INTERNSHIP REPORT

Retail Mastery: Analytics and Visuals from Adidas Sales Data

R. Shreya – 21951A67E3
Y. Yagna Sankeerthana – 22955A3307
Suhani Bahuwar – 21951A3354
B. Shiva Kumar – 22951A67D0

An Internship Report

on

Retail Mastery: Analytics and Visuals from Adidas Sales Data

by

R. Shreya – 21951A67E3
Y. Yagna Sankeerthana – 22955A3307
Suhani Bahuwar – 21951A3354
B. Shiva Kumar – 22951A67D0

INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad – 500 043, Telangana

May - June, 2024

DECLARATION

I certify that

a. The work contained in this report is original and has been done by me under the guidance of my supervisor (s).

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute for preparing the report.

d. I have conformed to the norms and guidelines given in the Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the report and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Place: Hyderabad   **Signature of the Student**

Date: **Roll No.**

CERTIFICATE

APPROVAL SHEET

ABSTRACT

Keywords: Data Analytics, Retail Optimization, Sales Metrics, Market Dynamics, Interpretability, Business Intelligence

In the highly competitive retail landscape, leveraging data analytics and visualization is crucial for informed decision-making and strategic planning. This report delves into Adidas sales data to demonstrate the power of analytics in optimizing retail operations, enhancing customer experience, and driving sales growth. Through the application of advanced data analytics techniques, we systematically analyze key sales metrics, identify trends, and uncover actionable insights. Our methodology encompasses data cleaning, exploratory data analysis, and visualization, aimed at providing a clear and insightful picture of Adidas's sales performance across various dimensions such as geography, product categories, and time periods. The analysis reveals significant findings, including seasonal sales patterns, regional preferences, and high-performing product lines, which offer a granular understanding of consumer behavior and market dynamics. We employ various visualization tools, including time-series plots, heat maps, and bar charts, to render complex data more accessible and interpretable for stakeholders. These visualizations not only highlight critical performance indicators but also facilitate scenario analysis and forecasting. The insights gleaned from the Adidas sales data enable the formulation of data-driven strategies to optimize inventory management, tailor marketing campaigns, and enhance customer engagement. By transforming raw sales data into actionable business intelligence, this report underscores the significance of data analytics and visualization in achieving retail mastery and sustaining a competitive advantage in the dynamic retail sector.

CONTENTS

Chapter 1: Introduction

In today's rapidly evolving retail landscape, the strategic use of data analytics and visualization has become essential for gaining a competitive edge and making informed business decisions. This report delves into the transformative potential of data analytics in the retail sector. By meticulously analyzing Adidas's sales data, we aim to uncover deep insights into sales trends, consumer preferences, and market dynamics that can drive impactful decision-making and operational efficiency.

Our comprehensive approach includes data cleaning, exploratory analysis, and the use of advanced visualization techniques to convert raw sales data into meaningful and actionable intelligence. Through this detailed analysis, we reveal key findings such as seasonal sales patterns, regional preferences, and high-performing product lines, providing a granular understanding of consumer behavior and market trends. These insights are crucial for optimizing inventory management, crafting targeted marketing campaigns, and enhancing customer engagement.

Furthermore, the report illustrates how various visualization tools, including time-series plots, heat maps, and bar charts, can make complex data more accessible and interpretable for stakeholders. By highlighting critical performance indicators and facilitating scenario analysis and forecasting, these visualizations enable data-driven strategies that ensure operational excellence and competitive advantage. Ultimately, this study underscores the importance of integrating data analytics and visualization into retail strategies to achieve mastery and sustain success in a dynamic and competitive market.

1.1 Background

Python, on the other hand, is a versatile programming language favored in the data science community for its simplicity and powerful libraries. Libraries such as Pandas (for data manipulation), Matplotlib and Seaborn (for data visualization), and Scikit-learn (for machine learning) make Python indispensable for data analysts. Python's ability to handle large datasets, perform complex calculations, and generate sophisticated visualizations makes it ideal for deep data analysis. When used together, Power BI and Python enable businesses to fully exploit their data's potential. Power BI's intuitive interface and powerful visualization capabilities complement Python's advanced analytical tools and flexibility. Utilizing these tools, businesses can perform thorough data analysis, create insightful visualizations, and make informed decisions that enhance operational efficiency and strategic planning.

- **Overview of Data Analytics and Visualization:**

  In the contemporary retail sector, data analytics and visualization are critical for maintaining competitiveness and making informed decisions. Data analytics involves examining raw data computationally to uncover patterns, trends, and insights that would otherwise remain hidden. By analyzing various data points like sales figures, customer behavior, and market trends, businesses can derive strategic insights that drive growth and operational efficiency. Visualization, meanwhile, transforms this complex data into graphical formats such as charts, graphs, and maps.

- **Data Visualization: Definition, Advantages, and Disadvantages:**

  Data visualization refers to the process of converting data into visual formats to facilitate easier interpretation and analysis. This includes creating charts, graphs, maps, and other visual tools that make data more comprehensible. The main advantage of data visualization is its ability to present complex data in an easily understandable way, allowing for the rapid identification of patterns, trends, and anomalies. It also improves communication of insights across various organizational levels, enhancing collaboration and decision-making. However, data visualization has drawbacks as well. Poorly designed visuals can mislead or misrepresent the data, resulting in incorrect conclusions. Additionally, crafting effective visualizations often requires expertise and can be time-intensive.

- **Importance of Power BI and Python in Data Analysis:**

  Power BI and Python are essential tools in the field of data analysis and visualization. Power BI is a business analytics service from Microsoft that provides interactive visualizations and business intelligence capabilities. Its user-friendly interface allows end-users to create their own reports and dashboards, and its robust data connectivity options facilitate the integration of a wide range of data sources. This makes Power BI a valuable tool for creating comprehensive, interactive visualizations that support decision-making processes.

1.2 Getting Started

A) Setting Up Power BI:

❖ Downloading Power BI Desktop:

      a) Visit the [Power BI website](#) and download Power BI Desktop.
      b) Follow installation instructions.

❖ **Connecting to Data Sources:**

      a) Open Power BI Desktop.
      b) "Get Data" to connect to data sources (Excel).
      c) Select the data source and load your data.

❖ **Creating Visualizations:**

      a) Drag and drop fields to create visualizations.
      b) Customize visual properties (colors, labels, legends).
      c) Add slicers and filters for interactivity.

❖ **Publishing and Sharing Reports:**

      a) Click "Publish" to publish the report to Power BI Service.
      b) Set permissions to share with others.

B) Installing Python and Necessary Libraries

❖ **Installing Python:**

      a) Go to the [Python download page](#).
      b) Download and run the installer.
      c) Ensure "Add Python to PATH" is checked.
      d) Choose "Install Now".
      e) Or can also download anaconda and use Jupyter python Kernal.
      f) Without any python idles shells you can also use Google Colab to execute the code.

❖ Installing Necessary Libraries:

      a) Open terminal or command prompt.
      b) Install libraries and import the libraries to verify
      c) Execute the code

❖ Libraries used for this project are:

a) **NumPy:** NumPy (Numerical Python) is a fundamental library for numerical computations in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.

b) **Pandas:** Pandas (Panel data) is a powerful data manipulation and analysis library built on top of NumPy. It introduces data structures like Data Frame and Series, which are designed to handle structured data intuitively and efficiently. Pandas provides a wide range of functions for data cleaning, transformation, aggregation, and visualization.

c) **Matplotlib:** Matplotlib (derived from "MATLAB" style plotting) is a comprehensive library for creating static, animated, and interactive visualizations in Python. It allows users to generate a wide range of plots, including line charts, bar charts, scatter plots, histograms, and more.

d) **Seaborn:** Seaborn (named after the seaborn seabird genus) is a statistical data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Seaborn simplifies the process of creating complex visualizations like heatmaps, time series, and violin plots.

e) **Plotly:** Plotly (a combination of "plot" and "ly" from the domain name used for the company's website (plot.ly)) is an open-source graphing library that enables the creation of interactive and visually appealing plots. It supports a wide variety of chart types, including line plots, scatter plots, bar charts, and 3D plots.

1.3 Requirements

➢ **Languages version:**
  a) Python: version 3.11.9 -April 2,2024
  b) PowerBI: version 2.117.984.0. Released: May 28,2023

➢ **Library version:**

  a) NumPy: version 1.26.4. Released: Feb 5, 2024

  b) Pandas: version 2.2. 2

  c) Matplotlib: version3.9.0

  d) Seaborn: v0. 12.0 (September 2022)

  e) Plotly: version 4.9

  f) Scikit-learn: version 1.5.0

➢ **Operating System:** windows 10

1.4 Data Storytelling

- ➢ In the dynamic world of retail, the ability to harness data effectively can be a game changer for business operations.
- ➢ This story centres on a mid-sized retail chain that embarked on a comprehensive retail management analysis and visualization project to address challenges in inventory management, customer retention, and market trend identification.
- ➢ The first phase involved extensive data collection from multiple sources, including sales transactions from point-of-sale systems, customer information from loyalty programs, inventory levels from warehouse management systems, and market trends from social media and online reviews. Integrating this data into a cohesive dataset was crucial, achieved through robust ETL (Extract, Transform, Load) processes to ensure consistency and accuracy.
- ➢ The final phase focused on transforming these analytical insights into intuitive visualizations using tools like Tableau and Power BI.
- ➢ Sales performance dashboards highlighted top-selling products, sales trends, and regional comparisons.
- ➢ Customer insights dashboards illustrated customer segmentation, purchase patterns, and loyalty program effectiveness. Inventory management dashboards showcased current stock levels, turnover rates, and predictive inventory needs.
- ➢ These visualizations made complex data easily interpretable, enabling swift and informed decision-making. For instance, a heatmap of regional sales performance quickly identified underperforming areas, prompting targeted marketing efforts.
- ➢ The project's impact was profound: inventory costs were reduced by 20% through optimized stock levels, customer retention improved by 15% with personalized marketing, and proactive market positioning was achieved by identifying and acting on emerging trends.

Chapter 2: Project

Description

This retail management analysis and visualization project exemplifies how data can transform business operations, turning raw data into strategic insights that drive better decision-making and improve overall business performance. It highlights the essential role of data in navigating the complexities of the retail industry and maintaining a competitive edge in an ever-evolving market. The project's goals were clear: optimize inventory to reduce costs and prevent stockouts, enhance customer retention through personalized marketing, and stay ahead of market trends.
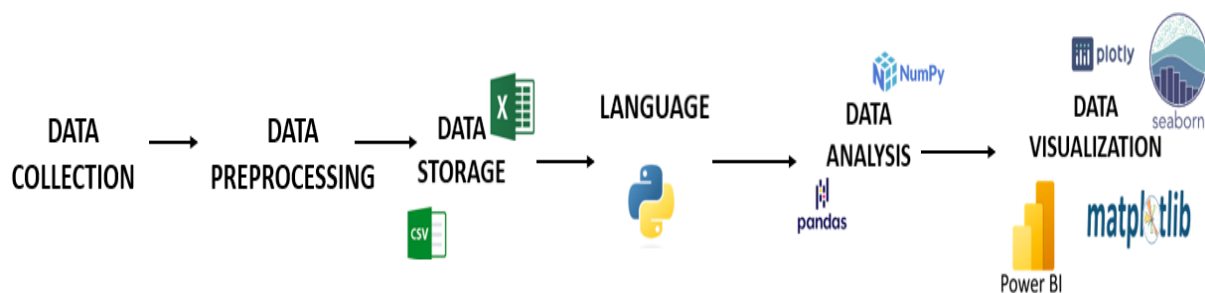

Figure 1: Flowchart of the project

2.1 Data Preparation

> ➤ The data preparation process for retail management analysis and data visualization using Python and Power BI involves several crucial steps to ensure data accuracy, relevance, and clarity.

> ➤ First, data collection is conducted, aggregating information from various sources such as sales transactions, inventory records, customer feedback, and market trends.

> ➤ Once collected, the data undergoes cleaning to rectify any inconsistencies, missing values, or errors, which is typically accomplished using Python libraries like Pandas and NumPy.

> ➤ This stage is essential for maintaining the integrity of the analysis. Following cleaning, data transformation is performed to convert raw data into a structured format suitable for analysis; this might include normalization, aggregation, and the creation of new calculated fields.

> ➤ Python's powerful data manipulation capabilities are extensively utilized here. After transformation, the data is loaded into Power BI, where it is organized into datasets and data models, enabling the creation of interactive and insightful visualizations.

- Power BI's robust visualization tools help in crafting dashboards and reports that provide comprehensive insights into retail performance, customer behavior, and market dynamics.

- Throughout this process, maintaining a focus on the business objectives ensures that the final visualizations and analyses are actionable and aligned with strategic goals.

- This integrated approach, leveraging Python's data handling prowess and Power BI's visualization strengths, facilitates effective retail management and decision-making.

- Data acquisition involves collecting and measuring information from diverse sources for analysis and decision-making.

- It is the first step in data preparation for analytics and visualization. Key sources include relational databases (such as MySQL, PostgreSQL, SQL Server), NoSQL databases (like MongoDB, Cassandra), web services and APIs (including RESTful APIs from social media platforms and financial data providers, as well as public APIs like OpenWeatherMap and Google Maps), and various file types (such as CSV, Excel, JSON, XML, and text files).

- Additionally, data can be gathered through web scraping with tools like Beautiful Soup and Scrapy, sensors and IoT devices that provide data from physical sensors, and manual data entry from surveys, forms, and direct input. After doing some research on different kinds of datasets, we found a dataset called Adidas Sales Data from Kaggle.

- In our project, we took data from direct source called Kaggle. In Kaggle website, we got a dataset on Adidas sales database as our project deals with retail management.

- URL - https://www.kaggle.com/datasets/vishwas199728/adidas-sales-data

- After collecting the data, we edit some part of it and continued to the next process

- After collection of data, we did preprocessing and removed unwanted data

Figure 2: Raw Data from the dataset collected for the project

2.2 Data Preprocessing

➢ In the Adidas sales database project, preprocessing involves a meticulous series of steps to ensure the data is clean and ready for analysis.

➢ The images illustrate critical preprocessing tasks. Initially, missing values, such as NA, NaN, NULL, and empty spaces, are addressed.

➢ This is done by iterating through object type columns and using the strip () method to remove extraneous spaces, followed by dropna () to remove rows containing NULL values, ensuring a clean dataset.

➢ This process is depicted in the first image. Subsequently, the info () function is employed to inspect the data frame, verifying that all columns contain non-null values and confirming the data types, as shown in the second image.

➢ This step confirms that the dataset consists of 9,648 entries across 13 columns, with no missing values remaining. This comprehensive preprocessing ensures that the data is accurate and reliable, forming a robust foundation for subsequent analysis and visualization.

STEP - 3 REMOVING IMPURE DATA

Impure data such as NA ,NaN , NULL , and empty spaces and values in the dataset

```
[ ]  # this code is used to remove missing values
     for col in adi.select_dtypes(include=['object']):
         adi[col] = adi[col].str.strip()
     adi.dropna(inplace=True)
     #dropna method removes the rows that contains NULL values.
```

Figure 3: Removing Null, empty spaces and unwanted data from the dataset

```
    #dataset information after removing missing values and white spaces
    adi.info()

    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 9648 entries, 0 to 9647
    Data columns (total 13 columns):
     #   Column            Non-Null Count  Dtype
    ---  ------            --------------  -----
     0   Retailer          9648 non-null   object
     1   Retailer ID       9648 non-null   int64
     2   Invoice Date      9648 non-null   object
     3   Region            9648 non-null   object
     4   State             9648 non-null   object
     5   City              9648 non-null   object
     6   Product           9648 non-null   object
     7   Price per Unit    9648 non-null   object
     8   Units Sold        9648 non-null   object
     9   Total Sales       9648 non-null   object
     10  Operating Profit  9648 non-null   object
     11  Operating Margin  9648 non-null   object
     12  Sales Method      9648 non-null   object
    dtypes: int64(1), object(12)
    memory usage: 980.0+ KB
```

From the above code we can understand that this data has no null and empty values

Figure 4: Information of the dataset after the data preprocessing

2.3 Data Analysis

➤ In this report, we present a comprehensive analysis of the Adidas sales database, focusing on uncovering trends, patterns, and insights that can drive informed decision-making. The raw sales data is meticulously prepared by addressing missing values, removing duplicates, and correcting inconsistencies. This process ensures the integrity and accuracy of the data, laying a solid foundation for subsequent analysis.

➤ We did descriptive statistics to summarize the data effectively. We calculated the key metrics such as total sales, average sales per transaction, and sales distribution across various regions and time periods are calculated. These statistics provide a snapshot of overall sales performance and highlight significant trends. For example, we observe seasonal variations in sales, peak sales periods, and regional differences in product popularity.

➤ In this project, Descriptive analysis provides insights on current sales performance, customer demographics, and inventory status. Predictive analysis uses time-series model to forecast the future sales and inventory needs, revealing patterns such as seasonal demand spikes.

➤ Prescriptive analysis also offers to show actionable recommendations for optimal stock levels and targeted marketing strategies. Customer lifetime value (CLV) analysis is another critical component of our report. This analysis focuses on understanding the long-term value of different customer segments by examining historical purchase data and customer interactions.

➤ Estimating the future value each customer brings to the company helps in identifying high-value customers, optimizing retention strategies, and improving targeted marketing efforts. Additionally, we conduct sentiment analysis on customer feedback, reviews, and social media interactions to gain insights into customer satisfaction and brand perception.

➤ By analyzing sentiments expressed in customer reviews and social media posts, we identify areas for improvement, address customer concerns, and enhance the overall customer experience.

➤ Variables of this project are:

- Adi – Variable containing the adidas sales dataset

- Columns of the dataset - Retailer, Retailer ID, Invoice Date, Region, State, City, Product, Price Per Unit, Units Sold, Total Sales, Operating Profit, Operating Margin, Sales Method

```
#returns number of unique val
adi.nunique(axis=0)

Retailer             6
Retailer ID          4
Invoice Date       724
Region               5
State               50
City                52
Product              6
Price per Unit      94
Units Sold         361
Total Sales       3138
Operating Profit  4187
Operating Margin    66
Sales Method         3
dtype: int64
```

Figure 5: Unique values in the dataset

2.4Data Visualization

Data visualization plays a crucial role in analyzing the Adidas sales database by transforming complex data into clear, actionable insights. Key visualizations include line charts for tracking sales trends over time, heatmaps for comparing regional sales performance, and bar charts for evaluating product category performance. Pie charts highlight customer segmentation, while stacked bar charts assess sales channel effectiveness. Scatter plots help identify high-value customers through Customer Lifetime Value (CLV) analysis, and forecasting graphs predict future sales trends for strategic planning. Additionally, word clouds and sentiment charts provide insights into customer feedback. These visualizations collectively enable Adidas to understand sales dynamics, optimize marketing strategies, and enhance decision-making processes across the organization.

2.4.1 Python

➢ Line Chart

This line chart displays the operating profits of Adidas across different states. The vertical axis represents the units sold by state, while the horizontal axis lists various states including Alabama, Hawaii, Massachusetts, New Mexico, South Dakota and many more. The chart shows significant fluctuations in sold goods, with peaks notably high in states like New Mexico and Hawaii, indicating these regions generate substantial profits compared to others.
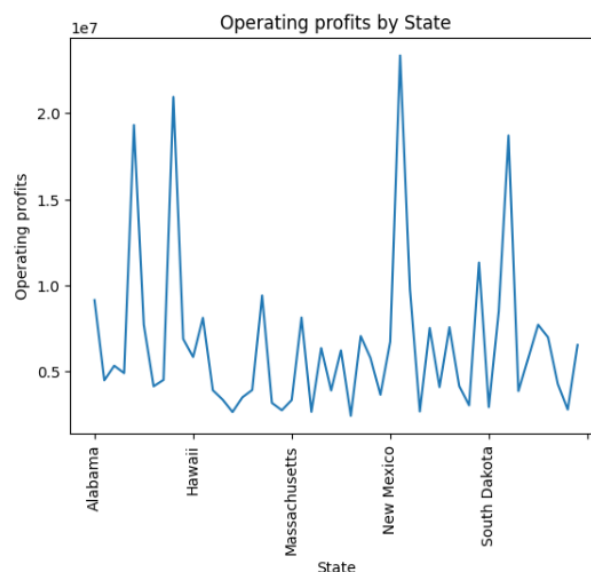

Figure 6: Operating Profits by State

➤ Histogram

This histogram shows the frequency distribution of sales values in the Adidas dataset. The x-axis represents the sales amounts, ranging from 0 to 800,000. The y-axis represents the frequency of these sales amounts. The histogram reveals that the majority of sales are concentrated at the lower end of the sales spectrum, with a significant number of sales falling between 0 and 100,000. As the sales amounts increase, the frequency of sales decreases sharply, indicating a right-skewed distribution.
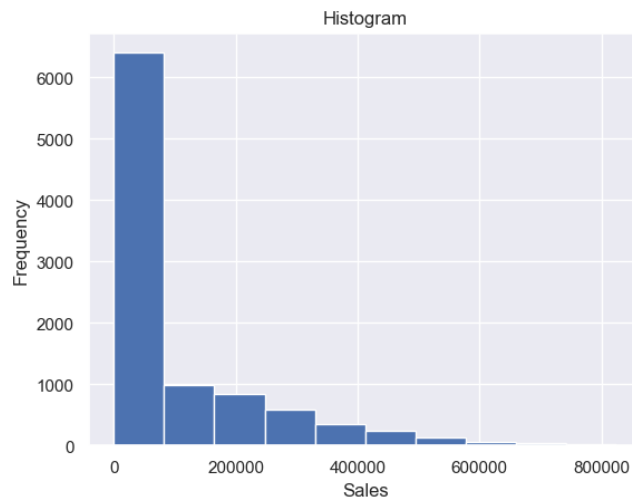

Figure 7: Histogram on sales and frequency

➤ Pie Chart

This pie chart illustrates the distribution of units sold by product category for Adidas. The categories include Men's Athletic Footwear, Men's Apparel, Women's Street Footwear, Women's Athletic Footwear, Women's Apparel, and Men's Street Footwear. The largest share of units sold belongs to Men's Street Footwear at 23.9%, followed by Men's Athletic Footwear at 17.6%, Women's Apparel at 17.5%, Women's Street Footwear at 15.8%, Women's Athletic Footwear at 12.8%, and Men's Apparel at 12.4%.
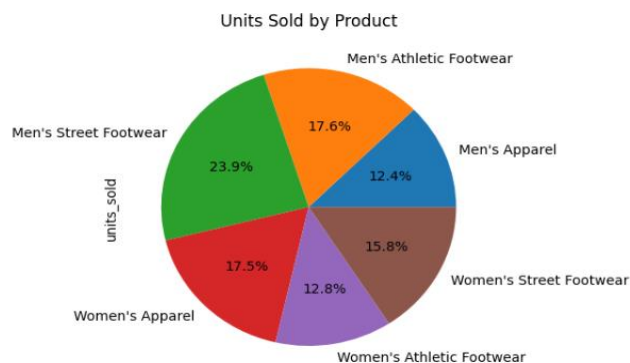

Figure 8: Units Sold by Product

➢ Bar Graph

This bar chart illustrates the total sales figures for Adidas across different states in the United States. The x-axis represents the states, while the y-axis indicates the total sales amount. States like California, New York, and Texas show significantly higher sales compared to others, with notable spikes indicating they are key markets for Adidas. This chart helps identify the top-performing regions in terms of sales volume.
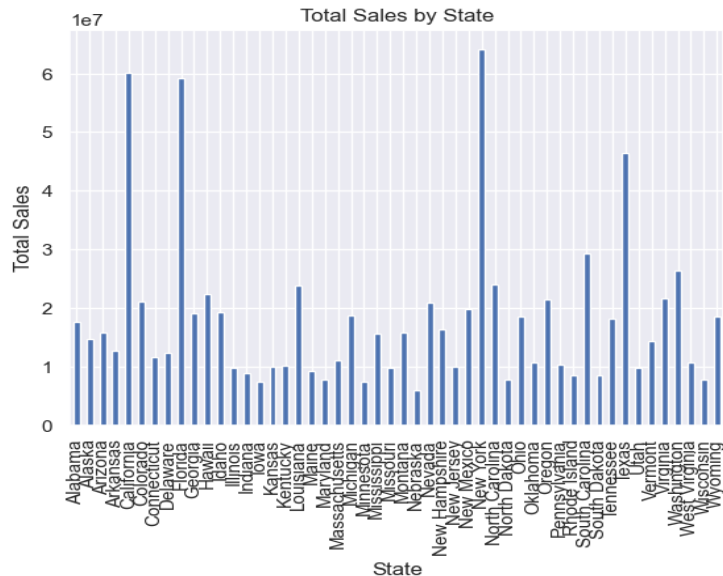

Figure 9: Bar Chart on Total Sales by State

➢ Scatter Plot

This plot illustrates the relationship between the units sold and the total sales in the Adidas dataset. The x-axis represents the number of units sold, and the y-axis represents the total sales amount. The plot shows a positive correlation between the units sold and the total sales, as expected. As the number of units sold increases, the total sales amount also increases. The data points are densely populated at the lower end of both axes, indicating that a large number of transactions involve relatively small quantities of units sold and corresponding total sales.
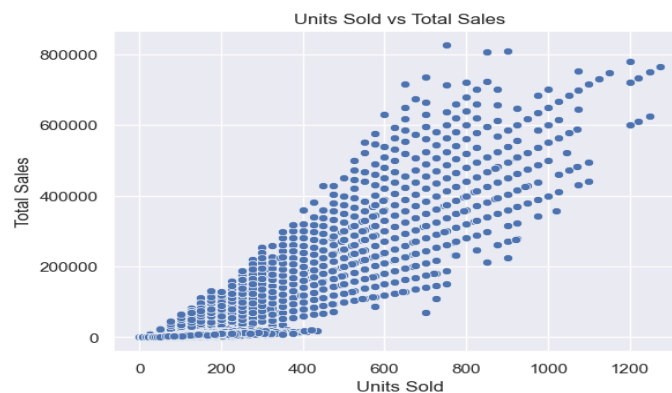

Figure 10: Scatter plot on units sold vs total sales

➢ Box Plot

This box plot provides a more detailed view of the sales distribution within each state. It shows the range, median, quartiles, and outliers of total sales data. This plot reveals variability in sales figures, with states like California, New York, and Texas again showing higher median sales and a wider range of sales values. The presence of numerous outliers indicates a high degree of variability in sales within these states, suggesting a mix of both high and low performing periods or regions within the states.
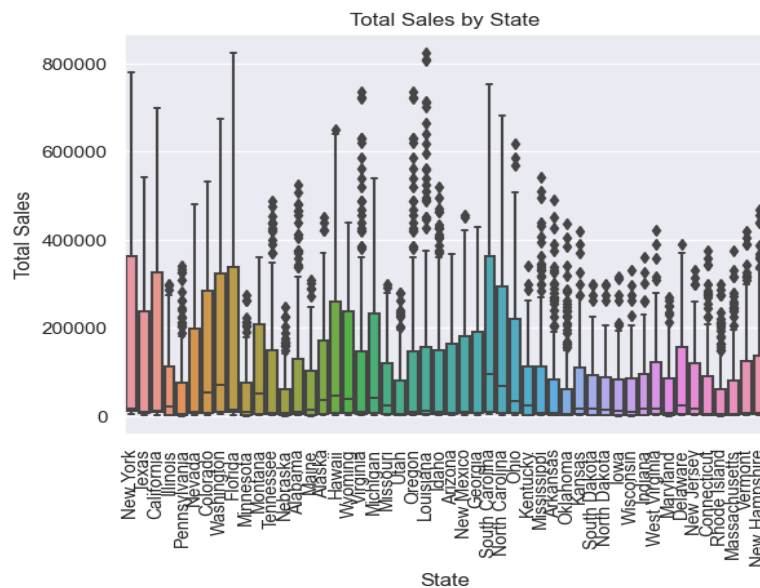

Figure 11: Box Plot on total sales by state

➢ Distribution Plot

This plot overlaid with a density plot, illustrates the distribution of units sold across all states. The x-axis shows the number of units sold, while the y-axis represents the density or frequency of sales. The distribution is right-skewed, indicating that while a large number of transactions involve relatively few units sold, there are fewer instances of high-volume sales. The peak around the lower end suggests that small to medium-sized orders are more common in Adidas' sales data.
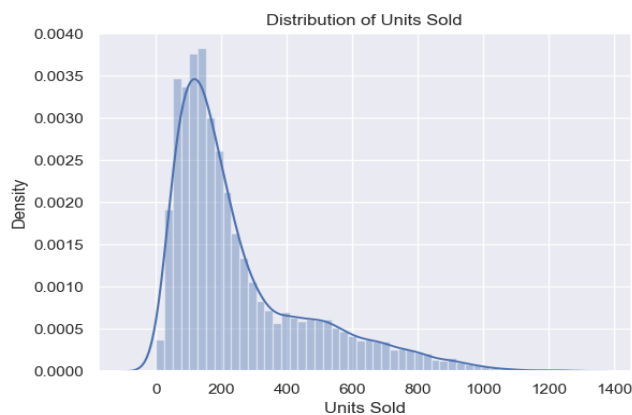

Figure 12: Distribution plot on density and units sold

➢ 3D Scatter Plot

In this 3D scatter plot, the three axes represent different key performance indicators: the x-axis denotes total sales, the y-axis signifies operating profit, and the z-axis reflects units sold. Each data point in the scatter plot represents a specific region, with its coordinates corresponding to the region's total sales, operating profit, and units sold. This graph identifies how different regions perform across these three metrics simultaneously. This visualization can help retail managers and analysts to quickly identify trends, anomalies, and potential areas for strategic focus.
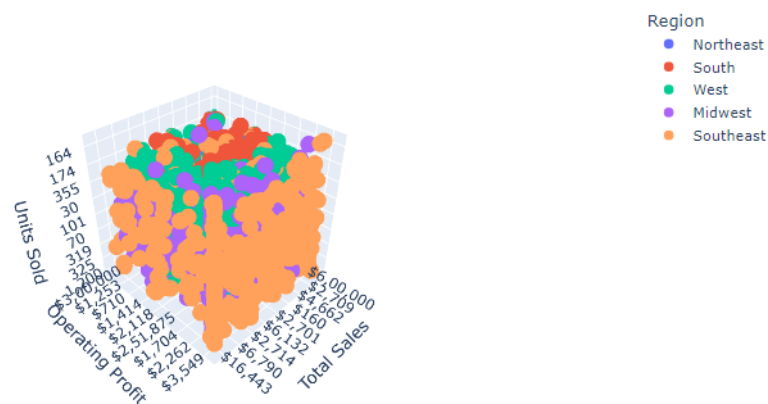

Figure 13: 3D Scatter plot using Plotly on Total sales, Operating profit and units sold by region

➢ Violin Plot

This plot generates a violin plot that provides a visual representation of the distribution of total sales across different regions, helping to uncover patterns and insights that are critical for effective retail management.
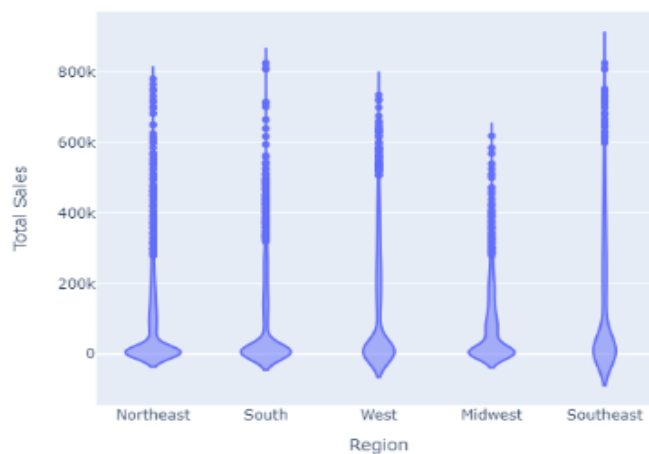

Figure 14: Violin Plot of total sales by region using Plotly

Scatter Plot

The scatter plot visualizes the relationship between total sales and operating profit across various regions, with each data point representing a unique sales entry. The plot is color-coded by region, including the Northeast, South, West, Midwest, and Southeast. This allows for easy comparison and analysis of regional performance. The x-axis represents total sales, while the y-axis indicates operating profit. Notably, the data points form distinct clusters at various operating profit levels, suggesting different tiers of profitability. This indicates that while higher total sales generally correlate with higher operating profits, other factors are influencing the profitability, creating some variability within the data. The distribution of points across regions appears relatively balanced, signifying that Adidas has a well-distributed sales presence across the United States. This plot is particularly useful for identifying high-performing regions and understanding the overall profitability landscape.
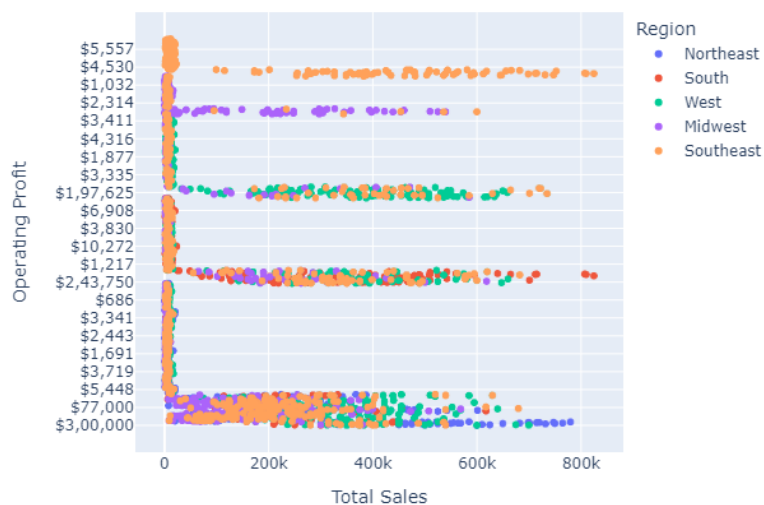

Figure 15: Scatter Plot of Total Sales vs. Operating Profit by Region using Plotly

2.4.2 Power BI

Dashboard

The below image is an elaborate Adidas US sales dashboard, providing a multi-faceted overview of sales performance across various dimensions. At the top, summary metrics display the total units sold (2M), total profit ($269M), total sales ($718M), and the count of products (6). Below these key figures, a time-series chart shows units sold by year, month, and day, highlighting sales trends over time. The dashboard also categorizes products into Men's Apparel, Men's Athletic Footwear, Men's Street Footwear, Women's Apparel, Women's Athletic Footwear, and Women's Street Footwear, allowing users to filter data accordingly.

Additional charts and maps offer insights into sales by retailer and product, sales by state, and sales methods (online, outlet, in-store). The top and bottom five retailers by profit are displayed, with Sports Direct and Foot Locker among the top performers, while Walmart and Amazon are in the bottom ranks.

A bar chart compares total sales and profit by product, showing Men's and Women's Street Footwear as top sellers. This comprehensive dashboard aids Adidas in analyzing sales performance, identifying top and bottom-performing segments, and making data-driven decisions to enhance their retail strategies.
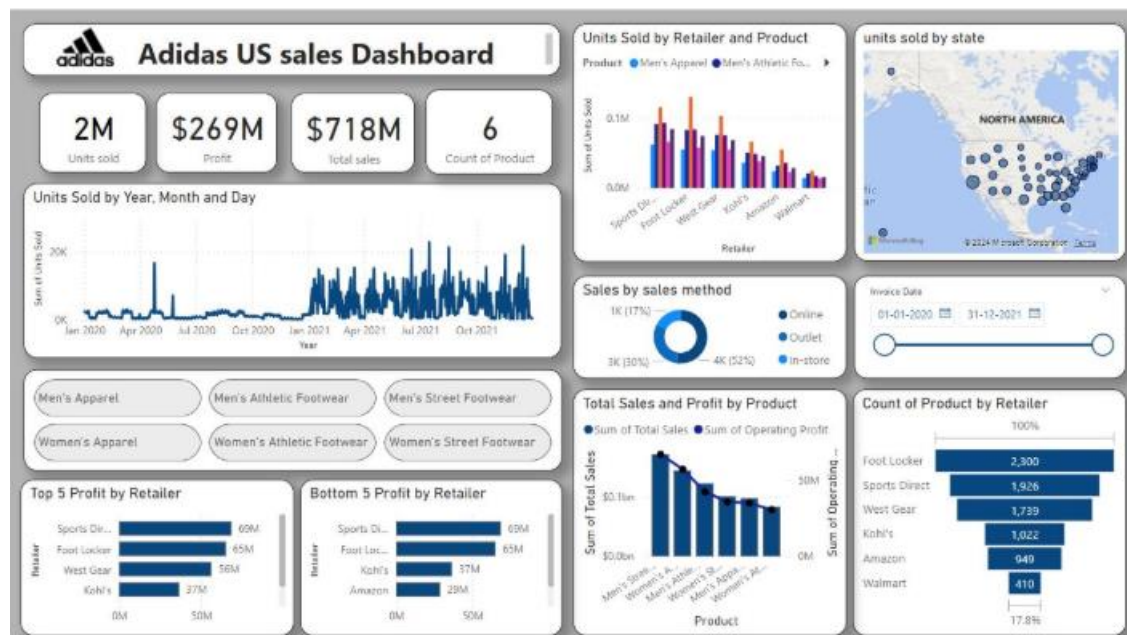

Figure 16: Power BI Dashboard on the Adidas Sales Dataset

➢ **Clustered Column Chart**

This chart provides a detailed breakdown of units sold across different retailers and product categories. The x-axis lists the retailers, including West Gear, Foot Locker, Sports Direct, Kohl's, Walmart, and Amazon, while the y-axis represents the sum of units sold. The bars are color-coded to differentiate between six product categories: Men's Apparel, Men's Athletic Footwear, Men's Street Footwear, Women's Apparel, Women's Athletic Footwear, and Women's Street Footwear. The chart reveals that Foot Locker and Sports Direct have the highest sales volumes, especially in Men's and Women's Street Footwear. West Gear also shows strong sales, particularly in Men's Street Footwear. In contrast, Kohl's, Walmart, and Amazon have lower overall sales volumes, with Amazon showing a notable preference for Women's Street Footwear. This visualization highlights the performance of different product categories across major retailers, providing valuable insights into consumer preferences and retailer strengths.
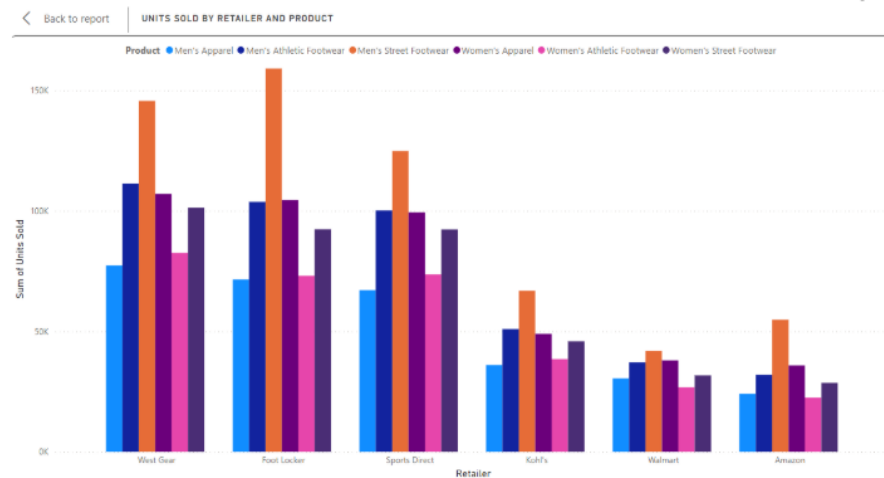
Figure 17: Clustered column chart on Units Sold by Retailer and Product

➢ Line and Stacked Column Chart

This visualization combines a bar chart and a line chart to show total sales and operating profit by product category. The x-axis lists the product categories, including Men's Street Footwear, Women's Apparel, Men's Athletic Footwear, Women's Street Footwear, Men's Apparel, and Women's Athletic Footwear. The y-axis on the left represents the sum of total sales, depicted by the bars, while the y-axis on the right represents the sum of operating profit, depicted by the line. This chart reveals that Men's Street Footwear generates the highest total sales, followed by Women's Apparel and Men's Athletic Footwear. Correspondingly, Men's Street Footwear also achieves the highest operating profit, indicating effective profitability management in this category. In contrast, Women's Athletic Footwear shows the lowest sales and operating profit. This visualization is essential for understanding the financial performance of each product category, allowing Adidas to make informed decisions about product focus and investment strategies.
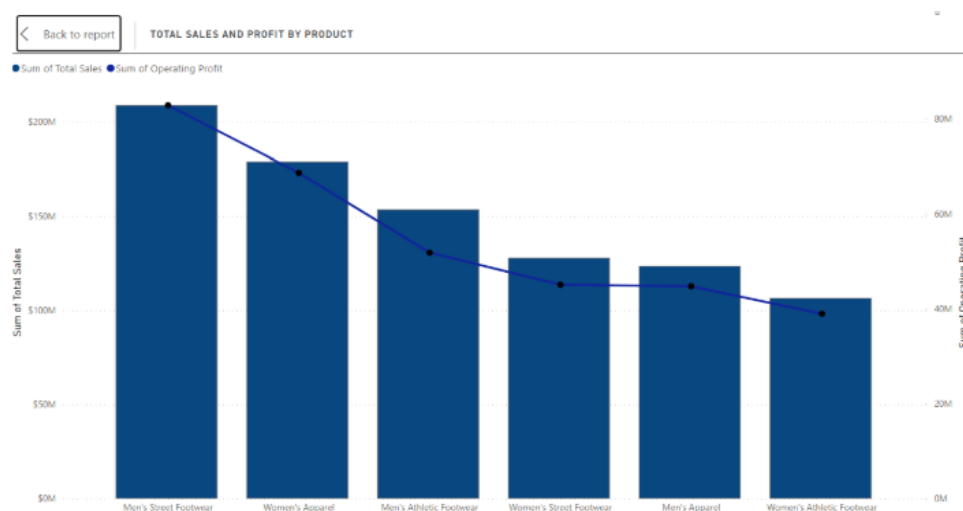

Figure 18: Line and Stacked column chart on Total Sales and Profit by Product

➢ Funnel Chart

This plot is a funnel chart illustrating the sales distribution among different retailers for Adidas. The retailers are listed vertically, with their corresponding sales figures represented by the horizontal width of the bars. The chart shows that Foot Locker has the highest sales with 2,637 units, followed by West Gear with 2,374 units, and Sports Direct with 2,032 units. The retailers with the lowest sales are Kohl's with 1,080 units, Amazon with 949 units, and Walmart with 626 units.
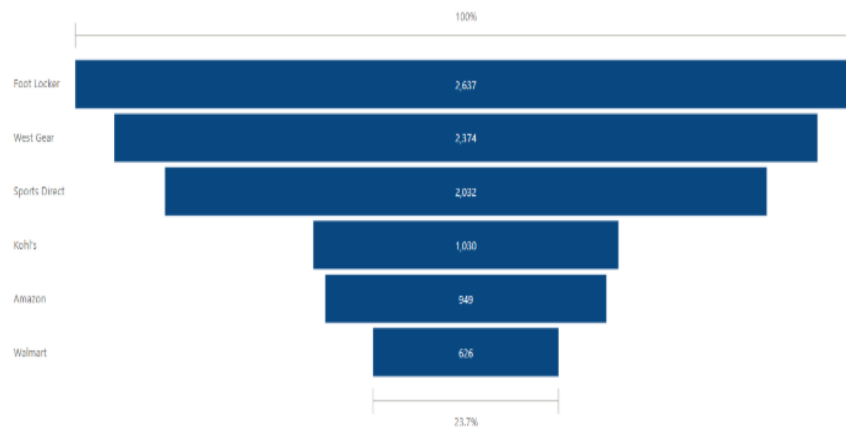
Figure 19: Funnel chart on Sales by Retailer

➢ Donut Chart

This plot is a donut chart depicting the distribution of Adidas sales across different sales channels. The chart segments are color-coded to represent online, outlet, and in-store sales. The largest segment, accounting for 51% of the sales (5K units), is from the online channel. This is followed by in-store sales at 31% (3K units) and outlet sales at 18% (2K units).
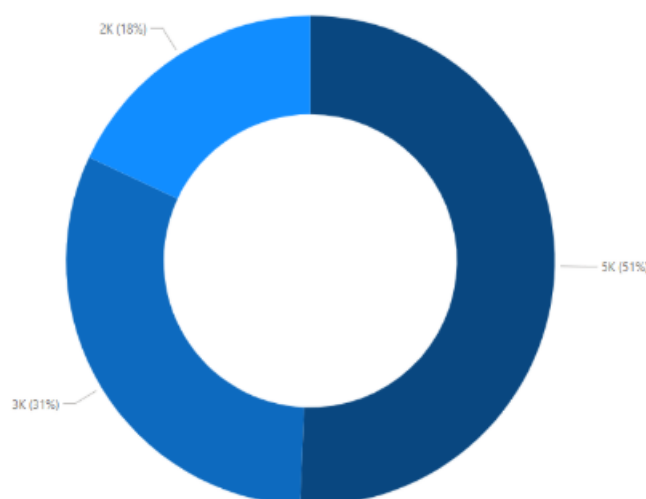
Figure 20: Donut chart on Sales Channel Distribution

➢ Map Chart

This plot is a map chart showing the geographic distribution of Adidas sales across the United States. Each blue circle on the map represents a sales location, with the size of the circle indicating the volume of sales. The highest concentration of sales is seen in the eastern part of the United States, particularly around the Great Lakes and along the East Coast. There are also notable sales volumes in California and a few locations in Hawaii and Alaska, indicating a wide geographic spread of Adidas sales.

Figure 21: Sales by Location using Map Chart

Chapter 3: Conclusion

- During this internship, our team conducted an in-depth analysis of Adidas sales data to provide actionable insights for retail management. The findings highlight several key areas for strategic improvement and optimization.
- Firstly, the sales data revealed clear seasonal trends, with significant peaks during the holiday season and back-to-school months.
- This insight suggests that optimizing inventory and marketing strategies around these periods can maximize sales and reduce stock issues.
- Additionally, specific product lines, such as sportswear and footwear, consistently showed higher sales performance.
- Focusing on expanding these high-performing lines and enhancing their marketing can drive overall sales growth, while underperforming products may require revaluation or rebranding.
- Regional sales analysis uncovered significant variations in demand across different areas, emphasizing the importance of tailoring marketing campaigns and inventory allocation to regional preferences.
- Empowering regional managers with these insights can enhance localized decision-making and improve sales efficiency.
- Understanding customer demographics, particularly age and gender, allowed for the identification of key customer segments.
- Targeted marketing efforts, personalized promotions, and loyalty programs aimed at these demographics can enhance customer engagement and retention.
- Promotional activities were found to significantly impact sales, though their effectiveness varied by product and timing.
- Strategically planning promotions to align with high-traffic periods and determining optimal discount rates can maximize sales without eroding profit margins. Conducting post-promotion analyses will be crucial for refining future promotional strategies.
- Finally, the analysis indicated a growing trend towards online sales, although in-store sales remain substantial.
- This trend highlights the importance of investing in a seamless omnichannel experience that integrates both online and in-store shopping. By doing so, Adidas can meet the evolving preferences and behaviours of its customers.
- Overall, this internship provided valuable insights into retail management strategies through the analysis of Adidas sales data. The recommendations derived from this analysis can help drive sales growth, optimize inventory management, and enhance customer satisfaction.

Chapter 4: Future Scope

➢ The future scope of retail management and analysis using Python and Power BI is vast and promising, driven by the increasing need for data-driven decision-making and personalized customer experiences.

➢ Retailers can leverage Python to implement predictive analytics, minimizing stockouts and overstock situations while tailoring marketing strategies to individual customer preferences. Power BI complements these capabilities by offering dynamic, real-time data visualization and interactive dashboards.

➢ This integration enables retailers to monitor key performance indicators, track sales trends, and gain actionable insights from diverse data sources, ultimately leading to improved operational efficiency and enhanced customer satisfaction.

➢ As the retail landscape continues to evolve, the combination of Python's analytical power and Power BI's visualization prowess will be instrumental in driving innovation and maintaining competitive advantages.

Chapter 5: References

[1] https://www.tutorialspoint.com/power_bi/index.htm

[2] https://docs.python.org/3/tutorial/index.html

[3] https://www.investopedia.com/terms/d/data-analytics.asp

[4] https://www.tableau.com/learn/articles/data-visualization#:~:text=Data%20visualization%20is%20the%20graphical,outliers%2C%20and%20%20patterns%20in%20data.

[5] https://www.indianretailer.com/article/retail-business/retail/understanding-core-objectives-retail-management

[6] https://retalon.com/blog/retail-data-analytics

[7] https://primathon.in/blog/revolutionizing-retail-with-data-visualization/

[8] https://www.youtube.com/watch?v=3Ua6lT7Ye0A

[9] https://www.youtube.com/watch?v=e6QD8lP-m6E&vl=en

[10] https://towardsdatascience.com/machine-learning-for-store-demand-forecasting-and-inventory-optimization-part-1-xgboost-vs-9952d8303b48

Chapter 6: Training Photos

During our summer internship workshop, we received comprehensive training in data analytics and data visualization using multiple industry-standard tools. We honed our skills in Python and R, mastering various libraries and techniques essential for analyzing and manipulating large datasets. Additionally, we delved into Tableau and Power BI, learning to create interactive and visually appealing dashboards that effectively communicate insights. This hands-on experience equipped us with the ability to translate complex data into actionable intelligence, making us proficient in leveraging these tools for real-world data analysis and visualization tasks.

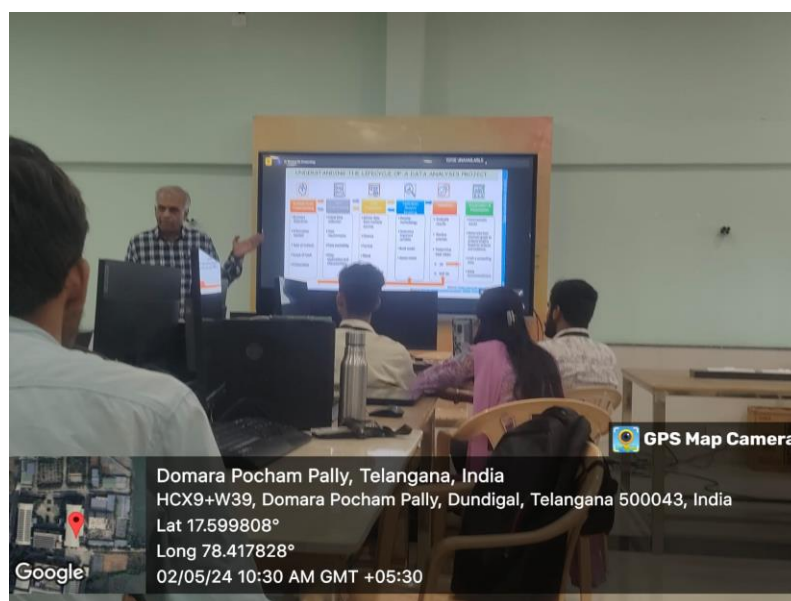Figure 22: Day one training at data analytics workshop

Figure 23: Learning Python, PowerBI, R Language and Tableau

Glossary

Average Transaction Value (ATV): The average amount spent by customers per transaction at Adidas US stores or online.

Bar Graph: A chart with rectangular bars representing data quantities, with bar lengths proportional to the values.

Box Plot: A graphical depiction of data through their quartiles, highlighting the median, outliers, and variability.

Business Intelligence: Involves using data analysis tools and techniques to provide actionable insights and support decision-making within an organization.

City: A large human settlement typically having extensive systems for housing, transportation, sanitation, utilities, and communication.

Clustered Bar Chart: A chart that compares different categories side-by-side, using bars of varying lengths.

Clustered Column Chart: A chart that compares values across different categories by displaying vertical bars side-by-side.

Consumer: Specifically refers to the end-user who consumes or uses the product or service.

Count Plot: A bar plot that shows the counts of observations in each categorical bin using bars.

Customer: Refers to anyone who buys products or services. They may purchase items for personal use, for use by others, or for resale.

Data Analysis: Analyzing sales data, market trends, and customer behavior to make informed decisions.

Data Analytics: The process of analyzing raw data to uncover patterns, trends, and insights that can inform decision-making and drive business strategies.

Data Cleaning: The process of identifying and correcting (or removing) errors, inconsistencies, and inaccuracies in a dataset to improve its quality and ensure it is accurate, consistent, and reliable for analysis.

Data Collection: The process of gathering and measuring information on variables of interest in a systematic manner to answer research questions, test hypotheses, and evaluate outcomes.

Data Preprocessing: The process of transforming raw data into an understandable format, including tasks such as cleaning, normalization, transformation, feature extraction, and selection.

Data Storytelling: The practice of using data to create compelling narratives that communicate insights clearly and effectively to inform and persuade an audience.

Data Visualization: The graphical representation of data to help people understand complex data patterns, relationships, and trends.

Donut Chart: A circular chart with a hole in the center, displaying data as percentages of a whole.

E-commerce Sales: Revenue generated from online sales through Adidas's official website or other online platforms.

Forecasting: Predicting future sales and inventory needs based on data and trends.

Funnel Chart: A chart used to visualize stages in a process, showing values decreasing progressively.

Gross Margin: The difference between net sales revenue and the cost of goods sold (COGS). It represents the percentage of revenue that exceeds the cost of producing the goods sold.

Histogram: A graphical representation of data distribution using bars of different heights.

Interpretability: Refers to the ability to understand and explain how a model or system arrives at its decisions or predictions in a clear and understandable manner.

Invoice Date: The date on which an invoice is issued.

Joint Plot: A plot that combines a scatter plot with univariate histograms for two variables.

Line and Stacked Column Chart: A chart that combines a line graph with stacked columns, useful for comparing trends across categories over time.

Map Chart: A geographical visualization that represents data across different locations using symbols or colors on a map.

Market Dynamics: Refers to the forces and factors that influence the behavior of markets, including supply and demand, competition, consumer preferences, and economic conditions.

Matplotlib: A Python library for creating static, animated, and interactive visualizations.

Net Sales: Total revenue generated by Adidas after subtracting returns, discounts, and allowances.

NumPy: A Python library for numerical computing.

Operating Margin: A profitability ratio that shows what percentage of revenue is left over after paying for variable costs of production.

Operating Profit: The profit earned from a firm's core business operations, excluding deductions of interest and taxes.

OS: Operating System; software that manages computer hardware and software resources.

Pandas: A Python library for data manipulation and analysis.

Pie Chart: A circular chart divided into sectors to illustrate numerical proportions.

Plotly: A data visualization library that allows users to create interactive plots and dashboards in various programming languages, including Python, R, and JavaScript.

Power BI: A business analytics service by Microsoft that provides interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards.

Power BI Dashboard: A customizable collection of visualizations and reports that give users a high-level view of their data and insights.

Price Per Unit: The cost assigned to a single unit of a product.

Product: An item or service offered for sale.

Region: A specific geographic area where business operations or sales occur.

Retail Optimization: Involves using data analysis and strategic planning to enhance various aspects of retail operations, such as inventory management, pricing, store layout, and customer experience, ultimately maximizing efficiency and profitability.

Retail Revenue: Revenue generated from selling Adidas products directly to consumers through its own retail stores, e-commerce platform, and other owned channels.

Retailer: A business or person that sells goods to consumers.

Retailer ID: A unique identifier assigned to each retailer.

Revenue: Total income generated by Adidas from the sale of its products. This includes revenue from both online and offline sales channels.

R: Another popular programming language used for statistical analysis and data visualization.

Rug Plot: A plot that adds small ticks along the bottom and left axes of a plot to show data distribution.

Sales Metrics: Quantifiable measures used to evaluate and track the performance of sales activities, such as revenue generated, number of leads, conversion rates, and average deal size, to assess the effectiveness of sales strategies and drive improvements.

Sales Method: The approach or technique used to sell products or services to customers.

Same-store Sales: Also known as comparable-store sales, this metric measures the sales growth of Adidas stores that have been open for a certain period of time (usually a year or more). It excludes sales from new store openings.

Scikit-learn: A machine learning library for Python that provides simple and efficient tools for data mining and data analysis.

Scatter Plot: A graph of plotted points that show the relationship between two sets of data.

Seaborn: A Python data visualization library based on Matplotlib, providing a high-level interface for drawing attractive statistical graphics.

Slicer: A Power BI visualization tool used for filtering data within reports and dashboards based on user-defined criteria.

State: A sub-national administrative division within a country.

Stacked Column Chart: A chart that displays data as stacked columns, where each column represents a different category and the height of each segment represents the value.

Strip Plot: A scatter plot where one of the variables is categorical.

Subplot: A section of a single figure in a grid layout that can contain a different plot.

Total Sales: The cumulative revenue generated from all sold products or services.

Units Sold: The number of individual items sold.

Violin Plot: A method of plotting numeric data that combines aspects of a box plot and a kernel density plot.

Wholesale Revenue: Revenue generated from selling Adidas products to third-party retailers, wholesalers, and distributors.

Appendix 1

POWER BI

Introduction to Power BI

Power BI is a collection of software services, apps, and connectors that work together to transform your unrelated sources of data into coherent, visually immersive, and interactive insights. Your data may be an Excel spreadsheet or a collection of cloud-based and on-premises hybrid data warehouses. Power BI is used for a variety of reasons, primarily revolving around its ability to transform raw data into actionable insights.

Why Use Power BI?

Power BI is USED for its powerful data visualization capabilities. Power BI provides robust tools for creating interactive visualizations such as charts, graphs, and maps, which help in understanding complex data sets quickly and making data-driven decisions.

Additionally, Power BI offers extensive data connectivity, allowing users to connect to a wide range of data sources, including databases like SQL Server, Oracle, and MySQL, cloud services like Azure, Google Analytics, and Salesforce, flat files like Excel and CSV, and online services like Facebook and Twitter. This connectivity enables users to aggregate data from multiple sources into a single, coherent view.

With Power Query, users can transform and clean data, making it ready for analysis by removing duplicates, filtering rows, and merging tables. Power BI also supports real-time analytics, allowing users to monitor live dashboards and make timely decisions based on the most current information.

Designed for business users, Power BI's user-friendly interface allows non-technical users to create reports and dashboards without extensive training or IT support. Power BI also supports advanced analytics capabilities, including DAX (Data Analysis Expressions) for complex calculations, integration with R and Python for statistical analysis and machine learning, and AI-powered features like Quick Insights and Natural Language Queries.

Practical Use Cases

a) **Sales and Marketing**: Track sales performance, analyse customer data, and measure campaign effectiveness.
b) **Finance**: Monitor financial performance, manage budgets, and forecast revenue.
c) **Operations**: Optimize supply chain management, track inventory levels, and improve operational efficiency.
d) **Human Resources**: Analyse workforce data, monitor employee performance, and track recruitment metrics.

e) **Healthcare**: Monitor patient outcomes, analyse clinical data, and improve service delivery.

Steps to Use

a) First, download and install Power BI Desktop from the Power BI website.
b) Then, connect to your data sources by selecting "Get Data" on the Home ribbon and following the prompts to load your data.
c) Next, transform and clean your data using Power Query Editor by applying necessary transformations to ensure it is ready for analysis.
d) Afterward, create visualizations by selecting the type of visualization you want and dragging fields from the Fields pane into the Visualizations pane.
e) You can then build a report by adding multiple visualizations, arranging them on the canvas, and adding filters and slicers for interactivity.
f) Once your report is ready, save it locally, publish it to the Power BI Service, and share it with colleagues by setting permissions for viewing or editing.
g) You can also create dashboards by pinning visuals from your reports and customizing the dashboard layout.
h) Power BI's mobile app allows you to view and interact with your reports and dashboards on

Advantages of Power BI:

- User-friendly interface.
- Robust data connectivity.
- Powerful data transformation tools.
- Interactive visualizations.
- Real-time analytics.
- Advanced analytics capabilities.
- Collaboration and sharing features.
- Mobile access.
- Scalability for small and large organizations.
- Cost-effective options.
- Seamless integration with Microsoft products.
- Robust security features.

Practical Benefits:

- Provides actionable insights.
- Streamlines data analysis and reporting processes.
- Enables real-time monitoring of key performance indicators (KPIs) and metrics.
- Customizable dashboards and reports to meet specific business needs.

Scope of Power BI:

- o Business intelligence and analytics, Finance, Sales and marketing.
- o Healthcare, Retail and Manufacturing, Human resources and Operations.
- o IT and data management, Advanced analytics and data science.
- o Integration and extensibility, Real-time and mobile BI.
- o Collaboration and sharing, Educational and training use.
- o Scalability from small businesses to large enterprises.
- o Extensive capabilities in data connectivity, transformation, visualization, and sharing, supports strategic decision-making and meets diverse business needs.

Appendix 2

PYTHON

Introduction to Python

Python is a versatile and powerful programming language known for its simplicity and readability. It is used for a wide range of applications, including web development, data analysis, artificial intelligence, scientific computing, and more. Python's extensive libraries and frameworks which make it an ideal choice for both beginners and experienced developers. Python's ability to handle large amounts of data and perform complex computations has made it a popular tool for data scientists and analysts.

Why Use Python?

Python is used for its simplicity and readability, making it an excellent choice for beginners and experienced programmers alike. Python's syntax is designed to be intuitive, reducing the learning curve and allowing developers to write clean and maintainable code. Additionally, Python offers a vast ecosystem of libraries and frameworks that extend its capabilities and support a wide range of applications. Key reasons to use Python include:

a) Data Analysis and Visualization: Python provides powerful libraries like Pandas, NumPy, Matplotlib, and Seaborn for data manipulation, analysis, and visualization, enabling users to extract insights and create informative visual representations of data.

b) Machine Learning and AI: Python's libraries such as TensorFlow, Keras, and Scikit-learn make it easy to develop and deploy machine learning models and perform complex computations for artificial intelligence applications.

c) Web Development: Frameworks like Django and Flask allow developers to build robust and scalable web applications quickly and efficiently.

d) Automation and Scripting: Python's simplicity and ease of use make it an ideal choice for writing scripts to automate repetitive tasks, improving productivity and efficiency.

e) Scientific Computing: Libraries like SciPy and SymPy enable scientific computing, solving mathematical problems, and performing simulations.

f) Extensive Community Support: Python has a large and active community, providing a wealth of resources, tutorials, and support for developers at all levels.

Practical Use Cases

a) **Data Analysis**: Analyze large datasets, perform statistical analyses, and visualize data trends.

b) **Machine Learning**: Develop and train machine learning models for predictive analytics and classification tasks.

c) **Web Development**: Create dynamic and interactive web applications with frameworks like Django and Flask.

d) **Automation**: Write scripts to automate tasks such as data extraction, file management, and system monitoring.

e) **Scientific Research**: Conduct simulations, solve complex equations, and analyze experimental data.

Steps to Use Python

a) **Install Python**: Download and install Python from the official Python website. Ensure you have the latest version for access to the latest features and libraries.

b) **Set Up an IDE**: Choose an Integrated Development Environment (IDE) like PyCharm, VS Code, or Jupyter Notebook for writing and running your Python code.

c) **Install Necessary Libraries**: Use pip, Python's package manager, to install libraries such as Pandas, NumPy, Matplotlib, and others required for your projects.

d) **Write and Run Python Code**: Create Python scripts or Jupyter notebooks to write and execute your code. Use Python's syntax and libraries to perform data manipulation, analysis, and visualization.

e) **Develop and Test**: Continuously test your code to ensure it works as expected. Use debugging tools provided by your IDE to troubleshoot and fix any issues.

f) **Share Your Work**: Share your Python scripts or notebooks with colleagues by using version control systems like Git or by exporting and sharing the files directly.

Advantages of Python:

- Simple and readable syntax.
- Extensive standard library and third-party modules.
- Strong support for data manipulation and analysis.
- Robust frameworks for web development.
- Powerful machine learning and AI capabilities.
- Strong community support and abundant resources.
- Cross-platform compatibility.
- High productivity and efficiency in development.
- Integration capabilities with other languages and tools.
- Cost-effective, as Python is open-source and free to use.
-

Practical Benefits:

- Streamlines data analysis and visualization.
- Facilitates rapid development and deployment of applications.
- Enhances productivity through automation and scripting.
- Enables sophisticated machine learning and AI model development.
- Supports a wide range of applications from web development to scientific research.

Scope of Python:

- Data analysis and visualization.
- Machine learning and artificial intelligence.
- Web development and backend services.
- Automation and scripting.
- Scientific computing and simulations.
- Financial modeling and analysis.
- Game development.
- IoT (Internet of Things) and embedded systems.
- Education and training.
- Cross-industry applications including healthcare, finance, retail, and more.
- Scalability from small projects to large enterprise applications.
- Extensive capabilities in data handling, computation, and application development

THANK
YOU