

# Nike Customer Analysis

Akash Prasad

## Environment Setup

```
# Environment reset
rm(list = ls())          # clear objects
graphics.off()            # close plots
cat("\014")               # clear console
```

## Load Libraries

```
library(tidyverse)
library(janitor)
library(ggplot2)
library(forcats)
library(scales)
library(dplyr)
library(lubridate)
```

---

## Part I - Exploring & Data Cleaning

### Step 1: Read the Product File

```
product <- read_csv("Nike_Sales_Uncleaned.csv")
nrow(product)

## [1] 2500

clean_names(product)

## # A tibble: 2,500 x 13
##   order_id gender_category product_line product_name    size units_sold    mrp
##   <dbl> <chr>          <chr>        <chr>      <chr>     <dbl> <dbl>
## 1 2000 Kids           Training       SuperRep Go M             NA   NA
## 2 2001 Women          Soccer        Tiempo Legend M             3 4958.
## 3 2002 Women          Soccer        Premier III M             4   NA
## 4 2003 Kids           Lifestyle     Blazer Mid L             NA 9674.
## 5 2004 Kids           Running       React Infinity XL            NA   NA
## 6 2005 Women          Training      Flex Trainer M             1 7364.
## 7 2006 Men            Training      SuperRep Go M             NA 6820.
## 8 2007 Kids           Training      SuperRep Go M             1   NA
## 9 2008 Women          Soccer        Premier III 11            -1   NA
## 10 2009 Men           Soccer        Mercurial Super~ M             4   NA
## # i 2,490 more rows
## # i 6 more variables: discount_applied <dbl>, revenue <dbl>, order_date <chr>,
## #   sales_channel <chr>, region <chr>, profit <dbl>
```

```
colnames(product)
```

```
## [1] "Order_ID"          "Gender_Category" "Product_Line"      "Product_Name"
## [5] "Size"               "Units_Sold"        "MRP"                "Discount_Applied"
## [9] "Revenue"            "Order_Date"        "Sales_Channel"     "Region"
## [13] "Profit"
```

## Understanding Data

```
# Understanding numeric data spread and identifying irrelevant columns
summary(product)
```

```
##      Order_ID   Gender_Category   Product_Line   Product_Name
##  Min.    :2000   Length:2500      Length:2500      Length:2500
##  1st Qu.:2535   Class  :character  Class  :character  Class  :character
##  Median  :3192   Mode   :character  Mode   :character  Mode   :character
##  Mean    :3194
##  3rd Qu.:3841
##  Max.    :4499
##
##      Size        Units_Sold       MRP      Discount_Applied
##  Length:2500      Min.   :-1.000   Min.   :2007   Min.   :0.0000
##  Class  :character  1st Qu.: 0.000   1st Qu.:4039   1st Qu.:0.3200
##  Mode   :character  Median  : 2.000   Median  :6110   Median  :0.6150
##              Mean   : 1.482   Mean   :6040   Mean   :0.6310
##              3rd Qu.: 3.000   3rd Qu.:8022   3rd Qu.:0.9625
##              Max.   : 4.000   Max.   :9996   Max.   :1.2500
##              NA's   :1235    NA's   :1254    NA's   :1668
##      Revenue      Order_Date      Sales_Channel      Region
##  Min.   :-7561.6   Length:2500      Length:2500      Length:2500
##  1st Qu.: 0.0     Class  :character  Class  :character  Class  :character
##  Median  : 0.0     Mode   :character  Mode   :character  Mode   :character
##  Mean    : 274.9
##  3rd Qu.: 0.0
##  Max.   :37169.3
##
##      Profit
##  Min.   :-1199.45
##  1st Qu.:  93.53
##  Median  : 1371.38
##  Mean   : 1376.01
##  3rd Qu.: 2660.64
##  Max.   : 3999.21
##
```

```
length(unique(product$Discount_Applied))
```

```
## [1] 127
```

```
view(product$Size)
```

## Identifying Character Columns

```
view(product$Order_Date)
table(product$Sales_Channel)
```

```

## 
## Online Retail
##    1255    1245

glimpse(select(product, where(is.character)))

## Rows: 2,500
## Columns: 7
## $ Gender_Category <chr> "Kids", "Women", "Women", "Kids", "Kids", "Women", "Me~
## $ Product_Line     <chr> "Training", "Soccer", "Soccer", "Lifestyle", "Running"~
## $ Product_Name     <chr> "SuperRep Go", "Tiempo Legend", "Premier III", "Blazer~
## $ Size             <chr> "M", "M", "M", "L", "XL", "M", "M", "M", "11", "M", "L~
## $ Order_Date       <chr> "2024-03-09", "2024-07-09", NA, "04-10-2024", "2024/09~
## $ Sales_Channel    <chr> "Online", "Retail", "Retail", "Online", "Retail", "Ret~
## $ Region           <chr> "bengaluru", "Hyd", "Mumbai", "Pune", "Delhi", "Delhi"~

table(product$Region)

## 
## Bangalore bengaluru      Delhi      Hyd Hyderabad hyderabad Kolkata      Mumbai
##      216          220        438       136        126       141        417        418
##      Pune
##      388

```

## Data Cleaning Steps

### Step 1: Remove Unwanted Fields

```
product <- select(product, -Profit, -Revenue, -Discount_Applied)
```

### Step 2: Clean Units Sold

```
# Replacing negative and NA with 0
product$Units_Sold <- ifelse((product$Units_Sold < 0) | is.na(product$Units_Sold),
                                0, product$Units_Sold)
```

### Step 3: Currency Conversion

```
# Current approximate rate: 1 USD = 88 INR
product <- product %>%
  mutate(Price_USD = round(MRP / 88, 2))
```

### Step 4: Create Sales Column

```
product <- product %>%
  mutate(Sales = round(Units_Sold * Price_USD, 2))
```

### Step 5: Standardize Region Field

```
product <- product %>%
  mutate(Region = case_when(
    grepl("^beng|^bang", Region, ignore.case = TRUE) ~ "Bengaluru",
    grepl("^hyd", Region, ignore.case = TRUE) ~ "Hyderabad",
    TRUE ~ Region
  ))
```

### Step 6: Create Product Category Column

```
product <- product %>%
  mutate(Product_Category = case_when(
    is.na(Size) ~ NA_character_,
    grepl("^[0-9]+\\.?[0-9]*$", Size) ~ "Footwear",
    grepl("(?i)[smlx]", Size) ~ "Apparel"
  ))
```

### Step 7: Convert Date Field

```
product$Order_Date <- parse_date_time(product$Order_Date,
                                         orders = c("ymd", "dmy"))
```

### Step 8: Drop NA Rows

```
product <- drop_na(product)
```

## Final Dataset

```
nrow(product)

## [1] 743

colnames(product)

##  [1] "Order_ID"          "Gender_Category"   "Product_Line"      "Product_Name"
##  [5] "Size"              "Units_Sold"        "MRP"              "Order_Date"
##  [9] "Sales_Channel"     "Region"           "Price_USD"        "Sales"
## [13] "Product_Category"
```

```
view(product)
```

---

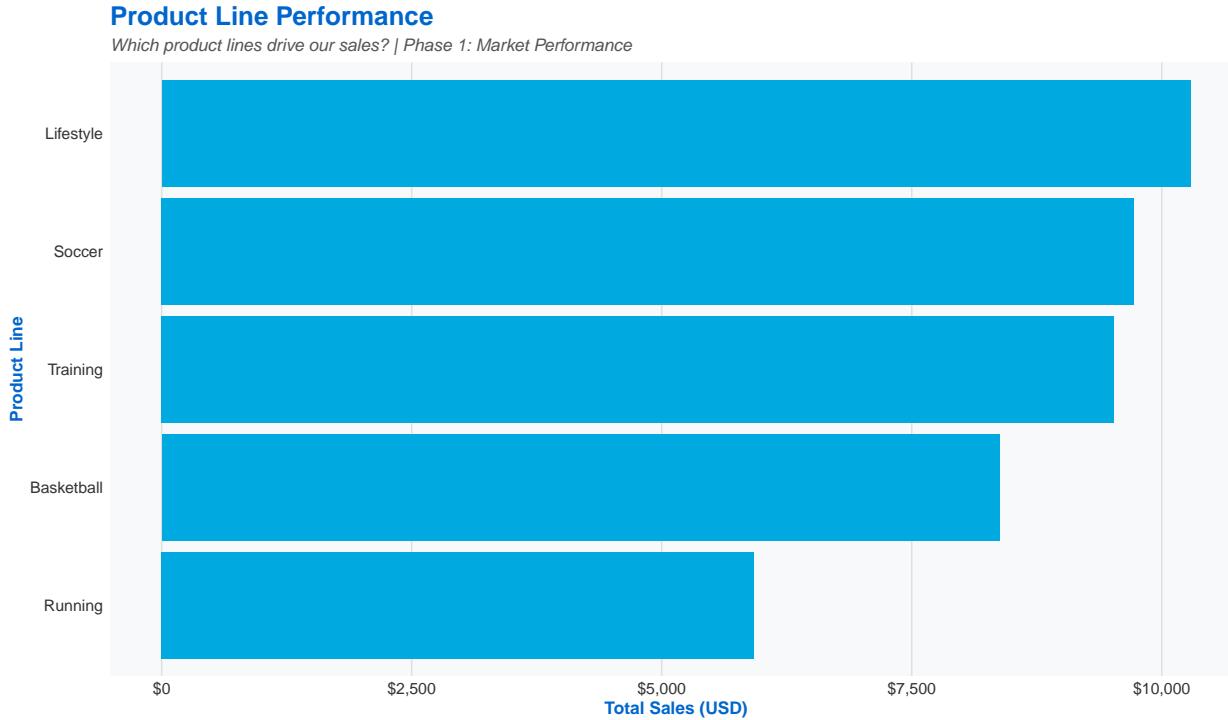
## PHASE I: Market Performance Overview (Blue Theme)

```
phase1_colors <- c("#0066CC", "#00A9EO", "#00BCD4", "#4FC3F7", "#81D4FA",
  "#0277BD", "#0288D1", "#039BE5", "#03A9F4")
```

### Visual 1: Product Line Performance

```
# Aggregate sales by product line
product_line_sales <- product %>%
  group_by(Product_Line) %>%
  summarise(Total_Sales = sum(Sales, na.rm = TRUE)) %>%
  arrange(desc(Total_Sales))

# Create the bar chart
viz1 <- ggplot(product_line_sales, aes(x = reorder(Product_Line, Total_Sales),
                                         y = Total_Sales)) +
  geom_bar(stat = "identity", fill = "#00A9EO") +
  coord_flip() +
  labs(
    title = "Product Line Performance",
    subtitle = "Which product lines drive our sales? | Phase 1: Market Performance",
    x = "Product Line",
    y = "Total Sales (USD)"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = dollar_format(prefix = "$")) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#0066CC"),
    plot.subtitle = element_text(size = 12, color = "#555555", face = "italic"),
    axis.text = element_text(size = 11, color = "#333333"),
    axis.title = element_text(size = 12, face = "bold", color = "#0066CC"),
    panel.grid.major.y = element_blank(),
    panel.grid.minor = element_blank(),
    panel.grid.major.x = element_line(color = "#E0E0E0"),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "#F8F9FA", color = NA)
  )
print(viz1)
```



Visual 2: Product Category Mix

```
# Aggregate sales by product category
category_sales <- product %>%
  group_by(Product_Category) %>%
  summarise(Total_Sales = sum(Sales, na.rm = TRUE)) %>%
  mutate(Percentage = Total_Sales / sum(Total_Sales) * 100)

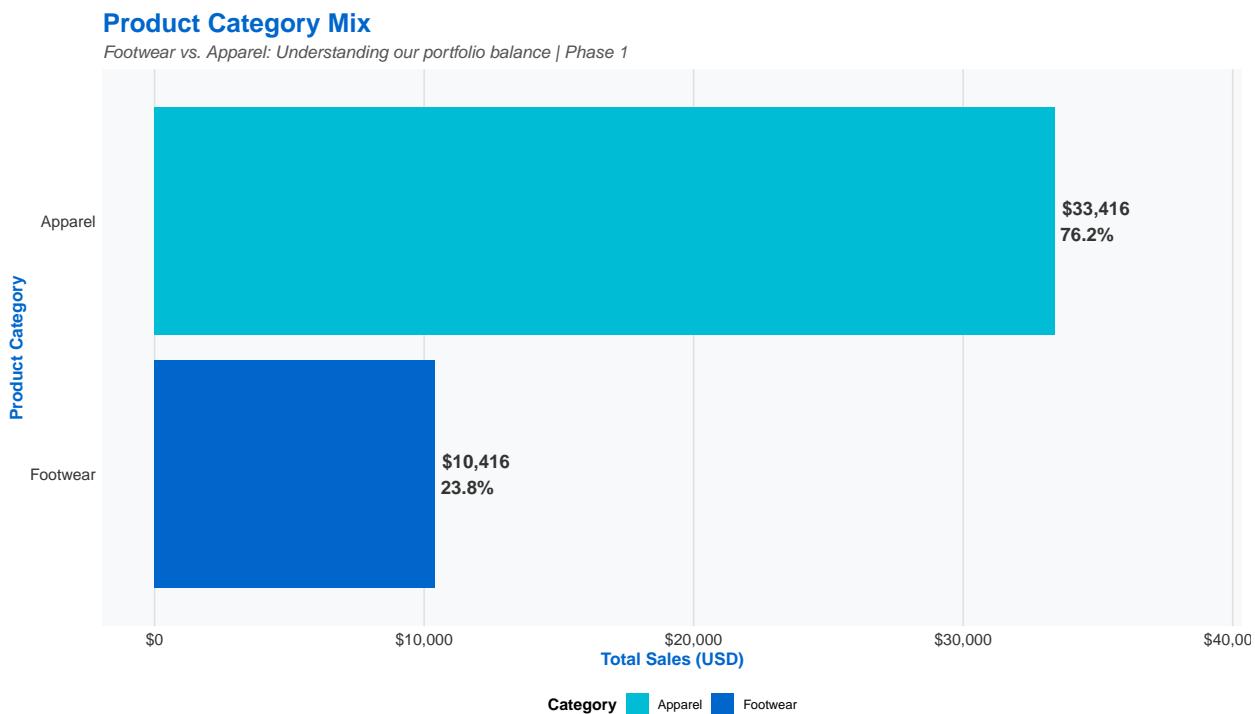
# Create bar chart
viz2 <- ggplot(category_sales, aes(x = reorder(Product_Category, Total_Sales),
                                      y = Total_Sales,
                                      fill = Product_Category)) +
  geom_bar(stat = "identity", show.legend = TRUE) +
  geom_text(aes(label = paste0("$", scales::comma(round(Total_Sales)), "\n",
                           round(Percentage, 1), "%")),
            hjust = -0.1, size = 4.5, fontface = "bold", color = "#333333") +
  coord_flip() +
  labs(
    title = "Product Category Mix",
    subtitle = "Footwear vs. Apparel: Understanding our portfolio balance | Phase 1",
    x = "Product Category",
    y = "Total Sales (USD)",
    fill = "Category"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = dollar_format(prefix = "$"),
                     limits = c(0, max(category_sales$Total_Sales) * 1.15)) +
  scale_fill_manual(values = c("Footwear" = "#0066CC", "Apparel" = "#00BCD4")) +
```

```

theme(
  plot.title = element_text(size = 18, face = "bold", color = "#0066CC"),
  plot.subtitle = element_text(size = 12, color = "#555555", face = "italic"),
  axis.text = element_text(size = 11, color = "#333333"),
  axis.title = element_text(size = 12, face = "bold", color = "#0066CC"),
  legend.position = "bottom",
  legend.title = element_text(face = "bold"),
  panel.grid.major.y = element_blank(),
  panel.grid.minor = element_blank(),
  panel.grid.major.x = element_line(color = "#E0E0E0"),
  plot.background = element_rect(fill = "white", color = NA),
  panel.background = element_rect(fill = "#F8F9FA", color = NA)
)

print(viz2)

```



### Visual 3: Regional Performance

```

# Aggregate sales by region
regional_sales <- product %>%
  group_by(Region) %>%
  summarise(Total_Sales = sum(Sales, na.rm = TRUE)) %>%
  arrange(desc(Total_Sales))

# Create bar chart
viz3 <- ggplot(regional_sales, aes(x = reorder(Region, Total_Sales),
                                      y = Total_Sales)) +

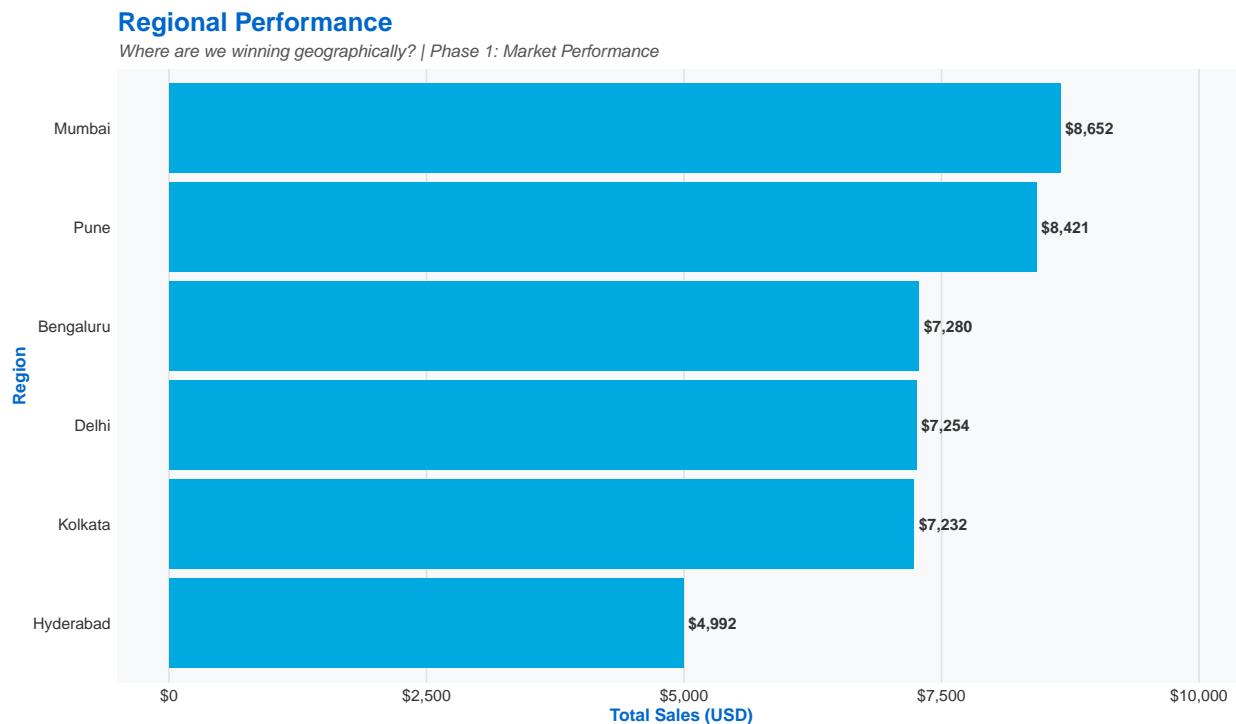
```

```

geom_bar(stat = "identity", fill = "#00A9E0") +
  geom_text(aes(label = paste0("$", scales::comma(round(Total_Sales)))), 
            hjust = -0.1, size = 4, fontface = "bold", color = "#333333") +
  coord_flip() +
  labs(
    title = "Regional Performance",
    subtitle = "Where are we winning geographically? | Phase 1: Market Performance",
    x = "Region",
    y = "Total Sales (USD)"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = dollar_format(prefix = "$"),
                     limits = c(0, max(regional_sales$Total_Sales) * 1.15)) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#0066CC"),
    plot.subtitle = element_text(size = 12, color = "#555555", face = "italic"),
    axis.text = element_text(size = 11, color = "#333333"),
    axis.title = element_text(size = 12, face = "bold", color = "#0066CC"),
    panel.grid.major.y = element_blank(),
    panel.grid.minor = element_blank(),
    panel.grid.major.x = element_line(color = "#E0E0E0"),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "#F8F9FA", color = NA)
  )
)

print(viz3)

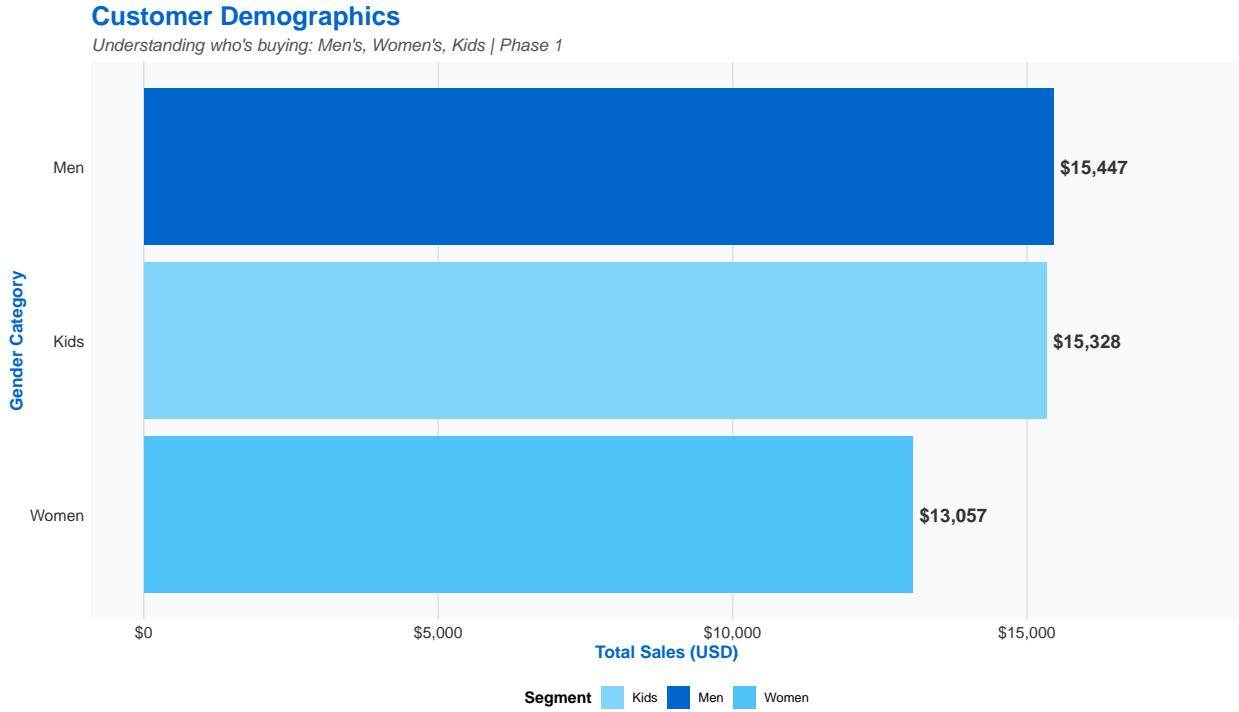
```



## Visual 4: Customer Demographics

```
# Aggregate sales by gender category
gender_sales <- product %>%
  group_by(Gender_Category) %>%
  summarise(Total_Sales = sum(Sales, na.rm = TRUE)) %>%
  arrange(desc(Total_Sales))

# Create bar chart
viz4 <- ggplot(gender_sales, aes(x = reorder(Gender_Category, Total_Sales),
                                    y = Total_Sales,
                                    fill = Gender_Category)) +
  geom_bar(stat = "identity", show.legend = TRUE) +
  geom_text(aes(label = paste0("$", scales::comma(round(Total_Sales)))),
            hjust = -0.1, size = 4.5, fontface = "bold", color = "#333333") +
  coord_flip() +
  labs(
    title = "Customer Demographics",
    subtitle = "Understanding who's buying: Men's, Women's, Kids | Phase 1",
    x = "Gender Category",
    y = "Total Sales (USD)",
    fill = "Segment"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = dollar_format(prefix = "$"),
                     limits = c(0, max(gender_sales$Total_Sales) * 1.15)) +
  scale_fill_manual(values = c("Men" = "#0066CC",
                               "Women" = "#4FC3F7",
                               "Kids" = "#81D4FA")) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#0066CC"),
    plot.subtitle = element_text(size = 12, color = "#555555", face = "italic"),
    axis.text = element_text(size = 11, color = "#333333"),
    axis.title = element_text(size = 12, face = "bold", color = "#0066CC"),
    legend.position = "bottom",
    legend.title = element_text(face = "bold"),
    panel.grid.major.y = element_blank(),
    panel.grid.minor = element_blank(),
    panel.grid.major.x = element_line(color = "#EOE0EO"),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "#F8F9FA", color = NA)
  )
print(viz4)
```



Visual 5: Channel Effectiveness

```
# Aggregate sales by sales channel
channel_sales <- product %>%
  group_by(Sales_Channel) %>%
  summarise(Total_Sales = sum(Sales, na.rm = TRUE)) %>%
  arrange(desc(Total_Sales))

# Create bar chart
viz5 <- ggplot(channel_sales, aes(x = reorder(Sales_Channel, Total_Sales),
                                      y = Total_Sales,
                                      fill = Sales_Channel)) +
  geom_bar(stat = "identity", show.legend = TRUE) +
  geom_text(aes(label = paste0("$", scales::comma(round(Total_Sales)))),
            hjust = -0.1, size = 4.5, fontface = "bold", color = "#333333") +
  coord_flip() +
  labs(
    title = "Channel Effectiveness",
    subtitle = "How are customers buying: Online vs. Retail? | Phase 1",
    x = "Sales Channel",
    y = "Total Sales (USD)",
    fill = "Channel"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = dollar_format(prefix = "$"),
                     limits = c(0, max(channel_sales$Total_Sales) * 1.15)) +
  scale_fill_manual(values = c("Online" = "#00A9E0", "Retail" = "#0066CC")) +
  theme(
```

```

plot.title = element_text(size = 18, face = "bold", color = "#0066CC"),
plot.subtitle = element_text(size = 12, color = "#555555", face = "italic"),
axis.text = element_text(size = 11, color = "#333333"),
axis.title = element_text(size = 12, face = "bold", color = "#0066CC"),
legend.position = "bottom",
legend.title = element_text(face = "bold"),
panel.grid.major.y = element_blank(),
panel.grid.minor = element_blank(),
panel.grid.major.x = element_line(color = "#E0E0E0"),
plot.background = element_rect(fill = "white", color = NA),
panel.background = element_rect(fill = "#F8F9FA", color = NA)
)

print(viz5)

```

### Channel Effectiveness

How are customers buying: Online vs. Retail? | Phase 1




---

## PHASE II: Channel Strategy (Purple Theme)

Visual 6: Channel vs Product Category

```

# Aggregate units sold by channel and product category
channel_category <- product %>%
  group_by(Sales_Channel, Product_Category) %>%
  summarise(Total_Units = sum(Units_Sold, na.rm = TRUE), .groups = 'drop')

```

```

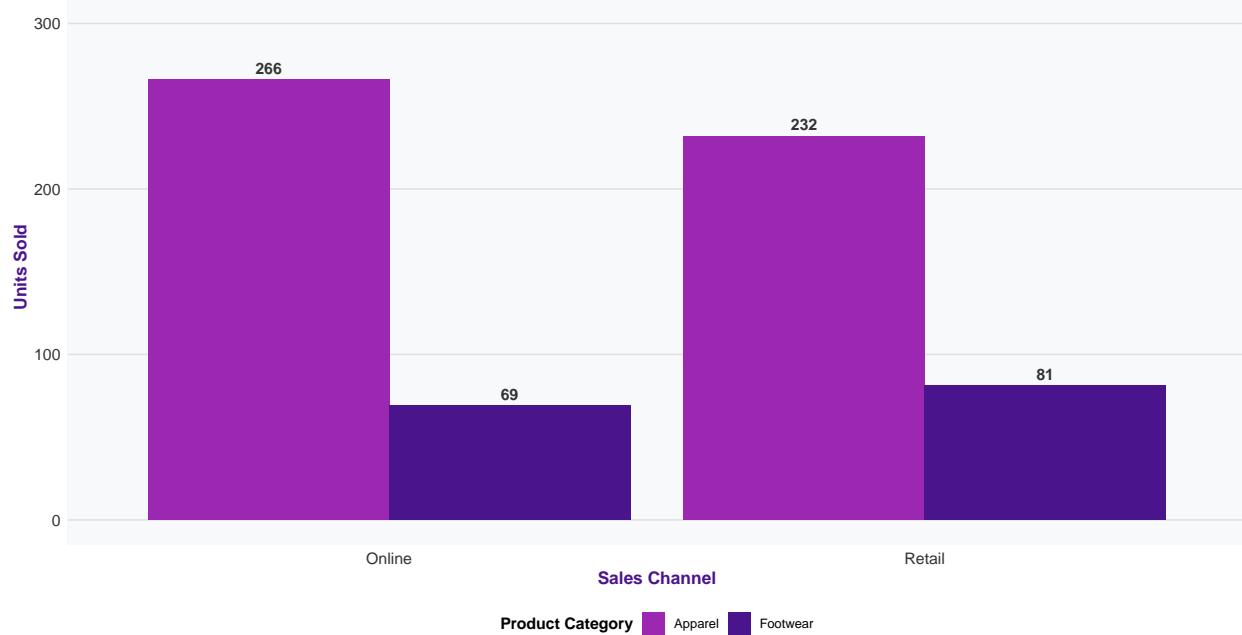
# Create grouped bar chart
viz6 <- ggplot(channel_category, aes(x = Sales_Channel,
                                         y = Total_Units,
                                         fill = Product_Category)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = scales::comma(Total_Units)),
            position = position_dodge(width = 0.9),
            vjust = -0.5, size = 4, fontface = "bold", color = "#333333") +
  labs(
    title = "Channel-Product Fit",
    subtitle = "Which channels work best for which products? | Phase 2: Channel Strategy",
    x = "Sales Channel",
    y = "Units Sold",
    fill = "Product Category"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = comma_format(),
                     limits = c(0, max(channel_category$Total_Units) * 1.15)) +
  scale_fill_manual(values = c("Footwear" = "#4A148C", "Apparel" = "#9C27B0")) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#4A148C"),
    plot.subtitle = element_text(size = 12, color = "#555555", face = "italic"),
    axis.text = element_text(size = 11, color = "#333333"),
    axis.title = element_text(size = 12, face = "bold", color = "#4A148C"),
    legend.position = "bottom",
    legend.title = element_text(face = "bold"),
    panel.grid.major.x = element_blank(),
    panel.grid.minor = element_blank(),
    panel.grid.major.y = element_line(color = "#E0E0E0"),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "#F8F9FA", color = NA)
  )
}

print(viz6)

```

## Channel–Product Fit

Which channels work best for which products? | Phase 2: Channel Strategy



Visual 7: Channel Performance Over Time

```
# Create monthly trends by channel
channel_trends <- product %>%
  mutate(Month_Year = floor_date(Order_Date, "month")) %>%
  group_by(Month_Year, Sales_Channel) %>%
  summarise(Units = sum(Units_Sold, na.rm = TRUE), .groups = 'drop')

# Create total sales (revenue)
total_sales <- product %>%
  mutate(Month_Year = floor_date(Order_Date, "month")) %>%
  group_by(Month_Year) %>%
  summarise(Total_Sales = sum(Sales, na.rm = TRUE), .groups = 'drop')

# Find max values for scaling
max_units <- max(channel_trends$Units)
max_sales <- max(total_sales$Total_Sales)
scale_factor <- max_units * 1.5 / max_sales

# Create the plot
viz7 <- ggplot() +
  geom_col(data = total_sales,
            aes(x = Month_Year, y = Total_Sales * scale_factor),
            fill = "#BDBDBD", alpha = 0.6, width = 25) +
  geom_line(data = channel_trends,
            aes(x = Month_Year, y = Units, color = Sales_Channel, group = Sales_Channel),
            size = 1.3) +
  geom_point(data = channel_trends,
```

```

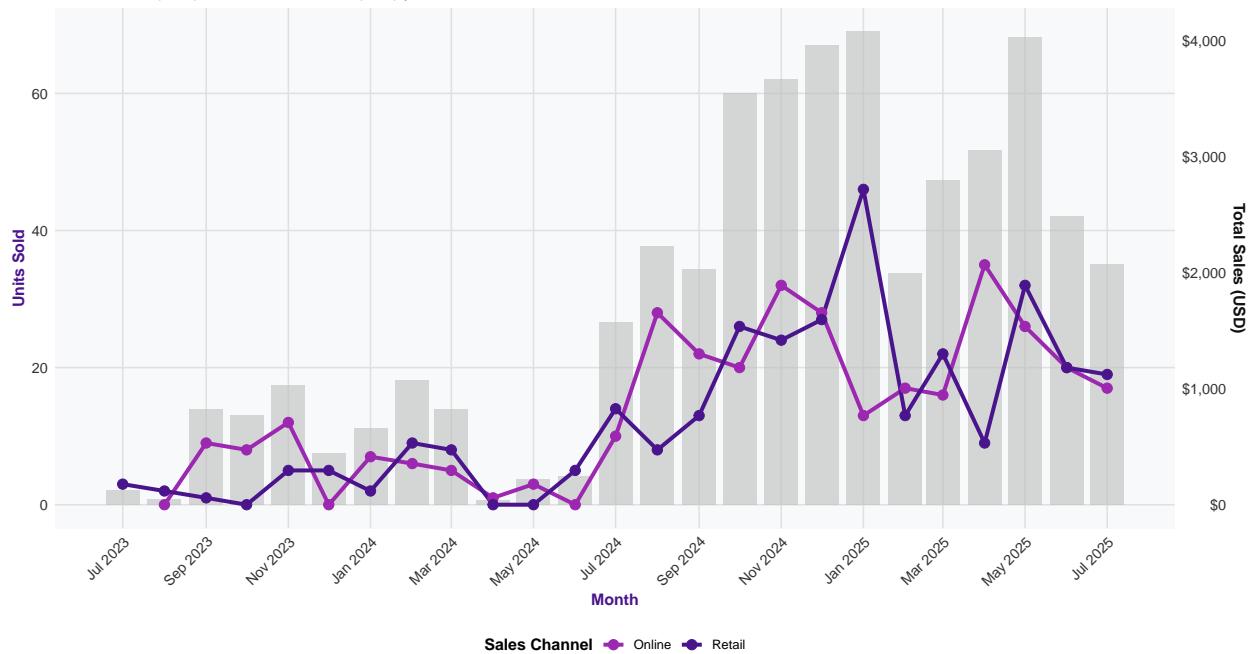
    aes(x = Month_Year, y = Units, color = Sales_Channel),
    size = 3) +
  scale_y_continuous(
    name = "Units Sold",
    labels = comma_format(),
    sec.axis = sec_axis(~ . / scale_factor,
      name = "Total Sales (USD)",
      labels = dollar_format(prefix = "$")))
) +
  labs(
    title = "Channel Performance & Total Sales Over Time",
    subtitle = "Channel units (lines) + Total sales revenue (bars) | Phase 2",
    x = "Month",
    color = "Sales Channel"
) +
  theme_minimal() +
  scale_x_date(date_labels = "%b %Y", date_breaks = "2 months") +
  scale_color_manual(values = c("Online" = "#9C27B0", "Retail" = "#4A148C")) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#4A148C"),
    plot.subtitle = element_text(size = 11, color = "#555555", face = "italic"),
    axis.text = element_text(size = 10, color = "#333333"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    axis.title.y.left = element_text(size = 11, face = "bold", color = "#4A148C"),
    axis.title.y.right = element_text(size = 11, face = "bold", color = "#111111"),
    axis.title.x = element_text(size = 11, face = "bold", color = "#4A148C"),
    legend.position = "bottom",
    legend.title = element_text(face = "bold"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "#E0E0E0"),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "#F8F9FA", color = NA)
  )
)

print(viz7)

```

## Channel Performance & Total Sales Over Time

Channel units (lines) + Total sales revenue (bars) | Phase 2



## PHASE III: Temporal & Pricing (Pink Theme)

### Visual 8: Seasonal Patterns

```
# Extract month and aggregate across all years
seasonal_data <- product %>%
  mutate(Month = month(Order_Date, label = TRUE, abbr = TRUE)) %>%
  group_by(Month, Product_Category) %>%
  summarise(Total_Units = sum(Units_Sold, na.rm = TRUE), .groups = 'drop')

# Create the plot
viz8 <- ggplot(seasonal_data, aes(x = Month, y = Total_Units,
                                      color = Product_Category,
                                      group = Product_Category)) +
  geom_line(size = 1.3) +
  geom_point(size = 3) +
  labs(
    title = "Seasonal Patterns by Product Category",
    subtitle = "Which categories peak in which months? (All years combined) | Phase 3",
    x = "Month",
    y = "Total Units Sold",
    color = "Product Category"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = comma_format())
```

```

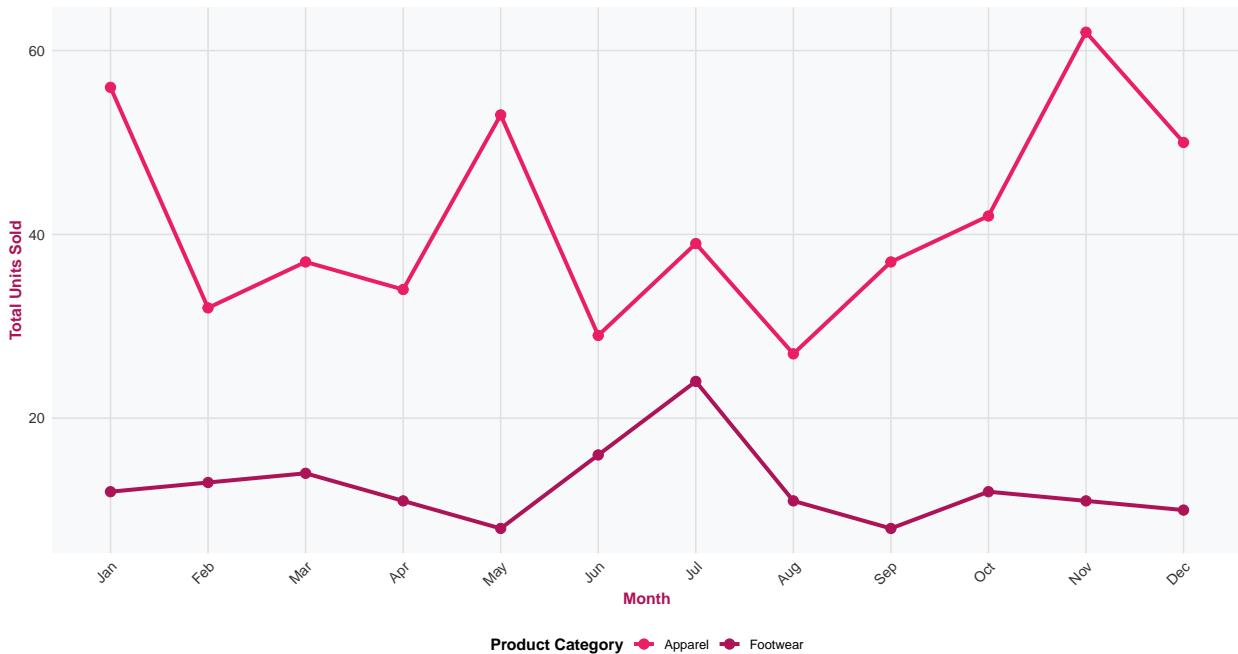
scale_color_manual(values = c(
  "Apparel" = "#E91E63",
  "Footwear" = "#AD1457"
)) +
theme(
  plot.title = element_text(size = 18, face = "bold", color = "#AD1457"),
  plot.subtitle = element_text(size = 11, color = "#555555", face = "italic"),
  axis.text = element_text(size = 10, color = "#333333"),
  axis.text.x = element_text(angle = 45, hjust = 1),
  axis.title.y = element_text(size = 11, face = "bold", color = "#AD1457"),
  axis.title.x = element_text(size = 11, face = "bold", color = "#AD1457"),
  legend.position = "bottom",
  legend.title = element_text(face = "bold"),
  panel.grid.minor = element_blank(),
  panel.grid.major = element_line(color = "#E0E0E0"),
  plot.background = element_rect(fill = "white", color = NA),
  panel.background = element_rect(fill = "#F8F9FA", color = NA)
)

print(viz8)

```

### Seasonal Patterns by Product Category

Which categories peak in which months? (All years combined) | Phase 3



### Visual 9A: Price Range vs Units - Apparel

```

# Create price bins for Apparel
apparel_binned <- product %>%
  filter(Product_Category == "Apparel") %>%
  mutate(Price_Bin = cut(Price_USD,

```

```

        breaks = c(0, 50, 100, 150, 200, 250, 300, Inf),
        labels = c("$0-50", "$50-100", "$100-150",
                   "$150-200", "$200-250", "$250-300", "$300+"),
        include.lowest = TRUE))

apparel_analysis <- apparel_binned %>%
  group_by(Price_Bin) %>%
  summarise(Total_Units = sum(Units_Sold, na.rm = TRUE), .groups = 'drop')

viz9a <- ggplot(apparel_analysis, aes(x = Price_Bin, y = Total_Units)) +
  geom_point(size = 6, color = "#E91E63", alpha = 0.8) +
  labs(
    title = "Price Range vs Units Sold - Apparel",
    subtitle = "Does price affect apparel sales volume? | Phase 3",
    x = "Price Range (USD)",
    y = "Total Units Sold"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = comma_format()) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#E91E63"),
    plot.subtitle = element_text(size = 11, color = "#555555", face = "italic"),
    axis.text = element_text(size = 10, color = "#333333"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    axis.title = element_text(size = 11, face = "bold", color = "#E91E63"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "#E0E0E0"),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "#F8F9FA", color = NA)
  )
}

print(viz9a)

```



Visual 9B: Price Range vs Units - Footwear

```
# Create price bins for Footwear
footwear_binned <- product %>%
  filter(Product_Category == "Footwear") %>%
  mutate(Price_Bin = cut(Price_USD,
    breaks = c(0, 50, 100, 150, 200, 250, 300, Inf),
    labels = c("$0-50", "$50-100", "$100-150",
      "$150-200", "$200-250", "$250-300", "$300+"),
    include.lowest = TRUE))

footwear_analysis <- footwear_binned %>%
  group_by(Price_Bin) %>%
  summarise(Total_Units = sum(Units_Sold, na.rm = TRUE), .groups = 'drop')

viz9b <- ggplot(footwear_analysis, aes(x = Price_Bin, y = Total_Units)) +
  geom_point(size = 6, color = "#AD1457", alpha = 0.8) +
  labs(
    title = "Price Range vs Units Sold - Footwear",
    subtitle = "Does price affect footwear sales volume? | Phase 3",
    x = "Price Range (USD)",
    y = "Total Units Sold"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = comma_format()) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#AD1457"),
    plot.subtitle = element_text(size = 11, color = "#555555", face = "italic"),
    plot.background = element_rect(fill = "#F0F0F0")
  )

```

```

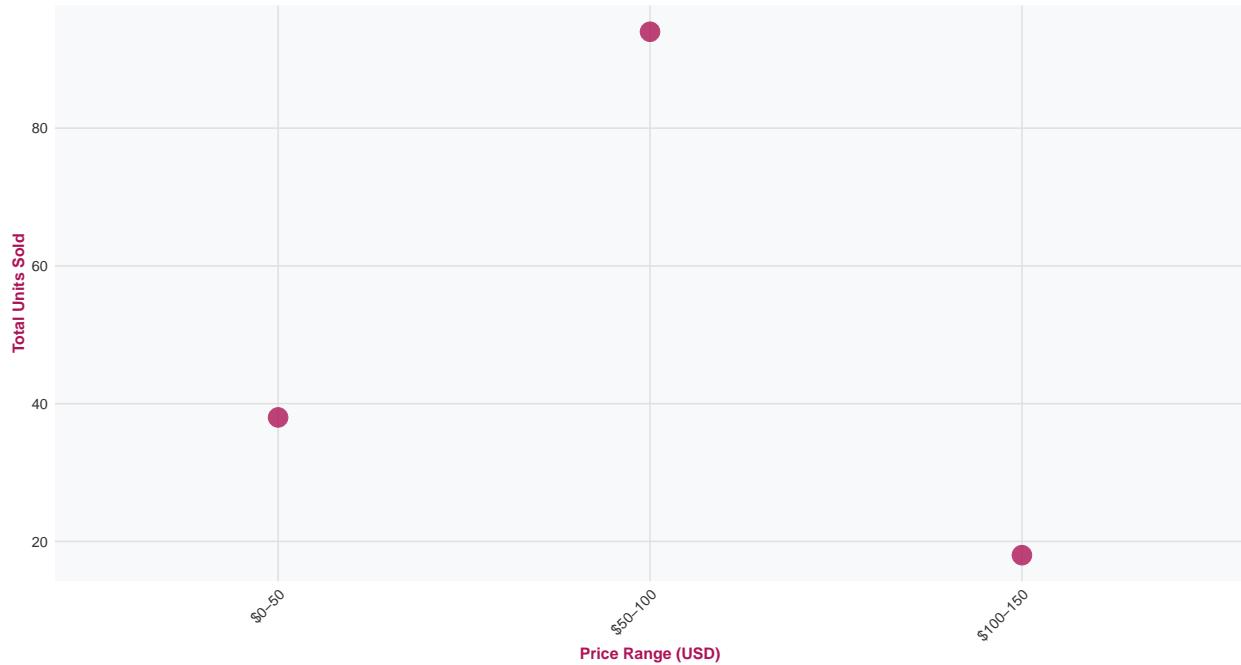
axis.text = element_text(size = 10, color = "#333333"),
axis.text.x = element_text(angle = 45, hjust = 1),
axis.title = element_text(size = 11, face = "bold", color = "#AD1457"),
panel.grid.minor = element_blank(),
panel.grid.major = element_line(color = "#E0E0E0"),
plot.background = element_rect(fill = "white", color = NA),
panel.background = element_rect(fill = "#F8F9FA", color = NA)
)

print(viz9b)

```

### Price Range vs Units Sold – Footwear

Does price affect footwear sales volume? | Phase 3




---

## PHASE IV: Strategic Opportunities (Coral/Orange Theme)

### Visual 10: Gender Performance by Category

```

# Aggregate by Gender_Category and Product_Category
gender_category <- product %>%
  group_by(Gender_Category, Product_Category) %>%
  summarise(Total_Units = sum(Units_Sold, na.rm = TRUE), .groups = 'drop')

# Create grouped bar chart
viz10 <- ggplot(gender_category, aes(x = Product_Category, y = Total_Units,
                                         fill = Gender_Category)) +
  geom_bar(stat = "identity", position = "dodge", width = 0.7) +

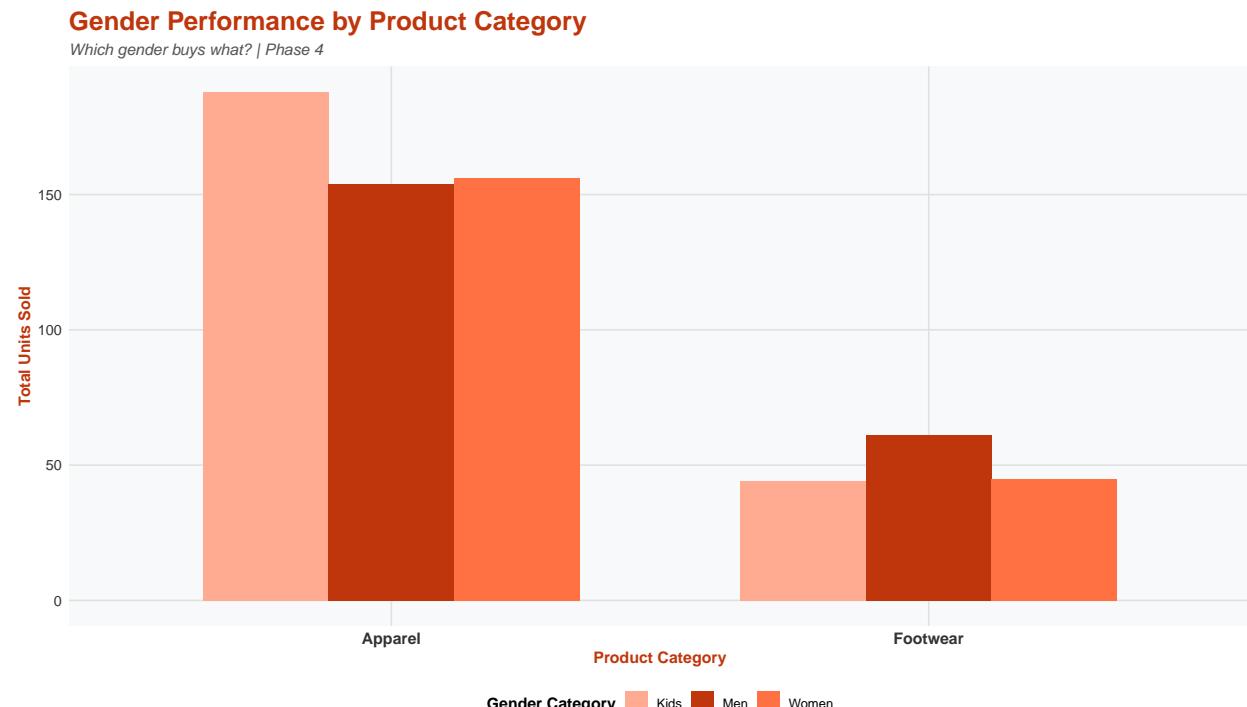
```

```

  labs(
    title = "Gender Performance by Product Category",
    subtitle = "Which gender buys what? | Phase 4",
    x = "Product Category",
    y = "Total Units Sold",
    fill = "Gender Category"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = comma_format()) +
  scale_fill_manual(values = c(
    "Men" = "#BF360C",
    "Women" = "#FF7043",
    "Kids" = "#FFAB91"
  )) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#BF360C"),
    plot.subtitle = element_text(size = 11, color = "#555555", face = "italic"),
    axis.text = element_text(size = 10, color = "#333333"),
    axis.text.x = element_text(size = 11, face = "bold"),
    axis.title = element_text(size = 11, face = "bold", color = "#BF360C"),
    legend.position = "bottom",
    legend.title = element_text(face = "bold"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "#E0E0E0"),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "#F8F9FA", color = NA)
  )
)

print(viz10)

```



## Visual 11: Regional Market Positioning

```
# Aggregate by Region
regional_bubble <- product %>%
  group_by(Region) %>%
  summarise(
    Avg_Price = mean(Price_USD, na.rm = TRUE),
    Total_Sales = sum(Sales, na.rm = TRUE),
    Total_Units = sum(Units_Sold, na.rm = TRUE),
    .groups = 'drop'
  )

# Create bubble chart
viz11 <- ggplot(regional_bubble, aes(x = Avg_Price, y = Total_Sales,
                                         size = Total_Units, color = Total_Units,
                                         label = Region)) +
  geom_point(alpha = 0.8) +
  geom_text(size = 4, fontface = "bold", color = "#8B0000", vjust = -1.8) +
  labs(
    title = "Regional Market Positioning",
    subtitle = "Premium vs Value Markets: Price × Sales × Volume | Phase 4",
    x = "Average Price (USD)",
    y = "Total Sales (USD)",
    size = "Total Units Sold",
    color = "Total Units Sold"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = dollar_format(prefix = "$")) +
  scale_x_continuous(labels = dollar_format(prefix = "$")) +
  scale_size_continuous(range = c(8, 30), labels = comma_format()) +
  scale_color_gradient(
    colors = c("#FFF3E0", "#FFB74D", "#FF7043", "#E64A19", "#BF360C"),
    labels = comma_format(),
    guide = guide_colorbar(barwidth = 1.5, barheight = 10, order = 1)
  ) +
  guides(
    size = guide_legend	override.aes = list(color = "#FF7043"), order = 2
  ) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#BF360C"),
    plot.subtitle = element_text(size = 11, color = "#555555", face = "italic"),
    axis.text = element_text(size = 10, color = "#333333"),
    axis.title = element_text(size = 11, face = "bold", color = "#BF360C"),
    legend.position = "right",
    legend.title = element_text(face = "bold", size = 10),
    legend.text = element_text(size = 9),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "#E0E0E0"),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "#F8F9FA", color = NA)
  )

print(viz11)
```



## Summary

This analysis examined Nike sales data across 4 strategic phases:

- **Phase 1 (Blue):** Market Performance - Product lines, categories, regions, demographics, channels
- **Phase 2 (Purple):** Channel Strategy - Channel-product fit and trends over time
- **Phase 3 (Pink):** Temporal & Pricing - Seasonal patterns and price-volume relationships
- **Phase 4 (Coral/Orange):** Strategic Opportunities - Gender preferences and regional market positioning

All 11 visualizations provide actionable insights for product management and strategic decision-making.