Team: Insights Innovators

Project Phase – 8

Final Report

Team Members:

1. **Krisha Patel** (kpate65@unh.newhaven.edu)
2. **Dhvani Patel** (dpate86@unh.newhaven.edu)
3. **Prem Kumar Sandireddy** (psand4@unh.newhaven.edu)

Table of Contents

Abstract

This study investigates the factors that influence Nike's sales performance in the United States, focusing on product categories, regional trends, and sales methods. Using the Nike_sales_nti15 dataset from Kaggle, we preprocessed, explored, and analyzed the data using statistical and machine learning techniques, including association rule mining, decision trees, and clustering. Key insights show strong regional and seasonal variations in sales, different customer preferences for online versus in-store purchases, and comparisons of footwear and apparel performance. Results put in light the dominance of e-commerce in certain regions, the leading retailers like Foot Locker, and seasonal patterns that affect sales. The findings provide actionable insights for optimizing marketing, inventory management, and sales strategies. Future work will expand the dataset to provide global and demographic insights and include predictive analytics to track sales in real time.

Introduction

- **Data set (Nike_sales_nti15) → (nike-sales-nti15)**

Indeed, the modern global sportswear marketplace has been registering tremendous growth lately, basically founded on shifting consumer preferences, enhanced technologies, and growing interest in lifestyles related to activity. These new changes create an environment that really forces sportswear brands into continuous innovation if they hope to win and retain market share. Nike has been a leading player among such companies in this dynamic industry, known for its innovative product offerings and strong market presence. Analyzing sales data is very pivotal in understanding market dynamics, as it highlights consumer behavior, regional market trends, and performance across different product categories. Such information is crucial for the companies to further develop their strategies and remain competitive. This project will be focused on analyzing sales patterns and performance in Nike's vast product line, using the Nike_sales_nti15 dataset.

This dataset includes sales invoices for Nike footwear and apparel sold across various diverse regions, retailers, and states within the United States. This research, therefore, tries to find the underlying determinants of sales success through the investigation of critical factors such as product category, sales method, and temporal trends. The focus is on understanding how these factors shape consumer purchasing behavior and influence overall sales performance. By exploring these dimensions, the study will seek to identify actionable insights that can optimize Nike's market strategies and further strengthen its competitive position. Results derived from this study are expected to develop a better understanding of consumers' preferences in the sportswear market. Analyzing seasonal sales trends in this research will also provide the outline necessary for mapping out patterns that will significantly lead to marketing campaigns and managing stock for each particular season.

Second, it tries to find out how e-channels and physical retail will bring about a sense of variations with digital transformation and traditional retail methodologies upon sales overall. This study will

try to shed light on the driving factors that explain Nike's success in varied regions and product categories by focusing on a comprehensive analysis of sales data, thus providing valuable perspectives for both academics and business strategy development.

Related Work

Various works have been done on different facets of Nike's strategies and operations that show how the company has maintained its lead within the global sportswear market. Among them, one of the most striking pieces of work is "Learning from Nike: Interactive Retail Spaces and Brand Engagement" by Williams and Harris (2020). The findings of this study have identified how Nike's innovative and interactive retail spaces are playing a very important role in the enhancement of brand engagement and customer loyalty. Offering immersive experiences through various means, including personalized shopping and innovative in-store technologies, Nike has increased its connection with consumers to give a boost to its bottom line. This study ascertains that these interactive spaces offer an opportunity to expose customers to more products, while at the same time giving them the space to know the Nike brand much closer than the existing way can allow, therefore helping shape their purchase decision in one way or another.

On the other hand, the study "Supply Chain Management of Nike", conducted by Liu, Ma, and Zhou (2021), largely deals with how Nike can efficiently manage its global supply chain, sustainably and effectively. It looks at how Nike integrated advanced technologies into data analytics and automation to facilitate production processes, optimize inventory management, and decrease operational costs. The study also investigates Nike's commitment to sustainability in the value chain: how the company has tried to reduce its carbon footprint, maintain ethical labor standards, and source materials in a responsible manner. These activities enable Nike to be competitive in response to the increasing demands of consumers for business responsibility.

The work " Strategic Marketing Plan of Nike " by Williams and Harris (2020) is also relevant; while considering Nike core determinant of its global success, they look at its marketing strategy. The authors outline a company approach regarding advertising activity, celebrity endorsements, and strategic partnerships that help Nike keep its leading position in the market. The study explains how Nike continues to be innovative in its marketing campaigns, using various inspiring advertisements related to diverse consumer groups of athletes or in partnership with major sports events. According to the authors, the ability of Nike to create emotionally powerful content in marketing, while adhering to social causes, has made its brand synonymous with empowerment and performance.

Stevens and Ross's "Nike's Flagship Stores and the Future of Retail Experience"(2020) gives insights into how Nike flagship stores shape the future of retail. This research focuses on flagship stores as epicenters for experience wherein digital and physical seamlessly blend. The authors focus on flagship Nike stores located in New York, London, and Tokyo, where one can test

products, do some fitness work, and even customize products. In so doing, Nike is improving the customer experience, and these showrooms also act as a testing ground for Nike's latest innovations in further enhancing brand image and connecting consumers with the company's mission.

Finally, "Interactive and Immersive Retail Spaces: Nike's Global Strategy" by Brown and Lee (2018), identifies how Nike has integrated emerging technologies within the retail setting to drive consumer interaction. The authors have discussed how augmented reality, virtual reality, and interactive displays are used to engage a consumer in an immersive shopping experience. This helps Nike reach those customers who are technologically savvy and seek novelty in their shopping experience. The study also highlights how such technologies attract customers and, on the other hand, help Nike collect valuable data about consumer preferences, thus enabling it to refine its marketing and product offerings.

These studies, therefore, compile a multidimensional outlook on the operation and strategy of Nike: from new retail experiences and state-of-the-art marketing to complicated chain management and the integration of technology. Each of these works underlines the ability of Nike to adapt to changes in market trends and increase consumer engagement, retaining a leader position in the sportswear industry.

Proposed Approach

The proposed approach for the analysis of Nike sales data will follow a systematic approach in terms of a combination of data preprocessing, EDA, mining techniques, and performance optimization to derive actionable insights from the data. This will help gain deeper insights into the sales trend of Nike and, hence, optimize marketing strategies.

1. **Dataset Overview and Accessibility**:
   The dataset is collected from Kaggle, nike-sales-nti15, which has Nike sales for footwear and apparel over different regions and retailers of the United States. The key attributes in this dataset are invoice date, product type-which would be either footwear or apparel, retailer, sales method as online or in-store, state, price per unit, total sales, units sold. This data is stored in a CSV file and can be imported into Python, analyzed using the help of libraries such as pandas, NumPy, and matplotlib.
2. **Preprocessing Data**:
   Dataset pre-processing is one of the approaches used to do the preprocessing of this dataset. It will consist of:
   - Cleaning the data: It may be performed by determining values that represent missing data, eliminating all kinds of duplication that may occur in this regard, and maintaining consistency to some extent-for instance, using datetime format for all the date columns and numeric type as it is.
   - Data Transformation: It may be made through normalization or scaling of the numerical features, such as total sales and price per unit, and encoding categorical variables like

product type, region, and sales method, using techniques like one-hot encoding or label encoding.

3. **Data Exploration (Visualization):**
   After cleaning the data, the next step is to do EDA. This phase helps identify patterns, trends, and anomalies in the data. Key techniques include:
   - Bar Charts: Distribution of Sales across Footwear and Apparel, Regions, and Sales Methods - Online vs. In-Store.
   - Heatmaps: Find the correlation between numerical features like Price per Unit, Total Sales, and Units Sold to know the potential relationship between various factors. Time Series Analysis: Assess how sales trends change over time, showing seasonal patterns or significant fluctuations.

4. **Data Mining Techniques**:
   Advanced data mining techniques will be applied to unlock further insight from the data after the exploratory analysis is complete, including the following:
   - Association Rule Mining: It helps discover how two products or any other attribute may co-occur-for example, which products tend to sell together or how different regions perform with various products.
   - Classification: The Decision Tree Classifier classifies the regions, retailers, or selling methods into poor, average, or good regarding sales performance. For example, it might predict the high-performing regions for a particular category and time.
   - Clustering: K-Means or Hierarchical Clustering groups regions or retailers with similar sales patterns to identify market segments with comparable sales behavior.
   - Support Vector Machines: These are used to classify regions or sales methods based on sales data and predict future sales trends based on historical performance.

5. **Data Optimization**:
   The performance of the models should be optimized for correct predictions. The following techniques will be used in optimizing model performance:
   - Hyperparameter Tuning: The proposed research work will also conduct the best settings for machine learning models using techniques such as Randomized Search Cross-Validation and Grid Search Cross-Validation.
   - Ensemble Methods: These will be developed on methods that enhance the prediction performance of the model, including a number of weak learners to come up with a robust model such as Random Forest or AdaBoost.

6. **Performance Metrics:**
   The following are some of the metrics that shall be used in the analysis of the performance of the proposed models:
   - Accuracy: Overall correctness of the model predictions.
   - Precision and Recall: To check how well the model identifies the positives (like sales spikes or high performing regions) while avoiding false positives or false negatives.
   - R-Squared (R^2): The quality of regression model explanation in variance of sales data.
   - F1 Score: To balance precision and recall, particularly useful for imbalanced datasets.

The mean squared error, or MSE, calculates the average of the squared discrepancies between the sales figures that were anticipated and real. This makes the model's performance in this regression challenge easier to see.

7. **Python Implementation Libraries:**
   Key Python libraries for data manipulation (like pandas), numerical operations (like NumPy), visualization (like matplotlib and seaborn), and machine learning algorithms and metrics (like scikit-learn) will be utilized in conjunction with Python.

Experimental Results

The experimental results highlight some interesting patterns related to sales behavior across regions in the United States. Among these, the South leads the charge in its significant preference for online shopping, further pinpointed by certain specific retailers such as Sports Direct and Foot Locker. This agrees with the association rules, which show a huge connection between the South and online sales methods, meaning that consumers here are more likely to use online platforms to make a purchase.

This observation is further enhanced by the network graph, where the South region is closely related to online sales methods, especially with retailers like Sports Direct and Foot Locker. This means that the South region might be very responsive to e-commerce strategies, including online promotions and digital marketing.

In contrast, the Midwest region shows a more varied approach, with both online and in-store purchases being equally significant. The association rules indicate that the Midwest region is associated with different retailers and sales methods, showing that consumers in this region are less fixed on one particular sales channel. This would suggest a more balanced and diverse shopping behavior in which physical stores and online platforms both play an essential role.

The Northeast has a strong penchant for large, more known online platforms, such as Amazon. The association rules further validate that consumers in this region have a very strong tendency to shop online via large e-commerce giants, reflecting the region's trend toward convenience and broad product availability.

In all, findings indicate that each region has its peculiarity in shopping behaviors. While the South region may be more inclined toward online sales, especially with certain retailers, other regions such as the Midwest and Northeast reflect a varied or platform-specific trend. These insights indicate that understanding regional shopping behaviors can be very important to businesses in terms of informing marketing strategies and optimizing sales channels based on unique consumer preferences across different regions.

Network Graph of Association Rules
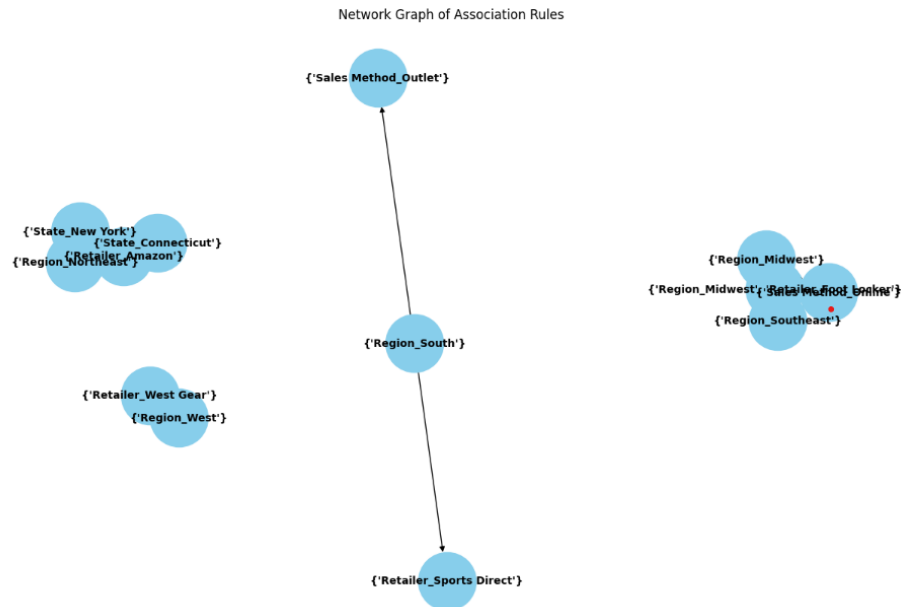
Regional Analysis of Most Purchased Categories

Each region showcases unique consumer preferences based on sales methods and product types. The Northeast region, focused on online sales, sees Men's Street Footwear as the top category with 1,226 units sold, followed by Women's Apparel and Men's Athletic Footwear. In the South region, where outlet sales dominate, Men's Athletic Footwear leads with 1,823 units, while Men's Street Footwear closely follows with 1,814 units.
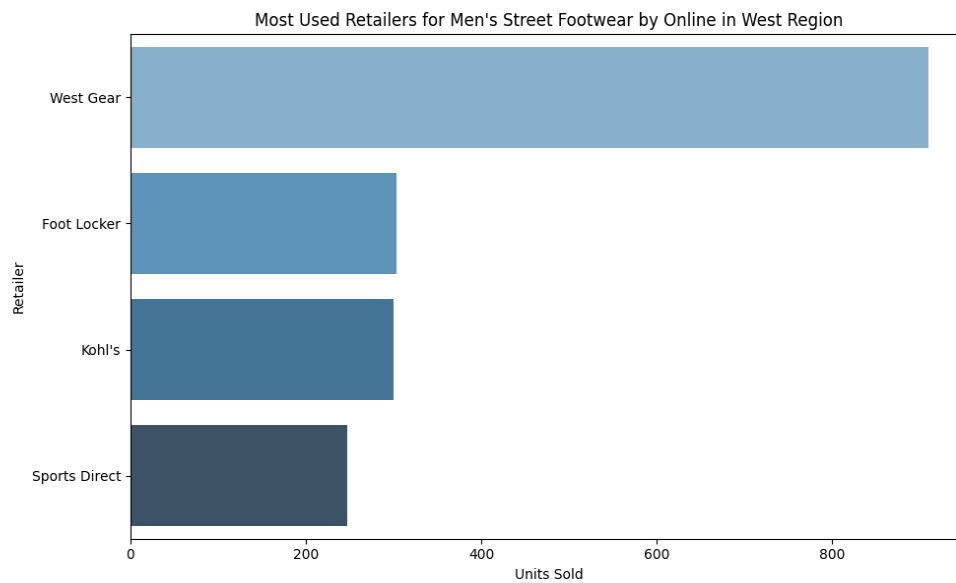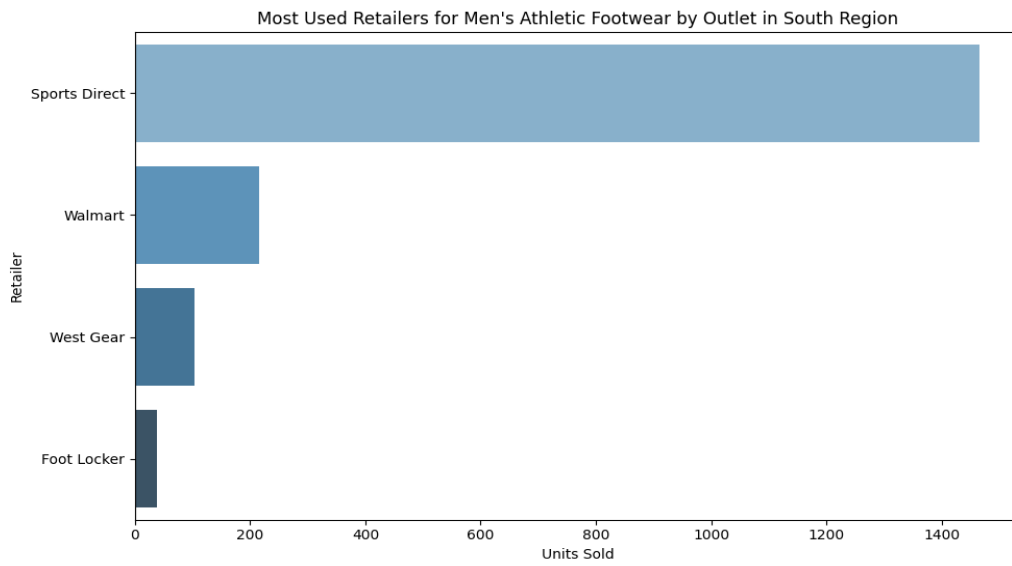
The West region, driven by online sales, also prioritizes Men's Street Footwear with 1,759 units, while the Midwest region shows an exceptionally high demand for the same category, achieving 2,951 units sold. Finally, in the Southeast region, Women's Street Footwear takes the lead with 2,518 units, indicating a strong preference for women's products in this area.
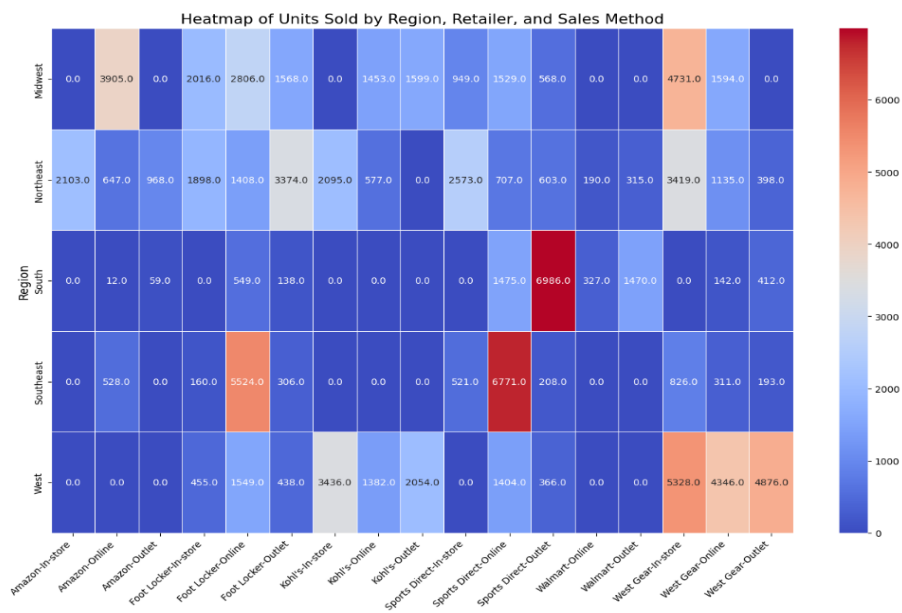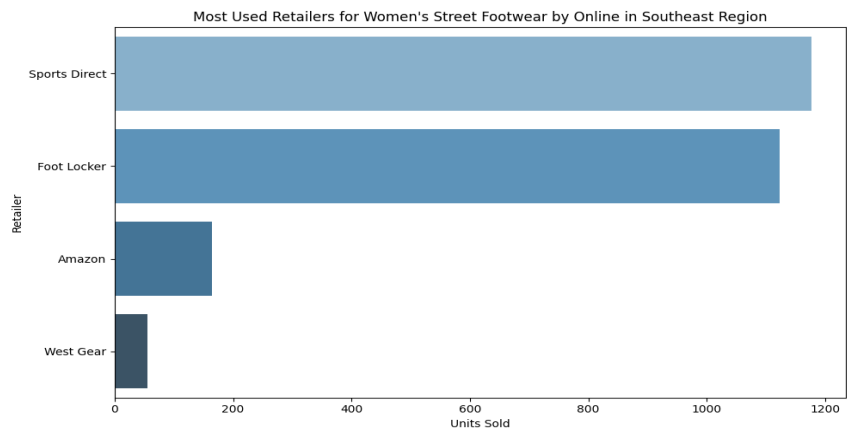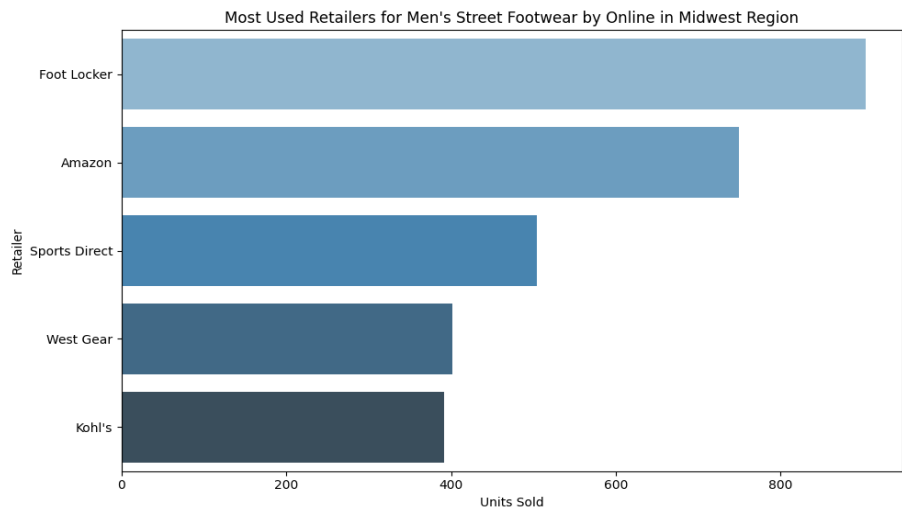
Retailer Analysis by Region

Retailer performance varies significantly based on product type and region. For Men's Street Footwear in the Northeast, Foot Locker leads the market with 363 units sold, followed by West Gear at 279 units. In the South, for Men's Athletic Footwear via outlet sales, Sports Direct dominates with 1,465 units, significantly outpacing other retailers like Walmart (216 units).

The West region sees West Gear leading online sales of Men's Street Footwear, contributing 909 units, more than triple the sales of competitors like Foot Locker (303 units). In the Midwest, Foot Locker remains the leader with 903 units, while Amazon makes a strong appearance with 750 units sold.

Most Used Retailers for Men's Street Footwear by Online in Northeast Region

Most Used Retailers for Men's Athletic Footwear by Outlet in South Region

Most Used Retailers for Men's Street Footwear by Online in West Region

Most Used Retailers for Men's Street Footwear by Online in Midwest Region

Most Used Retailers for Women's Street Footwear by Online in Southeast Region

Heatmap of Units Sold by Region, Retailer, and Sales Method

The heat map represents complex interactions between region, retailer, and sales method in driving unit sales. While Foot Locker and Kohl's are consistent performers across regions and sales channels, Sports Direct exhibits mixed performance, performing well in the Southeast and West but poorly in the Northeast and Midwest. Amazon and Walmart are also strong online, although they tend to be stronger in the West and Southeast than in other regions. In-store generally leads all others, online and outlet, for all but Foot Locker and Kohl's. The Northeast is the strongest region overall, with the Southeast next, while the Midwest and West are more mixed.

Most Purchased Categories by Region and Sales Method:

Region	Sales Method	Top Category	Units Sold	2nd Category	Units Sold
Northeast	Online	Men's Street Footwear	1,226	Women's Apparel	801
South	Outlet	Men's Athletic Footwear	1,823	Men's Street Footwear	1,814
West	Online	Men's Street Footwear	1,759	Men's Athletic Footwear	1,592
Midwest	Online	Men's Street Footwear	2,951	Women's Apparel	2,073
Southeast	Online	Women's Street Footwear	2,518	Men's Athletic Footwear	2,352

Top Retailers for Key Products by Region:

Region	Product	Retailer	Units Sold	2nd Retailer	Units Sold
Northeast	Men's Street Footwear	Foot Locker	363	West Gear	279
South	Men's Athletic Footwear	Sports Direct	1,465	Walmart	216
West	Men's Street Footwear	West Gear	909	Foot Locker	303
Midwest	Men's Street Footwear	Foot Locker	903	Amazon	750
Southeast	Women's Street Footwear	Sports Direct	1,176	Foot Locker	1,123

Discussion

1. **Regional Preferences Impact on Sales Performance:**
   The results show there is a significant regional difference in sales performance for Nike. For example, it can be ascertained that the Southern region has a strong preference for online sales, especially with retailers like Foot Locker. This would also mean that e-commerce efforts should be prioritized in this area to take full advantage of the increase in online shopping. On the other hand, the Midwest region reflects a balanced approach, wherein both online and in-store purchases are contributing significantly to sales. Understanding these regional preferences will enable Nike to adopt effective marketing strategies and inventory management to meet the specific demands of each region. This

indicates that regional targeting is very important to optimize sales and enhance customer satisfaction.

2. **Influence of Product Categories on Sales:**
   Product categories also determine overall sales performance.
   In the Northeast region, for instance, Men's Street Footwear is the top-selling category. This agrees with world trends of increasing demand for sports and casual footwear. On the other hand, in the South region, Men's Athletic Footwear shows strong sales. This further underline the need for detailed knowledge of the performance of product categories across different regions to enable the design of more focused product offerings and marketing campaigns. Such insights will also drive inventory decisions to ensure that the right categories of products are available in the right place at the right time.

3. **Online Sales Channels and Digital Transformation:**
   Growth through online sales, especially in regions like the South and West, indicates digital transformation gaining fast within retail.

Nike's online outreach-e-commerce strategy, online campaigns, and digital marketing-also greatly captures the attention of people when it comes to online purchasing.

It could also be interpreted that with this strong association of online sale in respective regions with specific retailers, Nike needs to strengthen its digital position in targeted regions. Besides, the integration of AR and virtual try-ons on online platforms can create an improved online shopping experience and further drive online sales. The ever-evolving nature of consumer shopping behavior means that, to stay competitive, Nike needs to be at the edge of digital trends.

Conclusion

This report analyzed Nike's sales data comprehensively to gain insight into sales performance, retailer influence, and product category trends. The following conclusions were made:

1. **Sales Performance by Region:**
   Men's street footwear leads e-commerce sales nationwide, with the highest sales volumes in the central and western U.S. Women's Street footwear demonstrates significant strength in the Southeast, with a pivotal role of regional preferences in driving sales strategies.

2. **Retailer Influence:**
   Retailers like Foot Locker and Sports Direct are influential in shaping sales performance. Foot Locker dominates Men's Street Footwear in the Northeast and Midwest, while Sports Direct holds the number one position in the South for Men's Athletic Footwear. This underlines how retailer-specific strategies drive regional sales success.

3. **Product Category Insights:**
   Men's and Women's are performing very strong across all geographies, but mostly via online channels in recent times. The outlet sales will still remain stronger in the South and maintain a good balance in demand between online and offline purchasing touchpoints.

4. **Overall Insights:**
   Regional variations in sales indicate that market development requires effective implementation of specific marketing and distribution campaigns. The collaboration with retailers acts as the key driver towards market establishment in the target geography. Online channels present significant growth opportunities, especially in regions where they dominate sales.

Future Work

From the analysis, the subsequent areas of future work that would help further in formulating Nike's sales strategy to enhance customer satisfaction are:

1. **Deep Dive into Customer Demographics:**
   It Conduct more in-depth analysis of customer demographics regarding specific age groups, income levels, and preferences influencing sales in different regions to further refine Nike's marketing and product offerings.
2. **Predictive Modeling and Trend Analysis:**
   Implement advanced predictive analytics to forecast future sales trends based on seasonal variations, product launches, and marketing campaigns. This will help in proactive inventory and supply chain management.
3. **Global Expansion of Regional Analysis:**
   To extend the regional sales analysis to international markets to identify global patterns and unique regional preferences, ensuring a more comprehensive understanding of Nike's worldwide market.
4. **Emphasize Emerging Sales Channels:**
   Explore new emerging sales channels like social media shopping and virtual reality stores to reach new customers and enhance their online experience.
5. **Sustainability Insights:**
   Add metrics of sustainability into sales data analysis in order to contribute toward Nike's goals regarding sustainability. This will show the trends that eco-friendly products portray in various regions to inform product development and marketing strategies in the future.
6. **Retailer Performance Optimization:**
   Evaluate the effectiveness of specific retailer partnerships and experiment with exclusive product launches or promotions to strengthen collaboration and maximize sales impact.
7. **Customer Sentiment Analysis:**
   Leverage text mining and sentiment analysis techniques to understand customer feedback from online reviews and social media platforms. This will provide actionable insights into product and service improvement.
8. **Dynamic Pricing Models:**
   Analyze sales performance by applying dynamic pricing models at various regions and product categories to establish the best pricing strategies.

By integrating these focus areas, Nike will be able to further build on strengths, work out challenges, and stay ahead in the competitive retail landscape. This forward-looking approach will enable the company to align with evolving market trends and maintain customer loyalty.

Appendix for link to GitHub repository

https://github.com/KrishaKPatel/Insight-Innovators

References

1. https://www.sciencedirect.com/science/article/pii/S105681902300249X
2. https://www.gapinterdisciplinarities.org/res/articles/(1-8)%20A%20COMPARATIVE%20STUDY%20ON%20ONLINE%20SHOPPING%20VS.%20OFFLINE%20SHOPPING.pdf
3. https://arxiv.org/abs/2203.06848
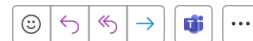4. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed_emp/documents/publication/wcms_835423.pdf

Proofreading with an email from writing Center

Writing Center Tutor: Carpentino, Zoe (E-mail: zcarp2@success.newhaven.edu)

Writing Center Client Report

Carpentino, Zoe <zcarp2@success.newhaven.edu>
To Patel, Krisha Ketankumar

7:18 PM

If there are problems with how this message is displayed, click here to view it in a web browser.

[EXTERNAL SENDER]

University *of* New Haven

Krisha came to me today with a research paper. She asked that I review and give general feedback. First, Krisha showed me the guidelines for the assignment then showed me the powerpoint that went along with the research paper. Krisha asked that I only review one section. I read the section aloud. The only suggestion I had for Krisha was that one sentence could be split into two as it was on the lengthier side. I also suggested that she could look for similar sentences and ensure she is avoidng run-ons. Krisha did not have any additional questions or concerns, so our session ended early. I wish Krisha the best of luck with this assignment and the best of luck with finals!