

# MACHINE LEARNING FINAL PROJECT

## Nike Unisex Footwear Trends: Machine Learning Analysis



Submitted by :

Rachita Harit

Blessy Chinthapalli

Yongija 'Will' Heng

## Executive Summary

In the face of intense competition within the athletic footwear industry, our project embarks on a groundbreaking journey to harness the power of advanced data analytics. This initiative is propelled by the necessity to transcend traditional market research and brand positioning strategies, which have been the cornerstone of Nike's approach to understanding and engaging its diverse consumer base.

### ***Business Questions:***

Our analysis is structured around critical business questions aimed at deepening our understanding of Nike's market positioning and consumer engagement. These include:

1. **Product Range Themes:** What are the prevalent topics or themes across Nike's product range, and how do they influence customer choice?
2. **Description Variation:** How do product descriptions vary across different categories, and what does this imply about Nike's marketing focus?
3. **Premium Product Differentiation:** Which topics are most associated with Nike's premium products, and how are these items differentiated in descriptions?

### ***Data Source:***

The backbone of our analysis is a comprehensive dataset meticulously compiled through web scraping techniques from Nike's official website in our DDR Project.

### ***Solution:***

Our project employs Latent Dirichlet Allocation (LDA) for discovering themes in product descriptions and Long Short-Term Memory (LSTM) networks for classifying texts, marking a shift towards leveraging data insights. The LDA model reveals key marketing language and features in Nike's products, while the LSTM network efficiently categorizes products based on this textual information. This combination of machine learning techniques provides fresh perspectives on Nike's product strategy and customer engagement.

By integrating these findings into Nike's strategy, we aim to improve product development, refine marketing messages, and increase consumer satisfaction.

## **Background and Context**

In the athletic footwear industry, Nike is recognized for its extensive range of products and a strong brand identity. This competitive market is driven by evolving consumer trends and technological advancements, making it crucial for companies like Nike to adapt and innovate continuously. Nike has traditionally leveraged its global presence and marketing prowess to maintain its market position. However, with the advent of data analytics and machine learning, there is an emerging shift towards these technologies to refine marketing strategies and enhance product development.

## **Existing Solutions**

Traditionally, Nike and its competitors in the athletic footwear industry have leaned on conventional strategies like extensive market research, brand positioning, and high-profile marketing campaigns to navigate market challenges and foster customer engagement. These time-tested methods have been pivotal in building strong brand recognition and a loyal customer base.

However, these traditional strategies may not fully leverage the nuanced insights available through advanced data analysis and machine learning technologies.

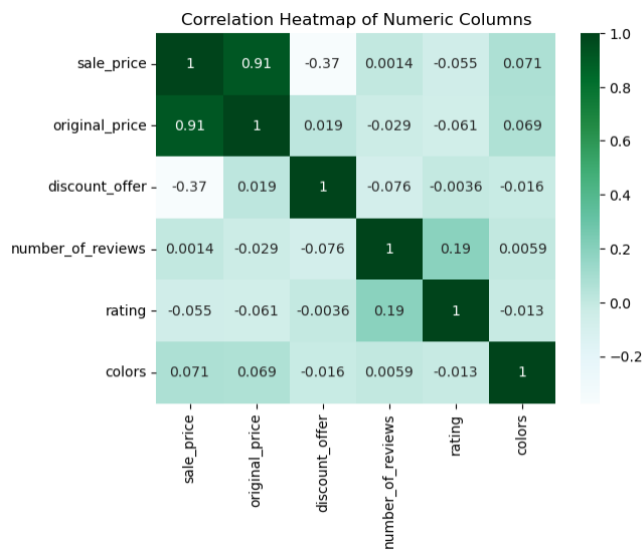
## **Analysis 1: Data Cleaning and Exploratory Data Analysis (EDA)**

In the data cleaning phase, we meticulously refined our dataset to ensure accuracy and relevance for analysis. This involved eliminating duplicates, correcting inconsistencies, and handling missing values, thus laying a solid foundation for our machine learning models to operate on reliable data. With the clean data, we conducted Exploratory Data Analysis (EDA).

Our EDA is a foundational step in understanding the nuances of Nike's unisex footwear data, serving as a precursor to more complex machine learning analyses. Through this process, we sought to uncover

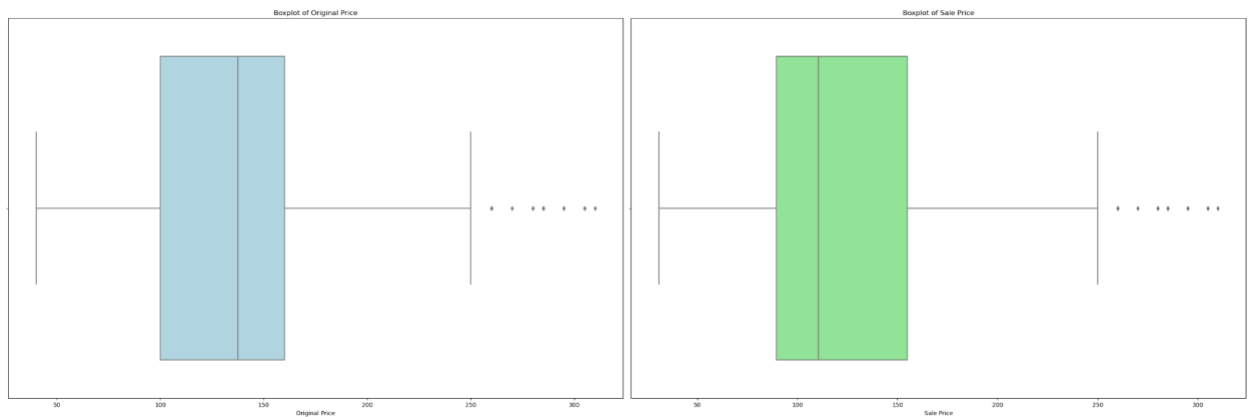
patterns, detect outliers, understand relationships, and generate hypotheses from the dataset obtained via web scraping.

**Correlation Heatmap:**



A strong positive correlation exists between sale and original prices, indicating consistent pricing strategies. Discounts show a negative correlation with sale prices, highlighting their impact on reducing prices. Additionally, a slight positive correlation between the number of reviews and ratings suggests more reviewed products may be better received, though this is not definitive.

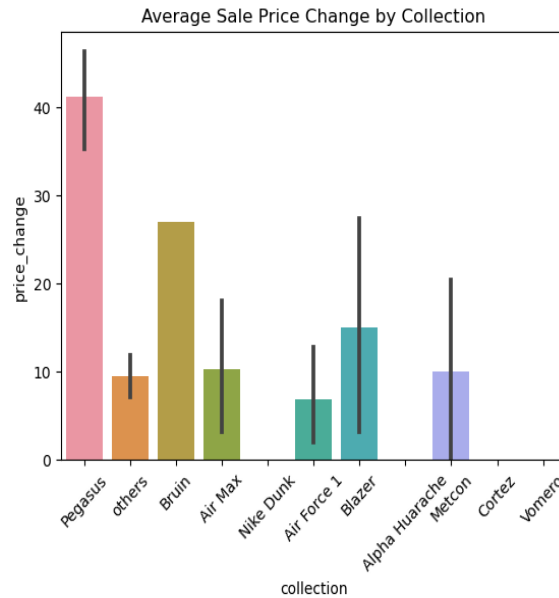
**Boxplot of Sale Price:**



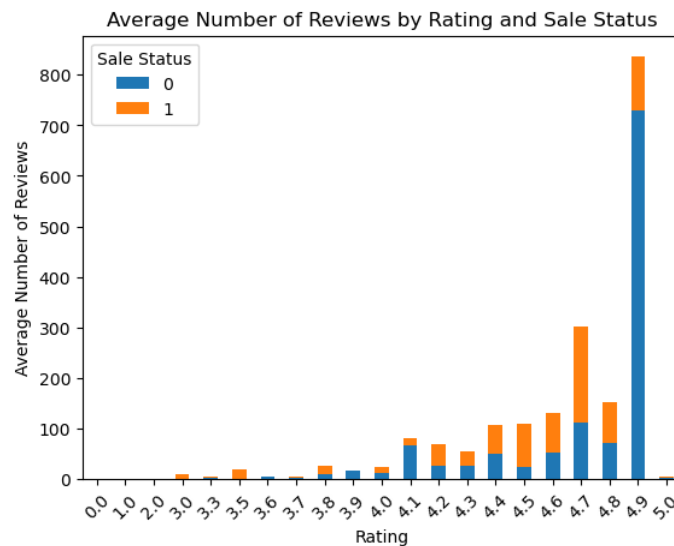
Original prices vary widely, indicating diverse pricing strategies and the presence of premium products

as outliers. The similar medians yet broad price range suggest strategic pricing worth deeper analysis within Nike's marketing strategies.

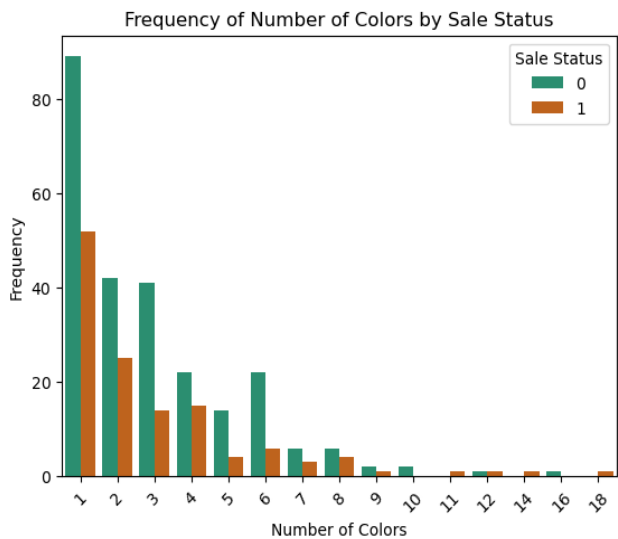
### Bar Charts:



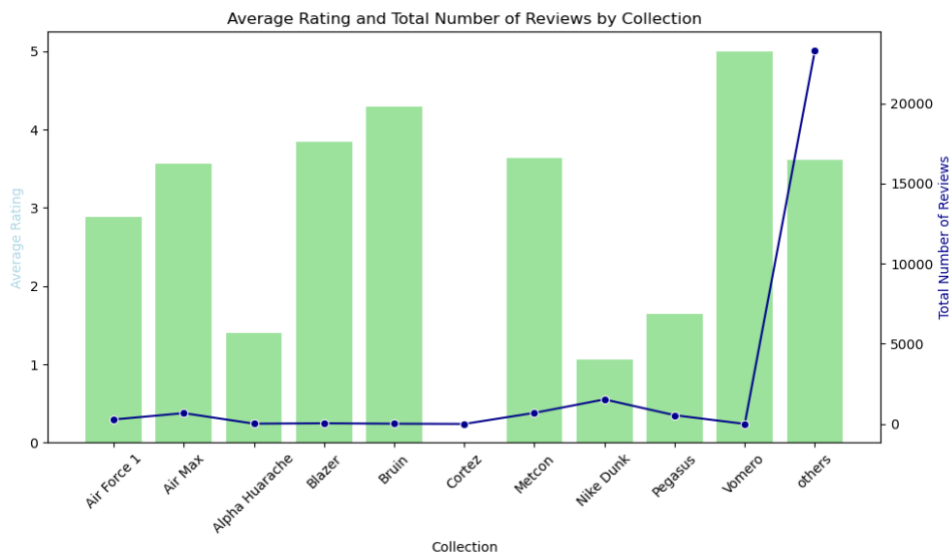
The Pegasus collection has the largest discounting, followed by "Others" and Bruin with less significant price changes. Air Max, Nike Dunk, and Air Force 1 show moderate changes, while Blazer and others have minimal variability. Pegasus sees the greatest fluctuation in discounts.



Products rated 4.9, especially non-sale items, have the most reviews. Sale items generally receive fewer reviews across all ratings, with customer engagement increasing alongside higher ratings, particularly for non-sale products.



Single-color products are most common, especially among non-sale items. As color options increase, product frequency decreases for both sale and non-sale items. Few products offer 10 or more colors. Most products come in 1 to 6 colors, regardless of sale status.



The "Pegasus" and "Others" collections top the average ratings, with "Pegasus" also receiving the most reviews. "Air Max" and "Cortez" have good ratings and a moderate review count. "Blazer" and "Bruin" feature decent ratings but fewer reviews, while "Alpha Huarache" scores lowest in both ratings and reviews, hinting at less positive feedback.

## **Analysis 2: Latent Dirichlet Allocation (LDA) Modeling**

### ***Model Training:***

Using LDA, we analyzed Nike's product descriptions to uncover underlying themes. This process involved preprocessing the text—tokenizing, lemmatizing, and removing stop words—and then creating a dictionary and corpus for analysis. With carefully tuned parameters, the LDA model identified themes related to performance, design, comfort, and materials, providing insights into Nike's product strategy.

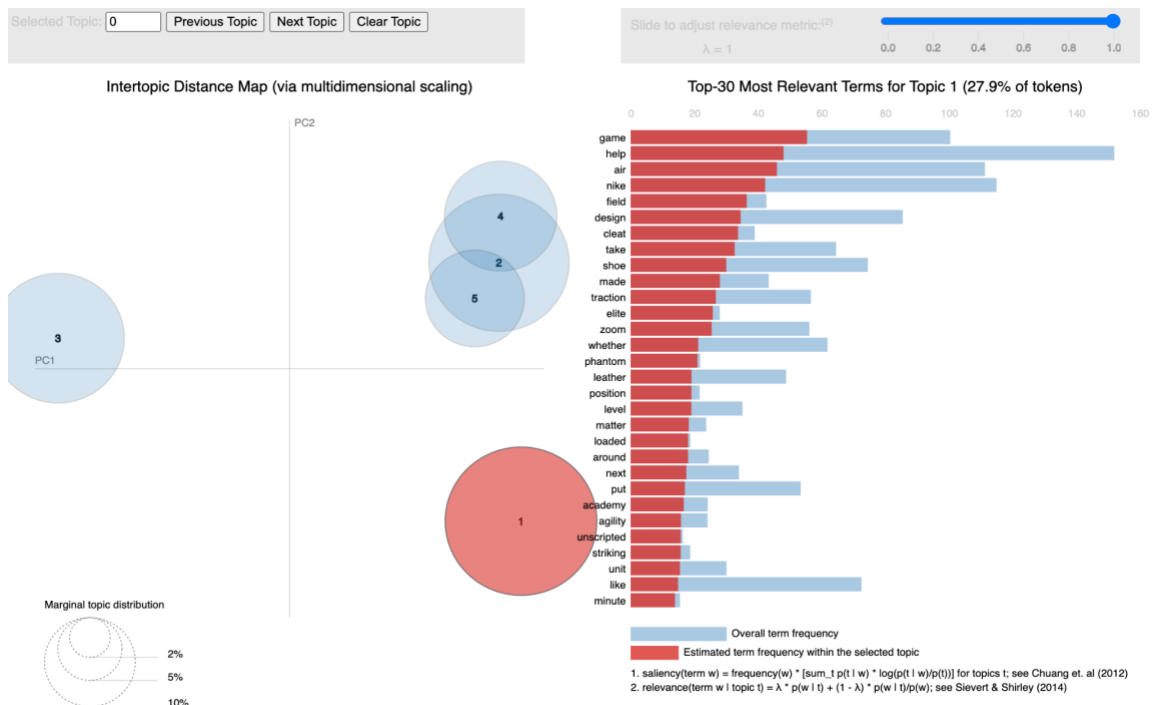
### ***Results of the LDA model:***

- **Topic 0** focuses on general attributes like "game," "design," "color," and "leather," suggesting a broad discussion around sportswear functionality and aesthetics.
- **Topic 1** zeroes in on running, with terms like "run," "speed," "comfort," and "support," indicating a focus on performance and endurance.
- **Topic 2** merges sport-specific terms like "air," "cleat," and "field" with design, implying a blend of performance gear tailored for field sports.
- **Topic 3** highlights technology with "air," "zoom," and "unit," suggesting emphasis on innovation in footwear design for performance.
- **Topic 4** emphasizes comfort and iconic branding with "air," "comfort," "jordan," and "iconic," showcasing Nike's blend of heritage and technology.

### ***LDA Visualization:***

Visual analyses, including the Intertopic Distance Map and Top-30 Terms charts, deepened our thematic insights. The map displayed topic relationships and prevalence, while bar charts highlighted key terms,

such as 'legend' and 'dunk', emphasizing themes of performance and style. The lambda-tuning feature allowed for adjusting term relevance, effectively balancing frequency within topics and the overall corpus.



### LDA Summary:

Theme Analysis: Through LDA, we have discovered specific themes such as performance, aesthetics, comfort, and design details that Nike emphasizes in their product descriptions.

Business Application: These insights are valuable for marketing, highlighting the language that resonates with consumers, and for product development, by identifying potential gaps in the features highlighted in the descriptions.

Future Strategy: By understanding these themes, Nike can tailor their product lines and marketing strategies to match the identified consumer preferences and trends.

### Analysis 3: LSTM Modeling



As part of our effort to classify Nike product offerings, we developed an LSTM model leveraging the Keras framework within TensorFlow, specifically tailored to handle the sequential nature of text data. This model's design was sequential, indicative of its layered architecture that processes data through a series of computational steps, each building upon the last.

### ***Model Structure:***

The model began with an embedding layer, transforming tokenized text into dense vectors, a vital step for the model to grasp linguistic nuances. With a substantial parameter count in the embedding layer alone, the model established a solid base for processing a comprehensive vocabulary embedded within Nike's product descriptions.

Subsequent LSTM layers, equipped to capture temporal dependencies, were complemented by dropout layers aimed at preventing overfitting—a crucial measure for maintaining the model's generalization ability. The presence of multiple LSTM layers signified the model's capacity to understand complex patterns in sequence data.

A final dense layer with 11 units corresponded to the number of product categories, acting as the decision-maker for classification based on learned textual features. The total parameter count exceeded 1.4 million, indicative of the model's sophisticated learning apparatus.

### ***Training Outcomes and Predictive Performance:***

The LSTM model achieved over 90% training accuracy, showcasing its ability to learn effectively from the dataset. However, a lower validation accuracy around 84% and misclassifications, such as mixing up "Air Max" with "Nike Dunk," indicate possible overfitting and the need for model optimization.

The model proved adept at identifying products from specific collections, like "Pegasus," but also underscored the importance of incorporating a broader and more nuanced dataset to enhance predictive accuracy. Achieving an overall accuracy of 89.63%, the LSTM model demonstrated its potential in automating product classification within an e-commerce setting.

## Recommendations and Business Value

### 1. Product Range Themes:

**Recommendation:** Highlight identified themes like performance and design more in marketing and product narratives.

**Business Value:** Enhances customer engagement and influences buying behavior, boosting loyalty and market share.

### 2. Description Variation:

**Recommendation:** Standardize product descriptions to reflect Nike's marketing focus, using insights to spotlight unique features.

**Business Value:** Improves perception and engagement, leading to higher conversion rates.

### 3. Premium Product Differentiation:

**Recommendation:** Emphasize unique attributes of premium products in descriptions and marketing.

**Business Value:** Attracts consumers to higher-priced items by highlighting value, increasing profitability.

### Overall Strategy:

Leveraging machine learning insights enables Nike to align products and marketing with consumer trends, refining offerings and boosting customer experiences. This strategic application of data-driven insights ensures Nike's continued industry leadership and growth.

## Summary and Conclusions

Our project used machine learning to analyze Nike's unisex shoes, identifying key themes, variations in descriptions, and premium product features. We recommended highlighting these themes in marketing, aligning descriptions with objectives, and emphasizing premium product uniqueness. A data-driven approach can boost Nike's customer engagement, optimize marketing, and enhance its premium segment, reinforcing its industry position, driving growth, and ensuring a competitive edge.