Nike Global Sales Data

Nike, a dominant player in the global athletic footwear and apparel market, continuously adapts its pricing strategies to maintain competitiveness and maximize profitability.

This project focuses on analyzing Nike's global sales data from 2024 to uncover key trends, market dynamics, and consumer purchasing behaviors.

By implementing advanced machine learning models, I aim to predict optimal pricing strategies that enhance revenue while maintaining market share.

Data Scientist

I was responsible for collecting and analyzing Nike's 2024 global sales data, building predictive models for price optimization, and developing an interactive dashboard to visualize insights.
My work involved data preprocessing, exploratory analysis, machine learning modeling, and delivering actionable insights to stakeholders.

This project analyzes Nike's 2024 global sales data to uncover pricing trends and optimize revenue.

Using a unified approach, I apply data analytics and machine learning to identify key sales patterns and predict optimal price points.

The goal is to develop data-driven pricing strategies that enhance Nike's market dominance.

Objective :

Analyze Global Sales Data : Identify key trends and patterns in Nike's sales data from 2024. - Understand Market Dynamics : Examine regional demand variations and their impact on sales. Consumer Behavior Insights : Investigate consumer purchasing behaviors and preferences.

Understanding market trends, empowering pricing decisions, driving informed strategies.

Developed an interactive dashboard to analyze and visualize trends in Nike's global sales, incorporating factors such as pricing, regional demand, and consumer behavior. The dashboard enables data-driven pricing decisions, enhances market insights, and supports strategic planning for revenue optimization.

PLATFORM
Visual Studio (Python), Tableau

WORK DONE
Data preprocessing, feature engineering, dashboard creation, machine learning modelling.

TIMEFRAME
2 Weeks

OVERVIEW
The Nike Global Sales Analysis & Machine Learning Prediction for Price Optimization project aims to :

1. Analyze Nike's 2024 global sales data to uncover pricing trends, regional demand variations, and consumer purchasing behaviors.
2. Develop predictive models to optimize pricing strategies and maximize revenue.

Data Preprocessing

Variable Description

Column	Description
Month	Month of the sales activity.
Region	Geographic region (e.g., Greater China, Europe, America).
Main Category	High-level product category (Footwear, Apparel, Equipment).
Sub Category	Specific sub-category within the main category.
Product Line	Specific product series or models.
Price Tier	Pricing segment (Premium, Mid-Range, Budget).
Units Sold	Number of items sold.
Revenue (USD)	Total revenue in US Dollars.
Online Sales Percentage	Percentage of sales through online platforms.
Retail Price	Retail price per unit in USD.

	feature	data_type	null_value(%)	neg_value(%)	0_value(%)	duplicate	n_unique	sample_unique
0	Month	object	0.0	0.0	0.0	0	12	[November, January, October, December, May]
1	Region	object	0.0	0.0	0.0	0	7	[India, Greater China, Japan, Europe, South Ko...
2	Main_Category	object	0.0	0.0	0.0	0	3	[Equipment, Apparel, Footwear]
3	Sub_Category	object	0.0	0.0	0.0	0	11	[Bags, Accessories, Tops, Cricket, Socks]
4	Product_Line	object	0.0	0.0	0.0	0	24	[Gym Sack, Hats, Tech Fleece, Vapor Cricket, P...
5	Price_Tier	object	0.0	0.0	0.0	0	3	[Budget, Mid-Range, Premium]
6	Units_Sold	int64	0.0	0.0	0.0	0	995	[48356, 9842, 25079, 41404, 33569]
7	Revenue_USD	int64	0.0	0.0	0.0	0	996	[14506800, 2066820, 1755530, 8694840, 5371040]
8	Online_Sales_Percentage	int64	0.0	0.0	0.0	0	41	[73, 50, 90, 58, 53]
9	Retail_Price	int64	0.0	0.0	0.0	0	26	[300, 210, 70, 160, 140]
10	Revenue_Zscore	float64	0.0	57.0	0.0	0	996	[2.8976416726702636, -0.9098742386746196, -1.0...

DATASETS

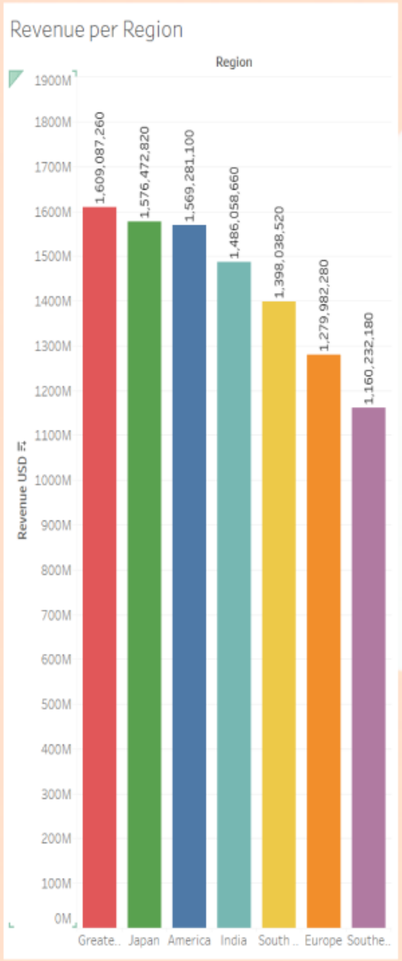
The dataset used in this project is sourced from kaggle and contains Nike's global sales data from 2024,

E.D.A

From the initial data exploration (as shown in the graph below), we observed that all columns have the correct data types, and there are no missing values in the dataset.

- Data Consistency: The dataset is clean, with no null values or duplicates across key features.
- Feature Overview:
  - The Region column contains seven unique geographic markets.
  - Main_Category and Sub_Category provide a structured breakdown of Nike's product offerings.
  - Units_Sold and Revenue_USD represent the key performance metrics.
  - Online_Sales_Percentage highlights the significance of digital sales.
  - Price_Tier categorizes products into Budget, Mid-Range, and Premium segments.

Data Analysis

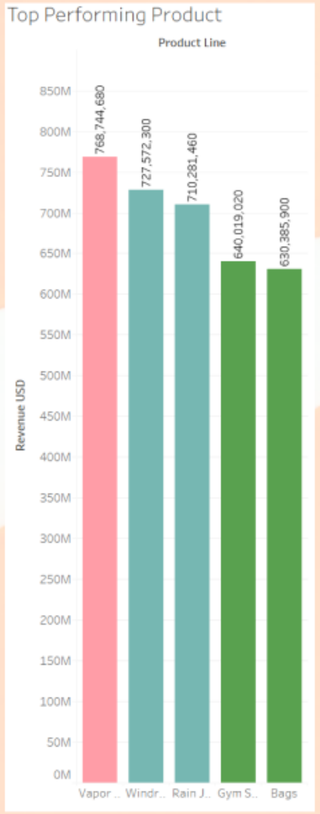


REVENUE ANALYSIS

Greater China: Leads in revenue due to strong brand presence and high demand.
Japan & America: Follow closely, benefiting from brand loyalty and an established market.
India: A rising market, driven by digital adoption and e-commerce growth.
Southeast Asia: Lags in revenue despite significant market potential.
Europe: Underperforms, likely due to economic conditions and competition.
South Korea: Strong sneaker culture, but sales remain low, suggesting market saturation or shifting trends.

OPTIMIZATION STRATEGY:

Nike should implement region-specific approaches based on consumer demand and economic conditions.



PRODUCT PERFORMANCE

Vapor Cricket: Leads with the highest revenue.
Windrunner & Rain Jacket: Follow closely, driven by brand recognition, premium pricing, and seasonal demand.
Gym Sack & Bags: Rank high, indicating strong demand for accessories

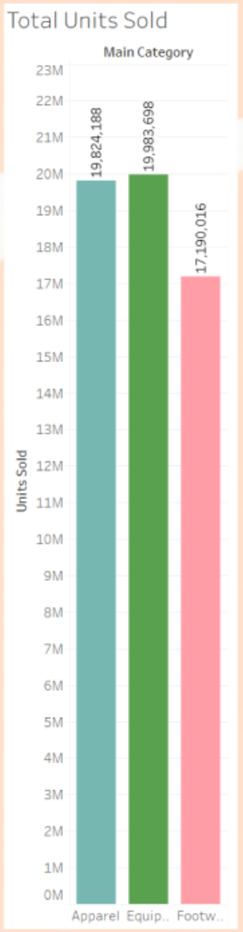
OPTIMIZATION STRATEGY:

Seasonal & Regional Strategy

- Leverage seasonal trends to boost sales for Windrunner & Rain Jacket.
- Tailor product offerings to region-specific demands and preferences.

Pricing & Positioning

- Assess if premium pricing strategy is sustainable or if adjustments can optimize revenue.
- Introduce tiered pricing for different customer segments.



PRODUCT PERFORMANCE

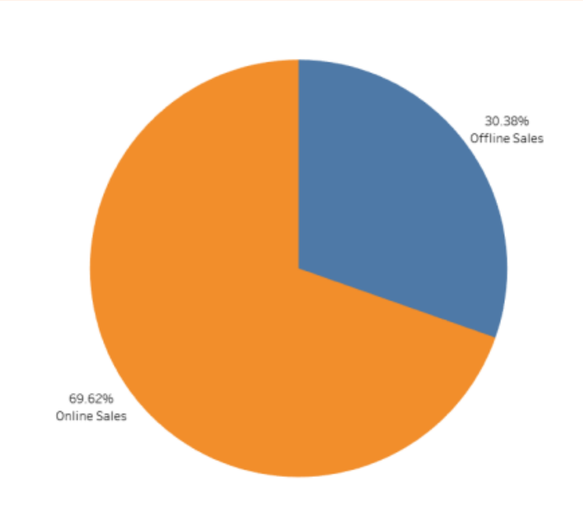
Equipment: Leads in total units sold, indicating strong demand for sports gear.
Apparel: Follows closely, suggesting frequent usage and replacement cycles.
Footwear: Trails behind, likely due to higher price points or longer product lifespans.

OPTIMIZATION STRATEGY:

Stock Optimization

- Adjust stock for Apparel & Equipment to accomodate higher product cycle.
- Increase the product cycle of Apparel and Equipment and product number

Data Analysis

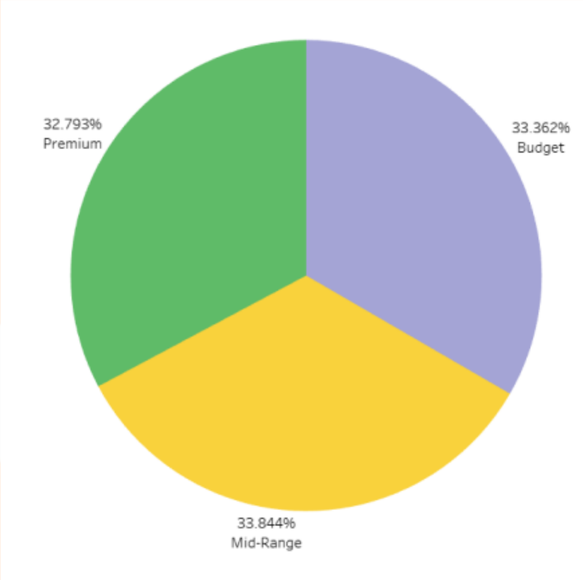


SALES CHANNEL INSIGHT

The data reveals a significant shift toward online sales, which account for 69.62% of total sales

OPTIMIZATION STRATEGY:

- Give incentive into offline store buyer
- Enhance Online Shopping Experience



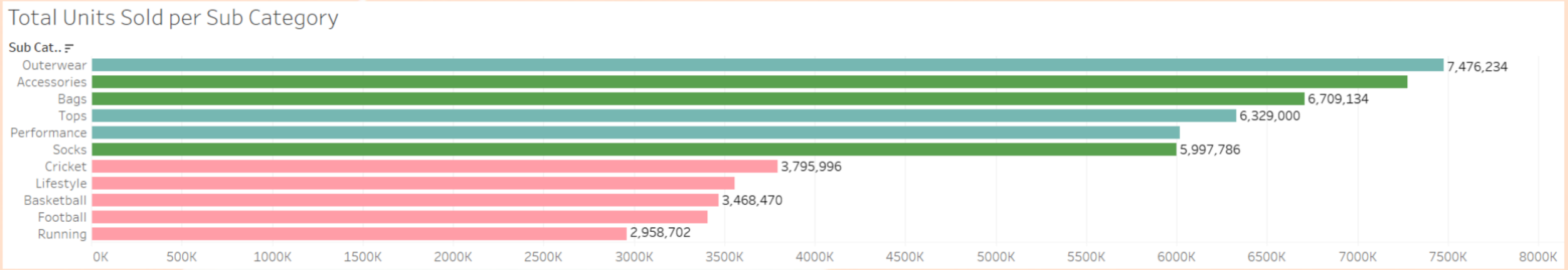
PRICE RANGE DISTRIBUTION

Sales are evenly spread across pricing tiers, with Budget (33.36%), Mid-Range (33.84%), and Premium (32.79%) products contributing almost equally.

OPTIMIZATION STRATEGY:

- Enhance Upselling & Cross-Selling Strategies
- Tiered Product Positioning & Marketing
- Strategic Pricing & Loyalty Programs

Data Analysis



TOTAL UNITS SOLD OVERVIEW

The highest-selling product categories are Outerwear (7.4M units), Accessories (6.7M units), and Bags (6.3M units), indicating strong consumer demand for these items.
Performance wear and Socks also contribute significantly, reinforcing the popularity of sports and activewear.
In contrast, Running, Football, and Basketball categories report lower sales, suggesting a potential opportunity for targeted promotions and product innovations in these segments.

OPTIMIZATION STRATEGY:

- Strengthen High-Performing Categories
- Leverage Growth in Performance Wear & Socks
- Revitalize Running, Football & Basketball Segments

Modelling

```python
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error

# Linear Regression
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
print("MAE (Linear Regression):", mean_absolute_error(y_test, y_pred))
```

✓ 1.6s

```
MAE (Linear Regression): 0.3111823709435364
```

```python
from xgboost import XGBRegressor

# XGBoost
model = XGBRegressor(random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
print("MAE (XGBoost):", mean_absolute_error(y_test, y_pred))
```

✓ 0.1s

```
MAE (XGBoost): 0.0932364008293229
```

Model Performance: Linear Regression vs. XGBoost

- Linear Regression yielded an MAE of 0.3113, providing a baseline but struggling with non-linear interactions in the data. Its assumption of linearity limits accuracy in capturing complex patterns.
- XGBoost achieved a far lower MAE of 0.0932, excelling at modeling non-linear relationships and feature interactions. Its gradient boosting approach enhances predictive accuracy, making it the superior choice for retail price prediction.

Model Selection & Hyperparameter Tuning

```
# Calculate metrics
mae_reduced = mean_absolute_error(y_test, y_pred_reduced)
rmse_reduced = root_mean_squared_error(y_test, y_pred_reduced)  # Updated to use root_mean_squared_error
r2_reduced = r2_score(y_test, y_pred_reduced)

print(f"MAE: {mae_reduced}, RMSE: {rmse_reduced}, R²: {r2_reduced}")
```

✓ 0.0s

```
MAE: 0.06627671368438172, RMSE: 0.08976296368952298, R²: 0.9923260463247945
```

REFINED MODEL PERFORMANCE & KEY INSIGHTS

Key Performance Metrics:
Lower MAE (0.06627) → More Accurate Predictions
- On average, the model's price estimates are off by just 0.066 units, meaning it predicts prices very close to actual values.

RMSE (0.10390) → Consistent but Some Larger Errors Exist
- While most predictions are highly accurate, occasional larger deviations still occur, but they remain within a reasonable range.

High R^2 (0.98972) → Strong Fit
- The model explains 98.97% of price variations, showing that it captures almost all key factors influencing retail prices.

Inventory Dashboard