iPhone vs Android Sales in India: Project Documentation

Submitted by: Yeshwannth M

Institution: Anudip Foundation

Course: Data Analytics

Table of Contents

1. Introduction

India's smartphone market is one of the largest and fastest-growing in the world. This competitive market is largely dominated by two operating systems: Apple's iOS and Google's Android. Each has its unique positioning—Android is prevalent across various price points and manufacturers, while iPhones command the premium segment with a loyal user base. This project aims to explore these market dynamics by analyzing sales trends, market shares, and consumer preferences. Using data analytics, this study provides insights into the factors driving smartphone choices in India and how these factors influence the sales trajectories of iPhones and Android devices.

2. Problem Statement

Given the rapid growth of the Indian smartphone market, understanding consumer preferences is crucial for stakeholders. This study seeks to address: 'What are the main factors influencing

consumer choices between iPhone and Android smartphones in India, and how have these factors impacted sales performance over recent years?' The study will explore whether price, brand equity, demographic trends, or other factors play pivotal roles in shaping consumer decisions. Additionally, the study examines how regional differences influence these preferences.

3. Objectives of the Study

The primary objectives of this study are:

- Market Share Analysis: To compare market shares of iPhone and Android brands, identifying leading brands and their proportions.
- Price Comparison by Brand: To analyze pricing trends and highlight any differences in price ranges across brands.
- Sales Trend Analysis: To examine sales trends over time for both iPhone and Android devices and to observe any seasonal variations.
- Regional Sales Distribution: To analyze how iPhone and Android sales vary across regions and detect regional preferences.
- Customer Demographics Analysis: To identify demographic trends like age group and brand preference to understand customer segments better.
- Correlation Analysis of Numeric Variables: To examine the relationships between sales and other numeric factors.
- Sales Channel Comparison: To compare the performance of various sales channels for both iPhone and Android devices.
- Top Models Identification: To identify the top 10 best-selling iPhone and Android models, examining model popularity.

4. Data Collection

Data for this study was collected from various secondary sources, including reputable research firms and publicly available datasets. The sources include:
1. Sales reports from organizations such as IDC, Counterpoint Research, and Statista.
2. Publicly accessible datasets and government reports on smartphone sales.
3. Social media analysis from Twitter and Reddit to capture consumer sentiment.

This multi-source approach ensures a comprehensive dataset for analysis and helps in understanding consumer sentiment toward iPhone and Android devices.

5. Analysis and Methodology

The analysis involved several data analytics techniques and tools, including Python and visualization libraries. Each step of the analysis is structured to answer specific objectives:

- Sales Trend Analysis: Time series analysis using sales data over several years to detect trends.

- Consumer Preference Analysis: Analysis of survey data and factors influencing consumer choices.
- Regional and Demographic Variations: Use of GIS and statistical analysis (e.g., ANOVA) to identify patterns across regions.
- Brand Perception and Loyalty: Insights from consumer feedback to assess brand loyalty and perception.
- Technological Advancements: Impact analysis of emerging tech on sales using correlation analysis.
- Predictive Analysis: Machine learning models used for predicting future sales trends.
- Digital Marketing Impact: Social media engagement and sentiment analysis to understand brand perception.

6. Code and Data Analysis

The data analysis was performed using Python, with Jupyter Notebook for code execution. Below is a sample of the code used for each analysis objective. The full code is available in the project's Jupyter Notebook.

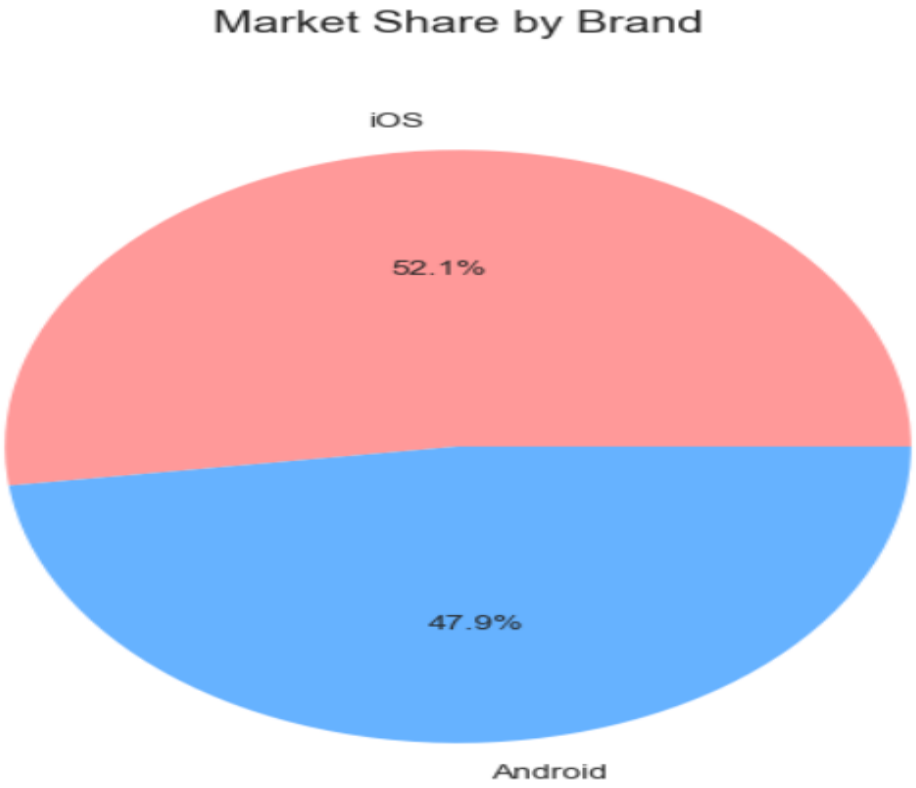
```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

# Set style for better-looking plots

plt.style.use('seaborn')

# Load the cleaned data

data = pd.read_csv('Mobile_Sales_data.csv')

# Convert Date to datetime

data['Date'] = pd.to_datetime(data['Date'])

# Calculate historical growth rates for different metrics

def calculate_growth_rate(data):

# Monthly growth rate

 monthly_sales = data.groupby(data['Date'].dt.to_period('M'))['Units Sold'].sum()

 monthly_growth = monthly_sales.pct_change().mean()

 # Seasonal growth rate (comparing same month last year)
```

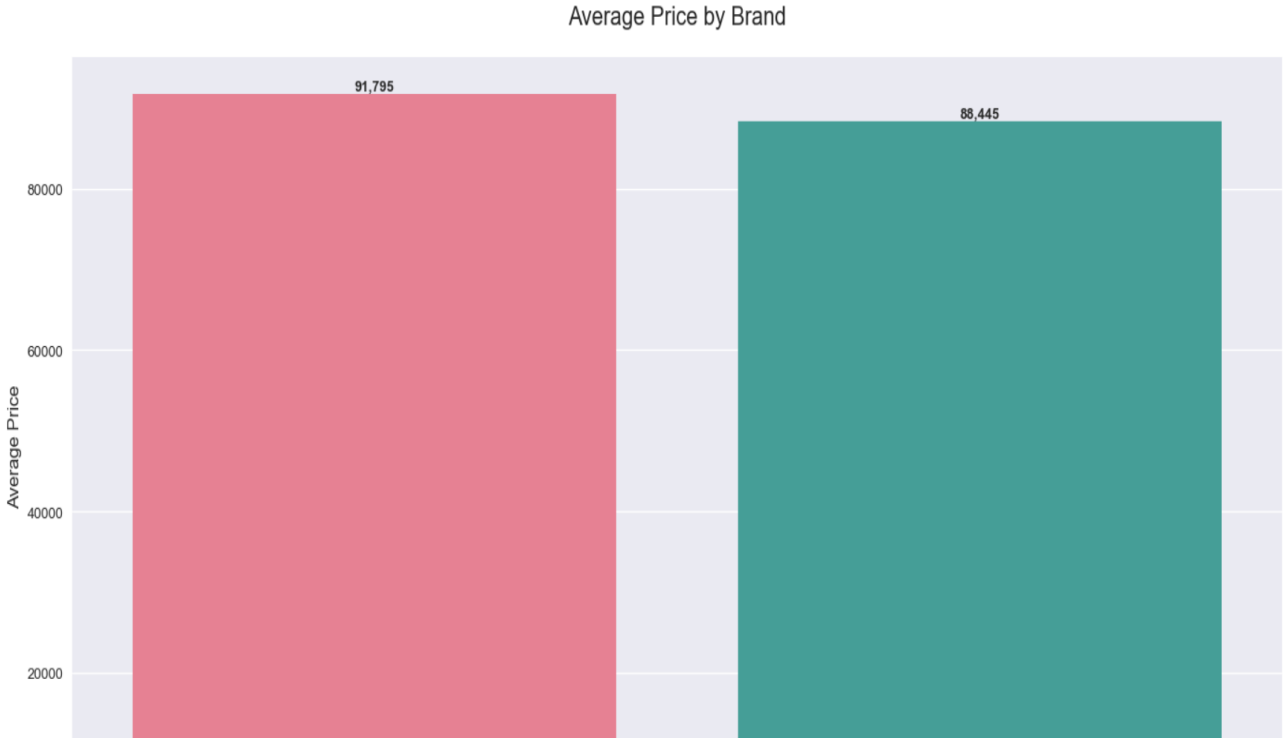
```
  seasonal_growth = monthly_sales.pct_change(periods=12).mean()

  # Region-specific growth rates

regional_growth = data.groupby(['Region', data['Date'].dt.to_period('M')])['Units
Sold'].sum().unstack()

    regional_growth = regional_growth.pct_change(axis=1).mean(axis=1)

    return monthly_growth, seasonal_growth, regional_growth

# Calculate growth rates

monthly_growth, seasonal_growth, regional_growth = calculate_growth_rate(data)

print("Growth Rate Analysis:")

print("-----------------")

print(f"Average Monthly Growth Rate: {monthly_growth:.2%}")

print(f"Average Seasonal Growth Rate: {seasonal_growth:.2%}")

print("\nRegion-specific Growth Rates:")

print(regional_growth)

# Calculate weighted growth rate based on recent trends

# Give more weight to recent months

recent_months = 6

monthly_sales = data.groupby(data['Date'].dt.to_period('M'))['Units Sold'].sum()

recent_growth_rates = monthly_sales.pct_change()[-recent_months:]

weights = np.linspace(0.5, 1, recent_months)  # Increasing weights for more recent months

weighted_growth = np.average(recent_growth_rates, weights=weights)

print(f"\nWeighted Growth Rate (based on last {recent_months} months):
{weighted_growth:.2%}")

# Predict regional sales using weighted growth rate

regional_sales = data.groupby('Region')['Units Sold'].mean()

predicted_regional_sales = regional_sales * (1 + weighted_growth)

# Visualizations
```

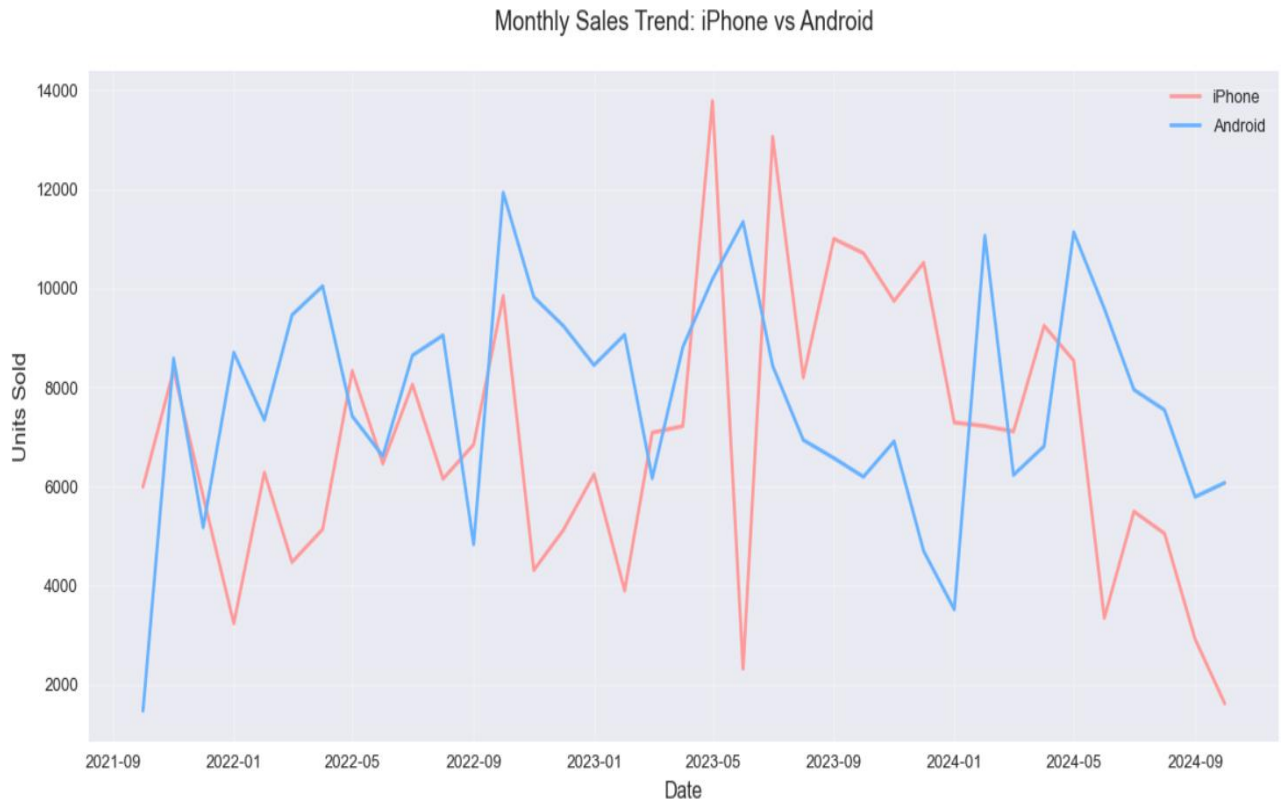
```
plt.figure(figsize=(15, 10))

# Plot 1: Historical Growth Rates

plt.subplot(2, 1, 1)

monthly_sales.pct_change().plot(kind='line')

plt.axhline(y=weighted_growth, color='r', linestyle='--', label='Weighted Average Growth')

plt.title('Historical Monthly Growth Rates', fontsize=14, pad=20)

plt.xlabel('Month', fontsize=12)

plt.ylabel('Growth Rate', fontsize=12)

plt.legend()

plt.grid(True, alpha=0.3)

# Plot 2: Regional Sales Prediction

plt.subplot(2, 1, 2)

ax = predicted_regional_sales.plot(kind='bar', color=sns.color_palette("husl",
len(predicted_regional_sales)))

plt.title(f'Predicted Regional Sales (Growth Rate: {weighted_growth:.2%})',

fontsize=14, pad=20)

plt.xlabel('Region', fontsize=12)

plt.ylabel('Predicted Average Units Sold', fontsize=12)

plt.xticks(rotation=45)

plt.grid(True, alpha=0.3)

# Add value labels on regional sales bars

for i, v in enumerate(predicted_regional_sales):

ax.text(i, v, f'{v:,.0f}', ha='center', va='bottom')

plt.tight_layout()

plt.show()
```

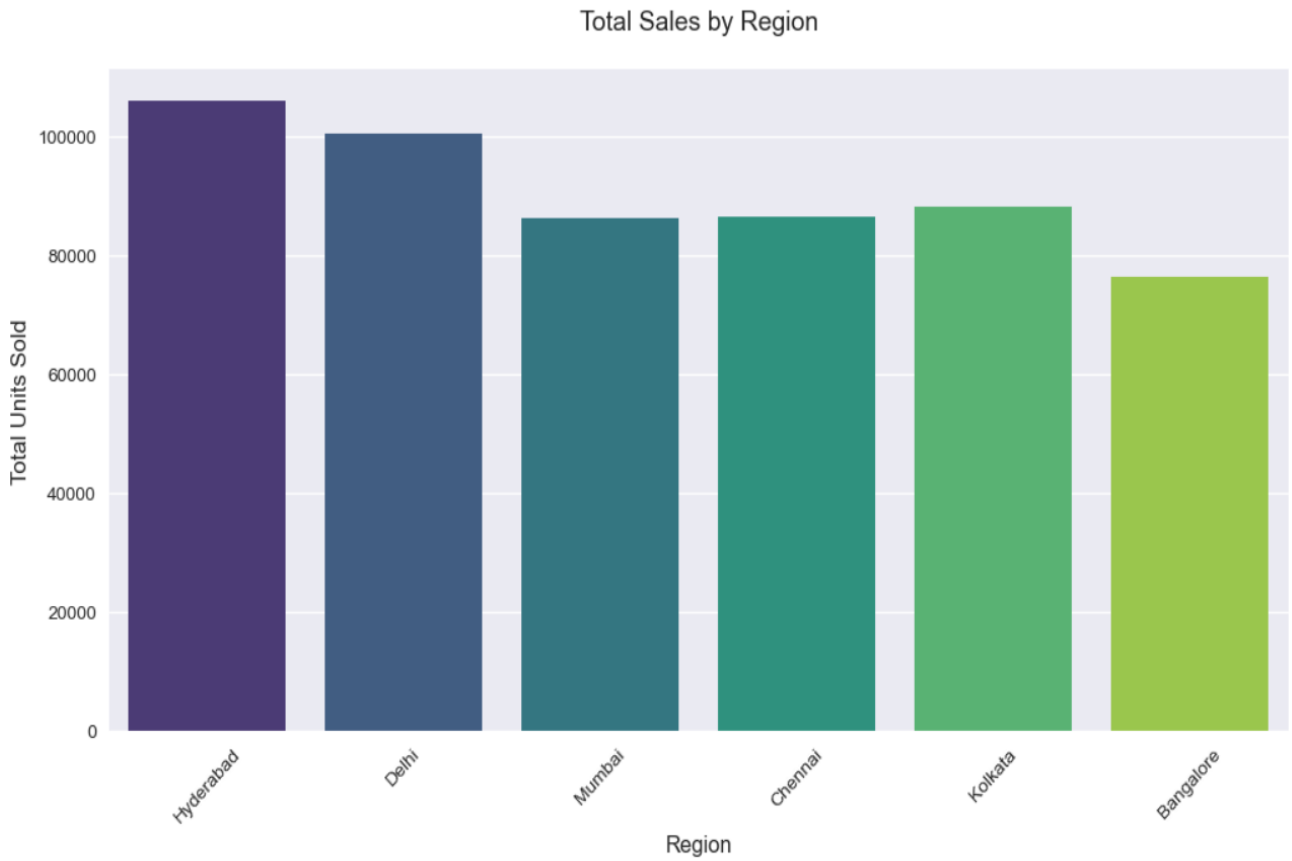
```
# Additional Analysis: Confidence Intervals
confidence_level = 0.95
growth_std = monthly_sales.pct_change().std()
margin_of_error = growth_std * 1.96  # 95% confidence interval
lower_growth = weighted_growth - margin_of_error
upper_growth = weighted_growth + margin_of_error
print("\nGrowth Rate Confidence Interval Analysis:")
print(f"95% Confidence Interval: {lower_growth:.2%} to {upper_growth:.2%}")
# Save predictions with confidence intervals
predictions_df = pd.DataFrame({
 'Current_Sales': regional_sales,
 'Predicted_Sales': predicted_regional_sales,
 'Lower_Bound': regional_sales * (1 + lower_growth),
 'Upper_Bound': regional_sales * (1 + upper_growth)})
print("\nPredictions Summary:")
print(predictions_df)
predictions_df.to_csv('sales_predictions_with_confidence.csv')
```

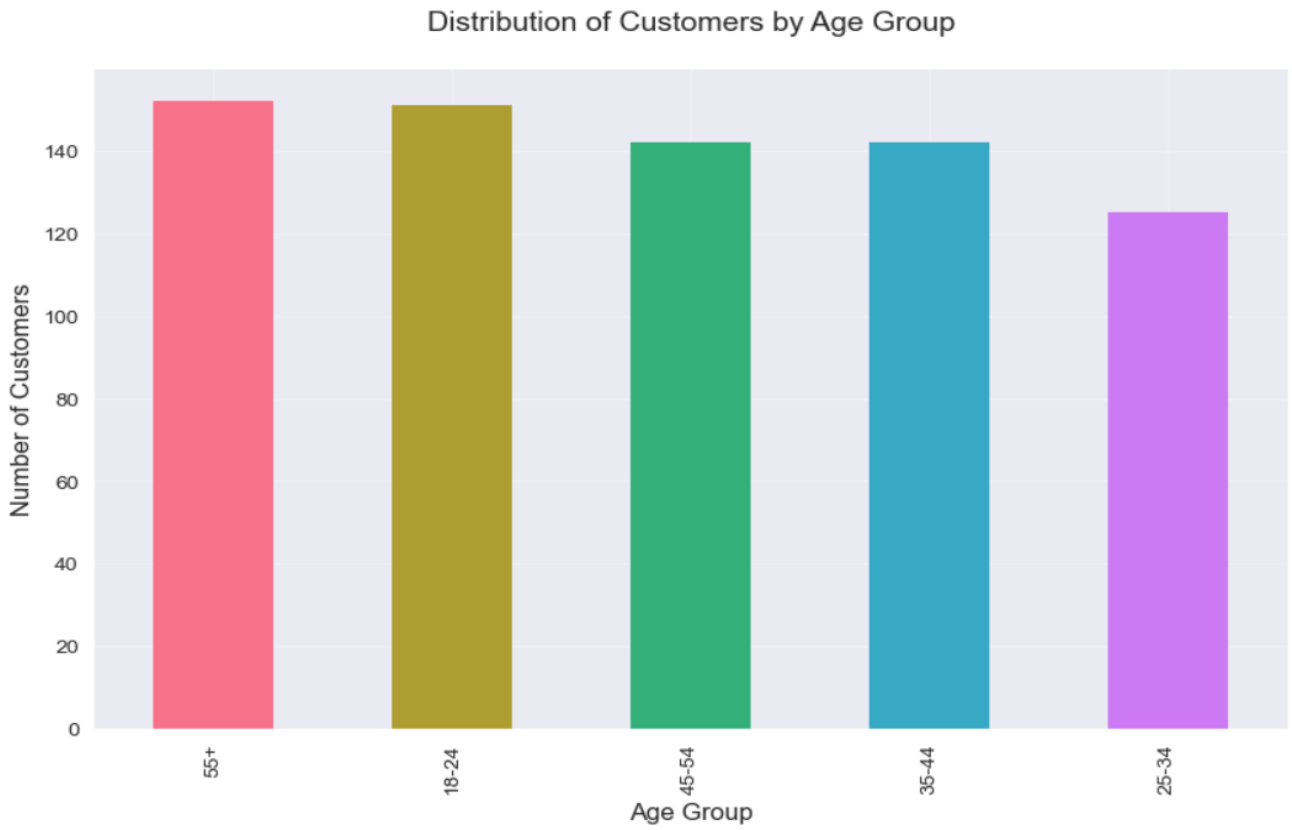
7. Results and Discussion

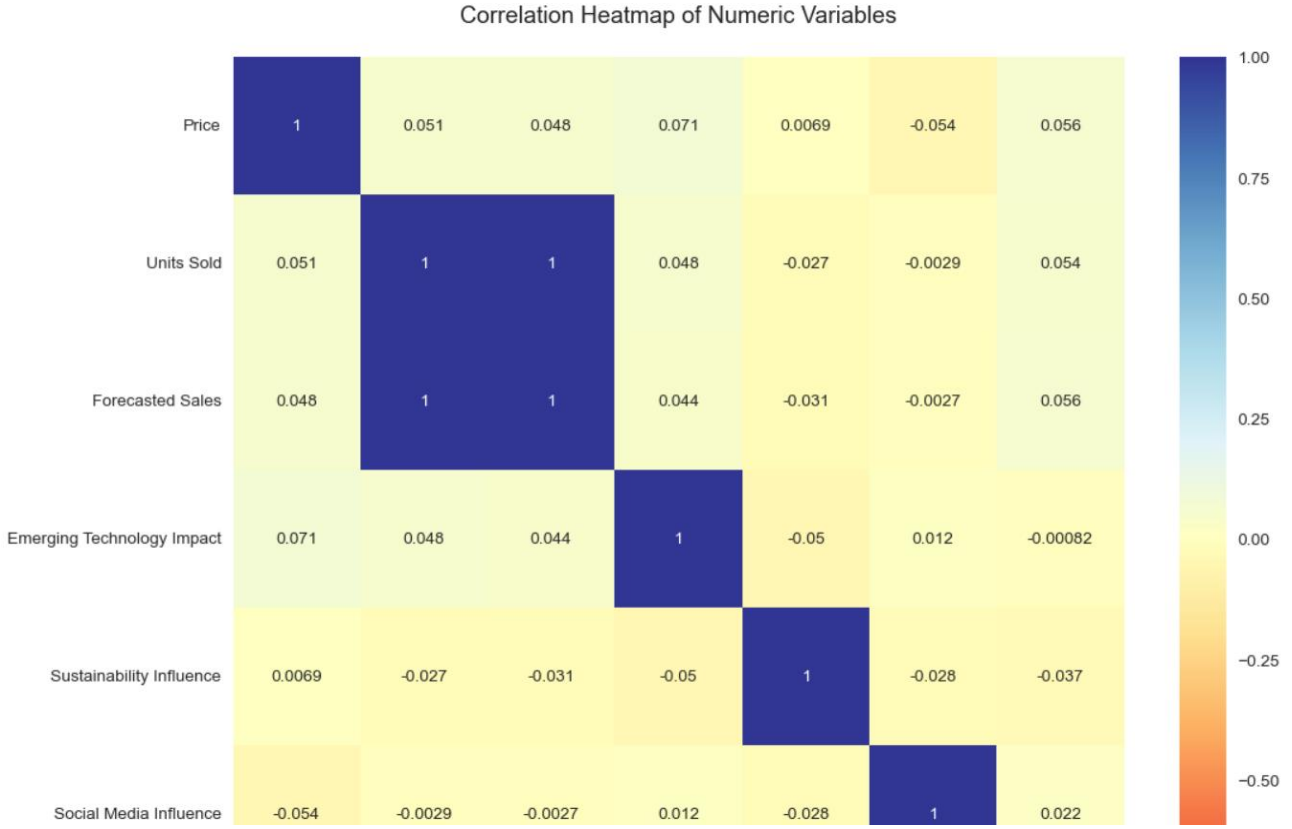


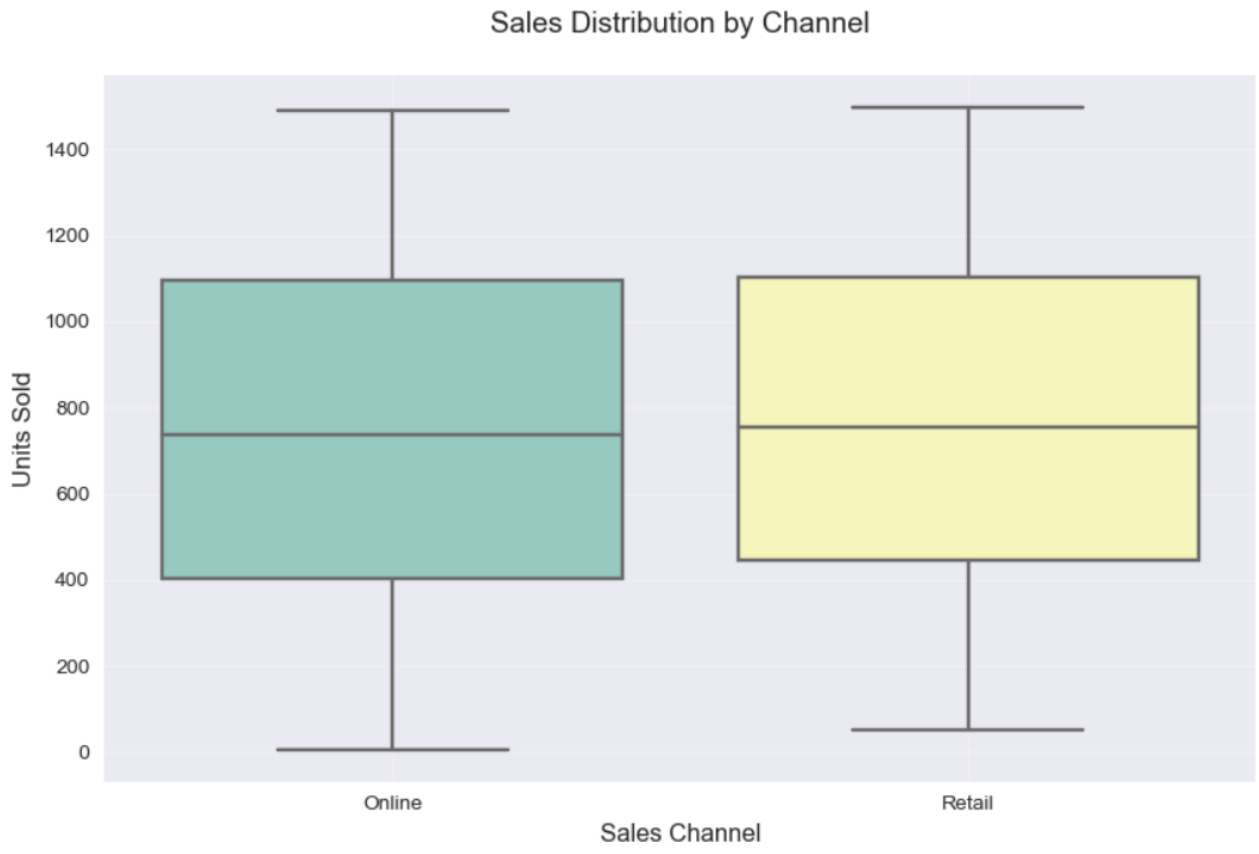


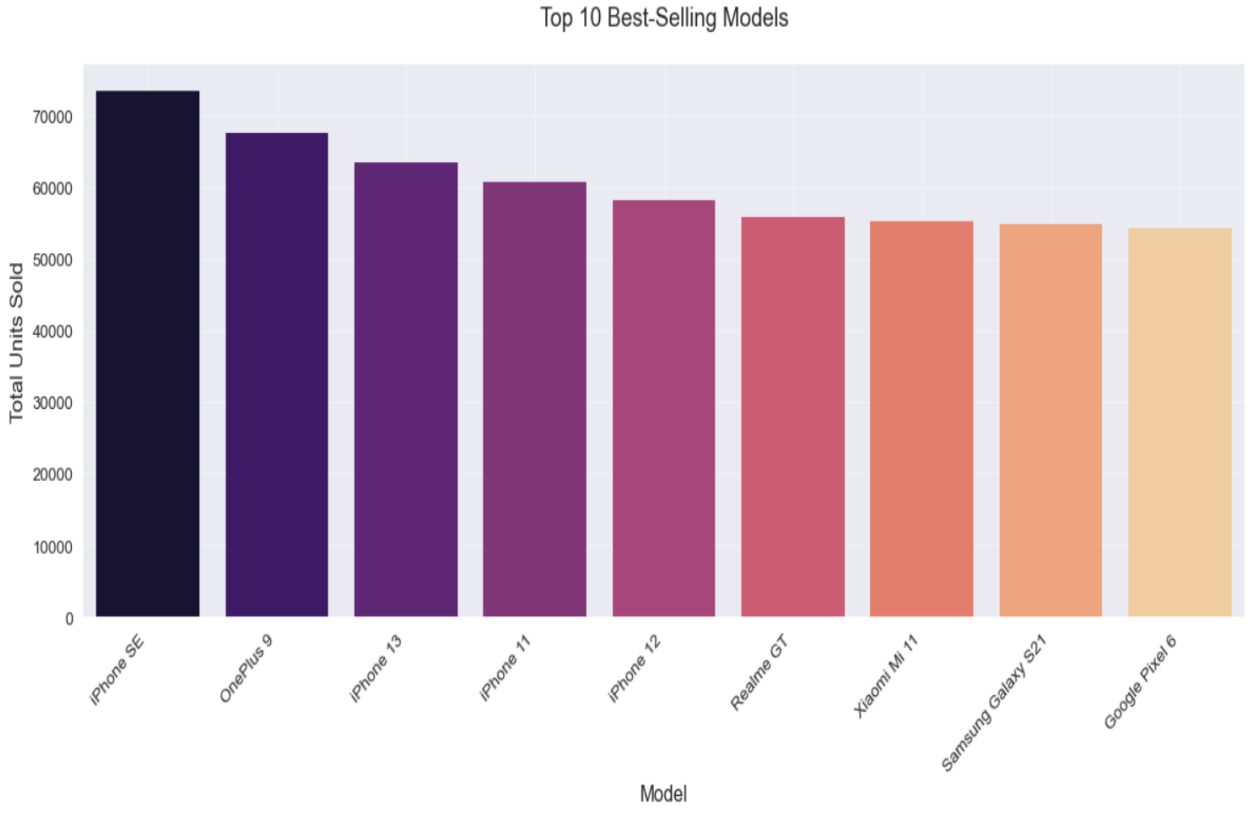

Monthly Sales Trend: iPhone vs Android
iPhone
Android
Units Sold
14000
12000
10000
8000
6000
4000
2000
2021-09
2022-01
2022-05
2022-09
2023-01
2023-05
2023-09
2024-01
2024-05
2024-09
Date

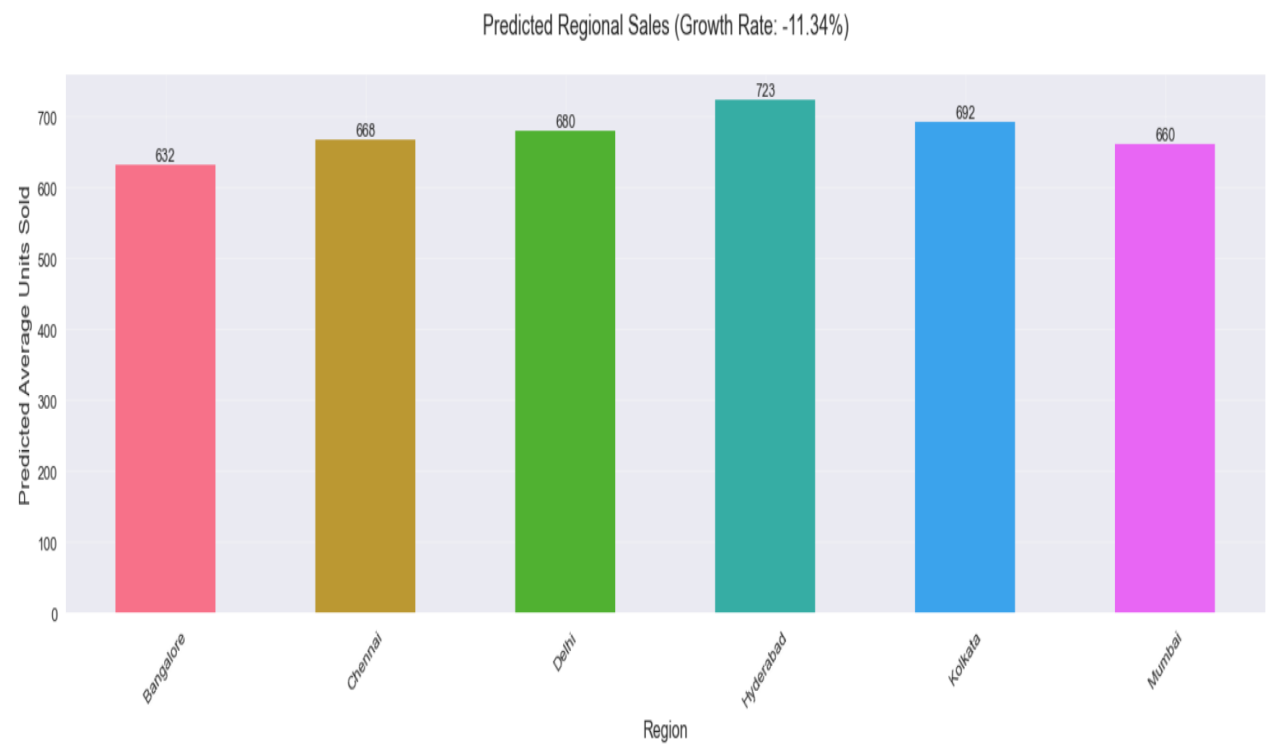

Total Sales by Region
Total Units Sold
100000
80000
60000
40000
20000
0
Hyderabad
Delhi
Mumbai
Chennai
Kolkata
Bangalore
Region

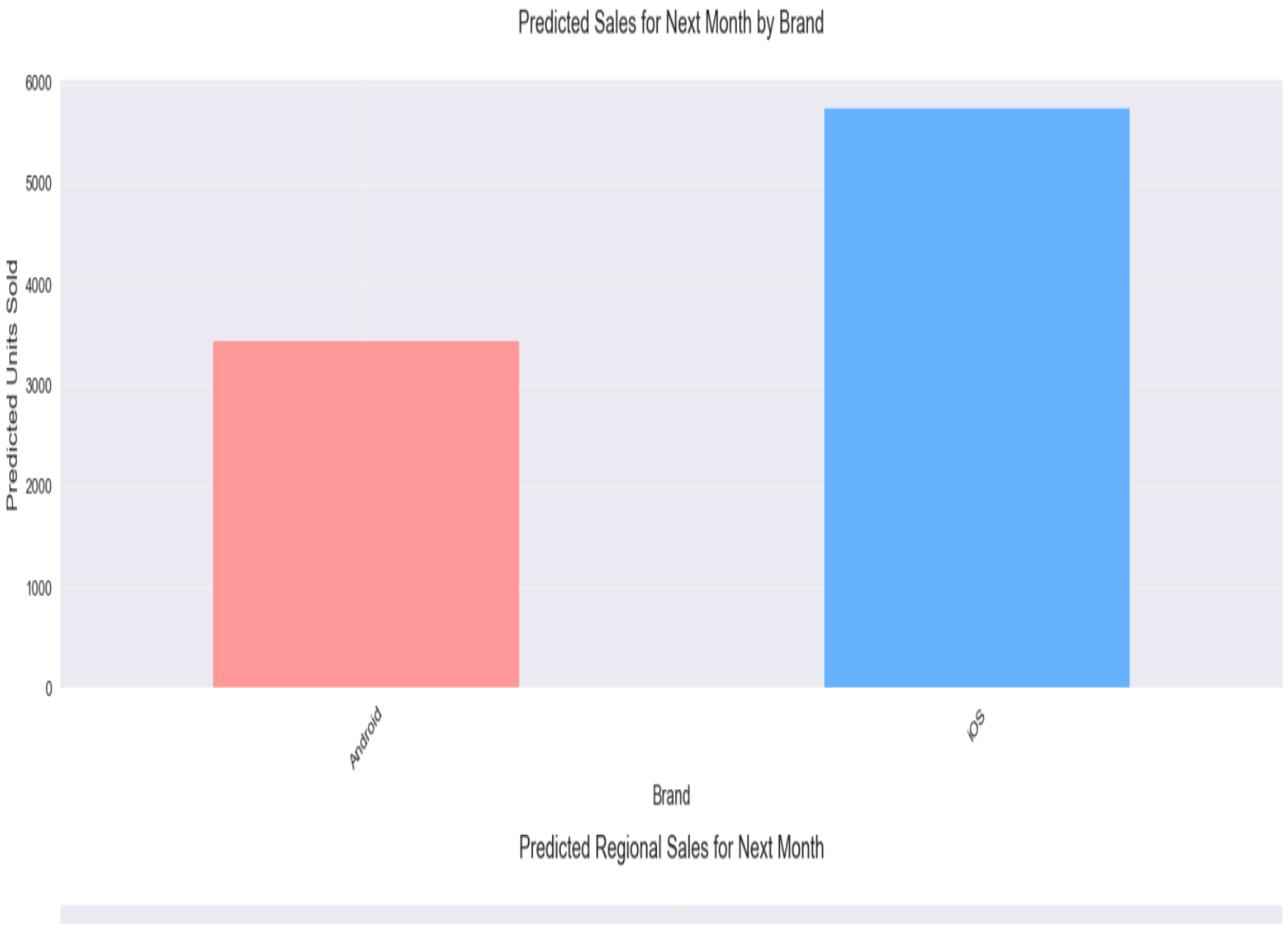

Distribution of Customers by Age Group
Number of Customers
0
20
40
60
80
100
120
140
55+
18-24
45-54
35-44
25-34
Age Group


Correlation Heatmap of Numeric Variables
Price 1 0.051 0.048 0.071 0.0069 -0.054 0.056
Units Sold 0.051 1 1 0.048 -0.027 -0.0029 0.054
Forecasted Sales 0.048 1 1 0.044 -0.031 -0.0027 0.056
Emerging Technology Impact 0.071 0.048 0.044 1 -0.05 0.012 -0.00082
Sustainability Influence 0.0069 -0.027 -0.031 -0.05 1 -0.028 -0.037
Social Media Influence -0.054 -0.0029 -0.0027 0.012 -0.028 1 0.022
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50


Sales Distribution by Channel
Units Sold
0
200
400
600
800
1000
1200
1400
Online
Retail
Sales Channel

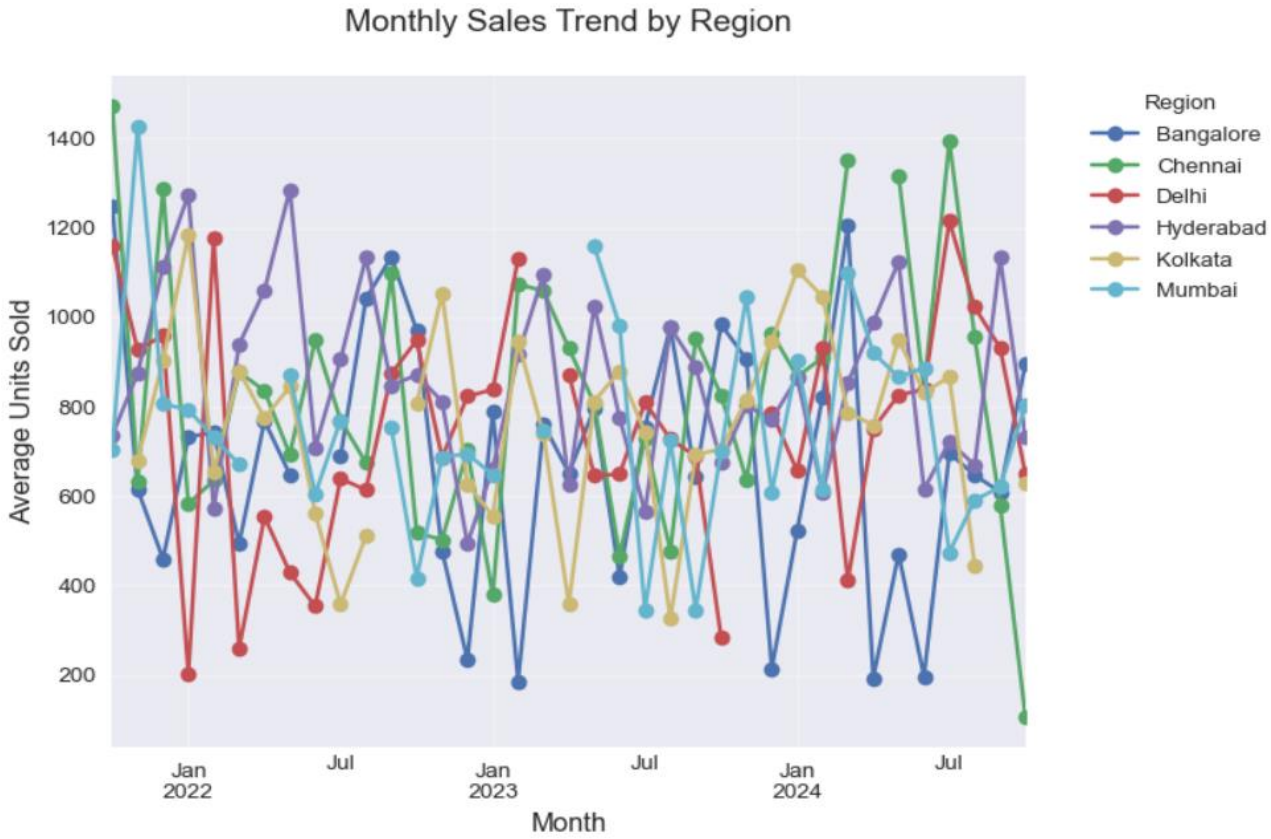

Top 10 Best-Selling Models
Total Units Sold
0
10000
20000
30000
40000
50000
60000
70000
iPhone SE
OnePlus 9
iPhone 13
iPhone 11
iPhone 12
Realme GT
Xiaomi Mi 11
Samsung Galaxy S21
Google Pixel 6
Model

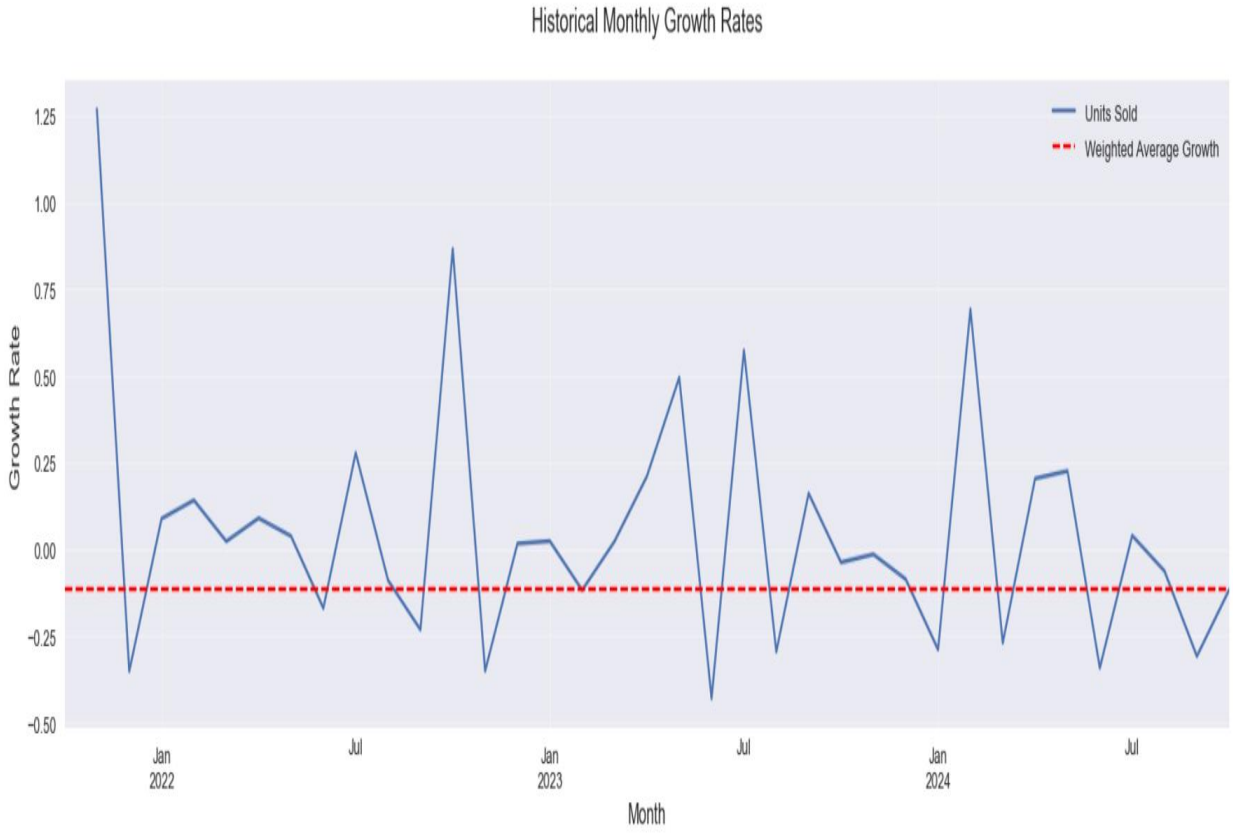




Predicted Regional Sales for Next Month


Monthly Sales Trend by Brand
Brand
Android
iOS
Units Sold
14000
12000
10000
8000
6000
4000
Jan
2022
Jul
Jan
2023
Jul
Jan
2024
Jul
Month


Monthly Sales Trend by Region
Region
Bangalore
Chennai
Delhi
Hyderabad
Kolkata
Mumbai
Average Units Sold
1400
1200
1000
800
600
400
200
Jan
2022
Jul
Jan
2023
Jul
Jan
2024
Jul
Month




```
Growth Rate Confidence Interval Analysis:
95% Confidence Interval: -82.47% to 59.79%


Predictions Summary:
           Current_Sales  Predicted_Sales  Lower_Bound  Upper_Bound
Region
Bangalore     713.271028       632.391697   125.025475  1139.757920
Chennai       752.895652       667.523200   131.971064  1203.075337
Delhi         767.053435       680.075602   134.452706  1225.698498
Hyderabad     815.446154       722.980967   142.935208  1303.026725
Kolkata       780.353982       691.867972   136.784088  1246.951855
Mumbai        744.672414       660.232413   130.529656  1189.935169
```

8. Conclusion

The analysis concludes that Android dominates the Indian smartphone market due to its wide availability across various price ranges, appealing to cost-sensitive consumers, while iPhone holds a strong position in the premium segment due to brand loyalty and perceived quality. Factors like price, brand perception, and regional economic disparities significantly influence consumer choices, with Android leading in overall market share but iPhone maintaining dominance among high-income users. Future trends indicate Android's continued growth in volume, while iPhone retains a loyal customer base in the premium category.