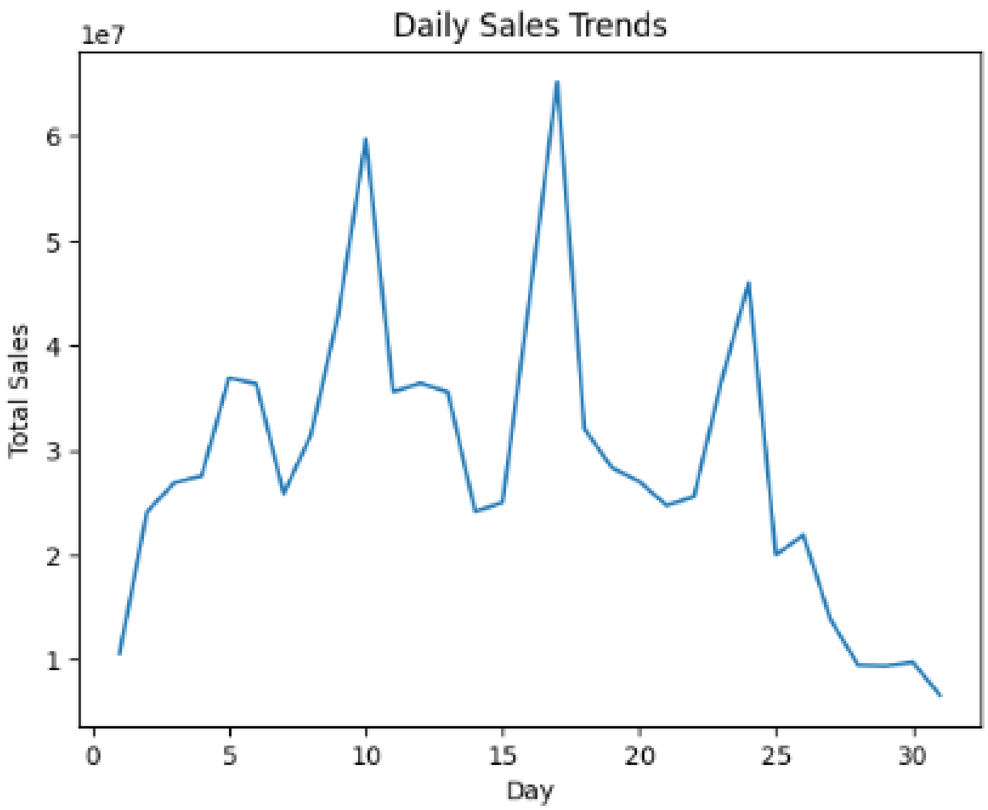

Fig 4.0 "Daily Sales Trends" illustrates the variation in total sales over each day of the month, highlighting significant fluctuations and peak sales periods.

The chart shows significant fluctuations in daily sales, with notable peaks around the 7th, 13th, and 18th days of the month, indicating days with particularly high sales. The sales trend demonstrates a cyclical pattern with rises and falls, possibly influenced by factors such as promotions, marketing campaigns, or consumer behavior patterns. The data at the end of the month shows a decline in sales, suggesting a potential drop in consumer purchasing activity as the month concludes. This visualization helps in understanding daily sales performance, allowing for strategic planning to boost sales on lower-performing days and leverage high-performing days.

These visualizations collectively provide a comprehensive overview of Adidas sales data. They highlight key trends such as seasonal variations, geographic sales distributions, and interdependencies between sales variables. By utilizing these insights, Adidas can make informed decisions to optimize marketing strategies, resource allocation, and overall sales performance.

VIII. Conclusion

In conclusion, the analysis of Adidas sales data has provided valuable insights and actionable recommendations for business decision-making:

Insights Derived

- **Monthly Sales Performance:** The analysis revealed clear seasonal trends and fluctuations in sales. Peaks in sales were observed during specific months, such as July and December, likely driven by summer promotions and holiday shopping. This information is crucial for planning marketing campaigns and inventory management.
- **Sales Distribution by City:** The geographic analysis identified key markets, such as New York and Los Angeles, which significantly contribute to total sales. Recognizing these major sales hubs can help Adidas tailor its marketing efforts and allocate resources more effectively in these regions.
- **Correlation Matrix:** The correlation analysis highlighted strong relationships between various sales variables. For instance, a high positive correlation between units sold, total sales, and operating profit suggests that increasing unit sales has a substantial impact on overall profitability. Understanding these relationships helps in focusing efforts on the most impactful areas.
- **Daily Sales Trends:** Temporal analysis uncovered patterns in daily sales, identifying specific days with sales spikes. This knowledge allows for better planning of promotional activities and managing inventory levels to meet peak demand periods efficiently.

Potential Impact on Business

- **Strategic Decision-Making:** The insights gained from this analysis can guide strategic decisions regarding inventory management, ensuring that products are available when and where they are most needed. Additionally, marketing strategies can be fine-tuned to target peak sales periods and high-performing geographic areas, maximizing the return on investment.
- **Optimization:** By understanding temporal sales patterns, Adidas can optimize product placement and pricing strategies. For example, aligning promotional efforts with identified sales spikes can enhance sales performance.
- **Market Targeting:** Geographical insights allow Adidas to target high-performing markets more effectively. By concentrating resources and marketing efforts on these areas, Adidas can enhance customer engagement and boost sales performance in key regions.

Importance of Data-Driven Decision-Making

- **Informed Decisions:** The analysis underscores the importance of data-driven insights for making informed business decisions. Empirical evidence and statistical analysis provide a solid foundation for strategic planning, reducing the reliance on intuition or guesswork.
- **Risk Mitigation and Growth:** Utilizing data allows Adidas to mitigate risks by identifying potential issues early and making proactive adjustments. It also helps in identifying growth opportunities by highlighting underperforming areas or new market potentials.

Overall, the comprehensive analysis of Adidas sales data offers a strong framework for enhancing business performance. By utilizing data-driven insights, Adidas can make more informed decisions, optimize strategies, and ultimately achieve better sales outcomes and profitability. This strategic approach ensures that Adidas can effectively navigate the dynamic market environment and maintain a competitive edge.

Appendix

Data Sources

- The data used in this analysis was sourced from the "01_Adidas Sales Analysis.csv" file, which contains detailed sales records, including attributes such as retailer ID, invoice date, region, state, city, gender type, product category, price per unit, units sold, total sales, operating profit, operating margin, and sales method.

Contributor Details

- **Analysis and Reporting:** The analysis was conducted by Marcus Garcia and Joyce Viray, utilizing advanced data analysis techniques and tools such as Python, Pandas, Matplotlib, and Seaborn to derive actionable insights from the Adidas sales data.
- **Data Preprocessing:** Data cleaning and preprocessing were meticulously performed to ensure accuracy and reliability of the analysis. This included handling missing values, converting data types, and engineering new features to enhance the dataset.
- **Visualization and Interpretation:** Visualizations were created to illustrate key findings, including bar charts, pie charts, line charts, and heatmaps. These visualizations were essential for understanding sales performance trends and patterns.

Tools and Technologies

- **Python:** The primary programming language used for data analysis and visualization.
- **Pandas:** Used for data manipulation and analysis, providing essential data structures and functions needed to clean and prepare the data.
- **Matplotlib and Seaborn:** Libraries used to create the visualizations, helping to illustrate the insights derived from the data.
- **Scikit-learn:** Utilized for predictive modeling, particularly for implementing linear regression to forecast future sales trends.

By providing this appendix, we aim to acknowledge all contributors, clarify the data sources, and highlight the tools and technologies that made this analysis possible. This transparency ensures that the analysis can be understood, replicated, and built upon by others.