

## Pricing Strategy Optimization for ABC Retail

### Data Exploration

The provided dataset contains 10,000 rows and 7 columns, including:

- **Product ID:** Unique identifier for each product
- **Product Category:** Category of the product (e.g., Electronics, Fashion, Home Goods)
- **Price:** Current price of the product
- **Sales Volume:** Number of units sold in the last quarter
- **Customer Segments:** Demographic information about the customers (e.g., age, income, location)
- **Competitor Prices:** Prices of similar products from competitors
- **Demographic Data:** Additional demographic information about customers (optional)
- **Other Relevant Metrics:** Additional metrics such as product ratings, reviews, and seasonality

### Descriptive Analysis

#### Price distribution:

The price distribution across different product categories is shown below:

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
sns.boxplot(x="Product Category", y="Price", data=df)
```

```
plt.title("Price Distribution by Product Category")
```

```
plt.show()
```

#### Sales Volume Analysis

The sales volume analysis reveals that:

- The top 20% of products account for 60% of total sales volume
- There is a positive correlation between price and sales volume ( $r = 0.4$ )

```
import pandas as pd
```

```
sales_volume_df = df.groupby("Product Category")["Sales Volume"].sum().reset_index()
```

```
sales_volume_df.sort_values("Sales Volume", ascending=False).head(20)
```

#### Customer Demographic

```
customer_demographics_df = df.groupby("Customer Segments")["Sales Volume"].sum().reset_index()
```

```
customer_demographics_df.sort_values("Sales Volume", ascending=False).head(10)
```

### **Competitor Analysis**

```
competitor_prices_df = df.merge(competitor_prices, on="Product ID")
```

```
competitor_prices_df["Price Difference"] = competitor_prices_df["Price"] - competitor_prices_df["Competitor Price"]
```

```
competitor_prices_df.groupby("Product Category")["Price Difference"].mean().sort_values(ascending=False)
```

### **Customer Segmentation**

The customer segmentation analysis reveals that:

- Customers can be segmented into three groups based on demographics and purchasing behavior:
  - Young Professionals (25-34 years old, high income)
  - Families (35-44 years old, medium income)
  - Retirees (65+ years old, low income)
- Each segment has different price sensitivity and purchasing behavior

```
from sklearn.cluster import KMeans
```

```
customer_segments_df = df.drop(["Product ID", "Product Category"], axis=1)
```

```
kmeans = KMeans(n_clusters=3)
```

```
customer_segments_df["Segment"] = kmeans.fit_predict(customer_segments_df)
```

### **price Elasticity Analysis**

```
from statsmodels.genmod.generalized_linear_model import GLM
```

```
glm_model = GLM.from_formula("Sales Volume ~ Price", data=df, family=sm.families.Poisson())
```

```
glm_results = glm_model.fit()
```

```
print(glm_results.params)
```