Adidas Sales: Technical Appendix

Dream Team 6B: Luisa Garate, Xinyi He, Naomi Hossain, Weinuo Jin, Tysean Haigood

Overview

This technical appendix outlines the research design, data preparation steps, and analytical methodologies used to examine Adidas U.S sales data set from 2020-2021. It expands information beyond our presentation including data sources, statistical modeling procedures, and full results. This gives a deeper understanding of the insights generated in our project. This project was completed by Team 6B, 'Dream Team'.

Introduction

Consumerism is a central force in the U.S. economy, and understanding the dynamics of sales is essential for evaluating business performance within this environment. Conducting detailed sales analyses allows organizations to better understand their target audiences, identify purchasing patterns, and refine their strategic decision-making.

In this research, we analyzed sales data from Adidas to examine the key factors that drive the company's performance. Our goal was to determine whether product attributes, regional differences, or other market characteristics have a significant impact on sales outcomes. By closely evaluating Adidas's performance across regions and product categories, we aim to provide data-driven insights and recommendations that can help the company enhance its competitive position and overall success. Our overall question answered in this report is: What is the impact of sales channel, region, and product category on Adidas operating profit?

Data Description

This dataset was found on Kaggle and was publicly posted as 'Adidas Sales Dataset'. Since this data is available to the public our group did not have to collect this dataset through methods such as surveys, scraping, or internal company systems. There is limited information on the raw data-collection of this dataset. The dataset provides detailed transactional information capturing sales activity across various retailers, product categories, and regions in the United States. The data was sourced from a Kaggle dataset that contains 9,648 records of Adidas purchases with 13 variables. The data focuses on U.S. regional sales from 2020 to 2021. The 13 variables included:

Variable Name	Description
Retailer	Name of the stores selling Adidas products.
Retailer ID	Unique numeric identifier for each retailer.
Invoice Date	The date the sales transaction was recorded.
Region	U.S. geographical region where the sales occurred (West, Northeast, Southeast, South, Midwest).
State	The U.S. state where the sale was made.
City	The city where the sale was made.
Product	The specific Adidas product category sold.

Price Per Unit	The selling price (USD) for one unit of the product.
Units Sold	Number of units purchased.
Total Sales	Total revenue from that transaction.
Operating Profit	Profit generated from the sales after subtracting operating costs.
Operating Margin	Profit ratio expressed as a decimal.
Sales Method	How the sale occurred (in-store, online, outlet).

Some thing to note, for the *Region* variable, regions are labeled as: West, Northeast, Southeast, South, and Midwest. *Total Sales* was calculated using *Price Per Unit* multiplied by *Units Sold*. *Operating Margin* was calculated by *Operating Profit* divided by *Total Sales*. Lastly, the sales methods used for the *Sales Method* variable: in-store, online, and outlet. No data was transformed for the processing part.

Data Processing

The dataset was scanned and in order to check if there were any null values. We found the dataset to be clean and contain no missing values. We also checked if there were any outliers and we did not find any outliers. We divided different predictors in our project. With independent Variables being: Sales Method, Region, Product Category, Price per Unit. Control Variables being: Retailer, Units Sold, Invoice Date. Dependent Variable: Operating Profit. A multicollinearity diagnosis was conducted since there was a concern that 'Price' may be correlated with 'Product' multicollinearity. Variance Inflation Factors (VIF) was computed and found that all VIFs fell between 1.2 and 2.1 which indicated no multicollinearity concern. This validated the inclusion of 'Price' as a control variable.

Analytical Methods

For our analysis we used a combination of Excel and Python in order to obtain the best results. Our primary analytical method for this project was a multiple linear regression. We decided that regression was the best method for our project because operating profit is a continuous variable. The multiple regression analysis was conducted to identify the key factors influencing Adidas's operating profit across U.S. sales transactions. Overall, the model is statistically significant ($p < 0.001$), indicating that the included predictors reliably explain variation in operating profit. The model yields an R^2 of 0.34, suggesting that approximately 34% of the variation in operating profit is explained by the variables included in the model. While this represents a moderate level of explanatory power, the significance of the F-statistic ($F = 419.9$, $p < 0.001$) demonstrates that the model as a whole is strong and valid for understanding drivers of profitability.

We included an Exploratory Data Analysis (EDA) in order to understand the differences in operating profit across categories. We compared the sales method and found that the in-store method had the highest profit, then followed by profit and lastly online. A regional comparison was also done with Southeast having the highest-performing region. The Midwest and Northeast were the weakest performing regions.A product category comparison was also done with Men's Street Footwear generating the highest average operating profit. Women's categories leaned towards generating less. Lastly, we completed a heat map analysis. A cross-tab of heat map of region x sales revealed the most and least profitable combinations. The cell that was most profitable was the Southeast In-store while the Northeast Online was the lowest profitable

Results

Derived from our multiple linear regression model, the model highlights that sales channel, region, product type, and pricing are all significant drivers of Adidas's operating profit. We saw that the sales method of Online -55,020 ($p < 0.001$) and Outlet -32,290 ($p < 0.001$) performed significantly worse than In-store. For regional effects we see Southeast 8,630 ($p < 0.001$) and Southeast -16,530($p < 0.001$). With this we saw Southeast being the strongest region and Northeast the weakest. With Category Effect, we saw Men's Street Footwear: 33,380 ($p < 0.001$) and Men's Athletic footwear 14,900 ($p < 0.001$) This shows that Men's categories drive more profit than women's categories. With Price per Unit we saw that 1,603($p < 0.001$), meaning higher prices are associated with higher operating profit. We also created a cross-tab heatmap to obtain the most and least profitable combinations. We found that the profitable combination was In-store sales in the Southeast region. This had the highest profit as 131,869. While in comparison the combination with the least profitable combination was online sales in the Northeast, producing an average profit of 2,400. With our regression findings and heatmap insights, we are able to conclude that profitability is shaped by interactions between channel, region, and product strategy. In-store sales, specific regions such as the Southwest, and Men's Street Footwear product category contribute the most to profitability. Positive price effects indicate that maintaining strong price posits is beneficial for profit growth. Despite a moderate R^2 , the statistical strength and clarity of the findings show that the model effectively captures the most impactful determinants of profit, providing meaningful insights for strategic decision making.

References

Adidas. (2025, November 8). *In Wikipedia*. https://en.wikipedia.org/wiki/Adidas

Chaudhari, H. (2022). *Adidas sales dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/heemalichaudhari/adidas-sales-dataset