SAMSUNG

Superstore Sales Analysis

Team: 10

Our Team

Yousef Emad

Somaia Ahmed

Nour-Eldeen Anwar

Haneen El-Daly

Overview

- Introduction
- Dataset Overview
- Business Goals
- EDA
- Data Preprocessing
- Modeling
- Deployment

Introduction

Data analysis has become a critical tool for businesses seeking to drive profitability, optimize sales, and streamline operations.
Retailers generate vast amounts of data through every customer interaction, transaction, and product movement.

Harnessing this data allows businesses to gain deeper insights into customer behavior, product performance, and operational efficiencies.

Dataset Overview:

- The "Superstore Sales" dataset captures transactional data from a retail store, covering various regions, product categories, and customer segments.
- It includes information such as sales, profit, discounts, shipping modes, and customer demographics.

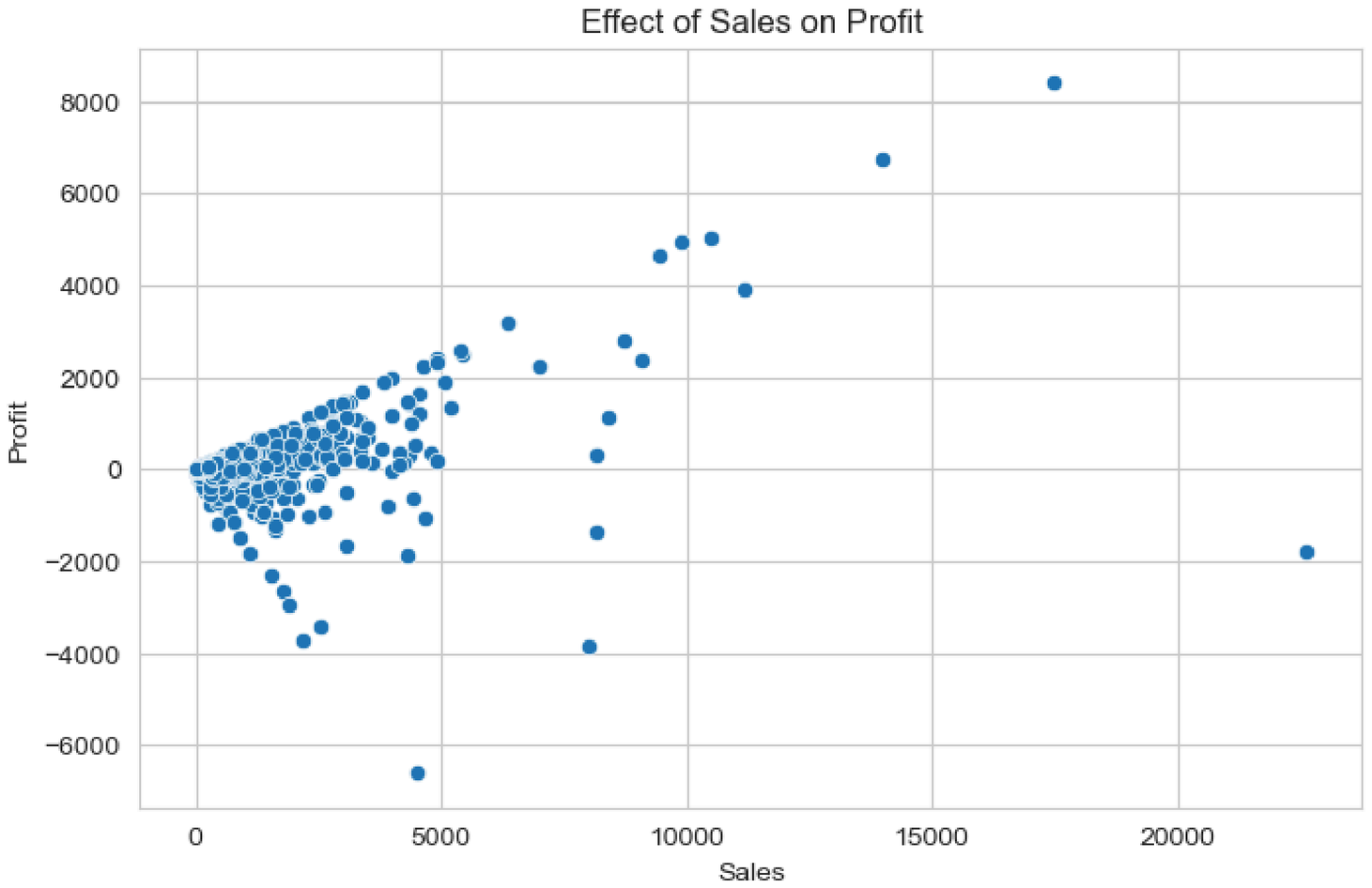
Business Goals:

- Understand how discounts impact profitability.
- Evaluate the impact of seasonality on sales trends.
- Identify the most profitable product combinations.

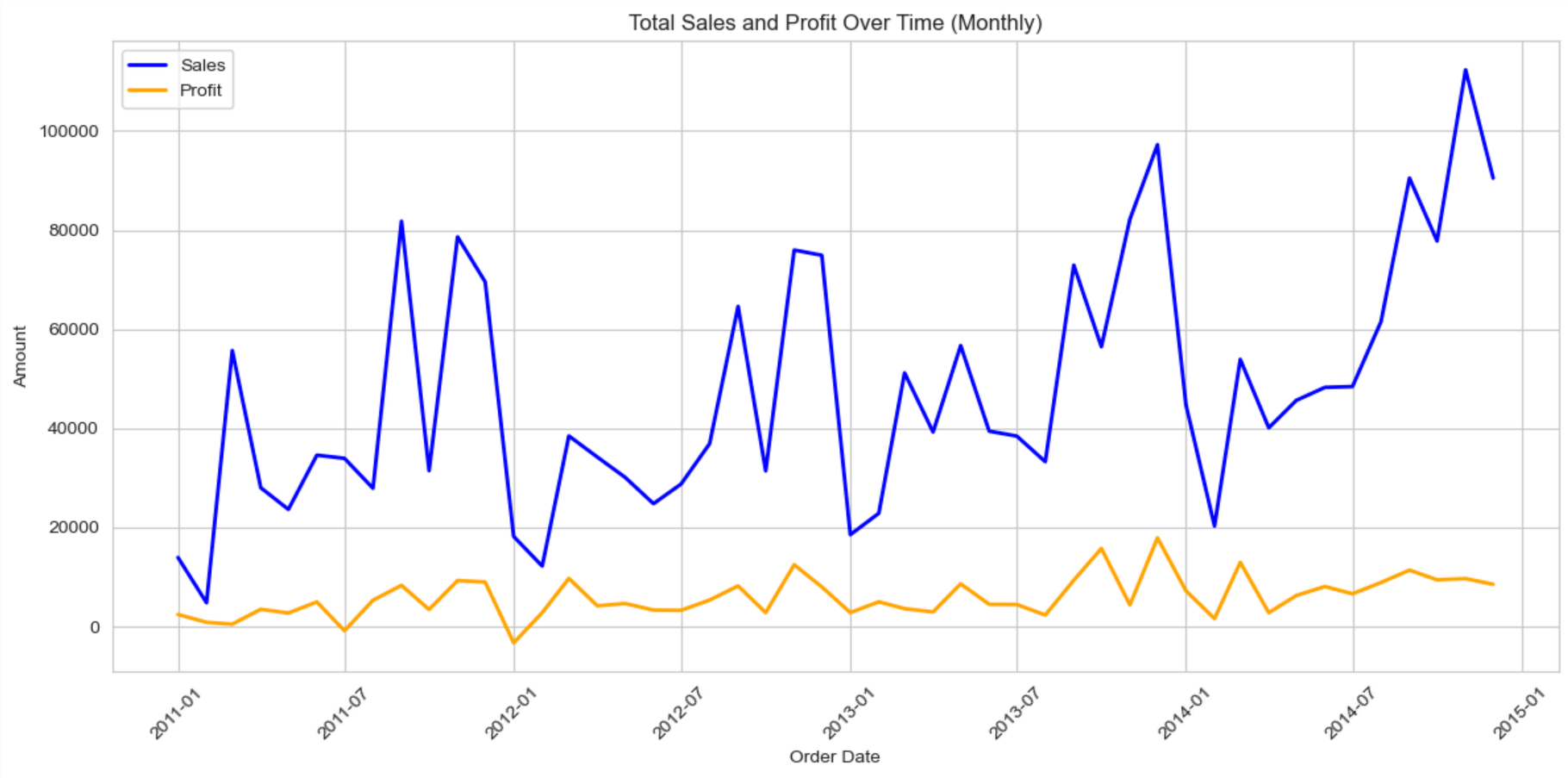
SAMSUNG

SALES PERFORMANCE

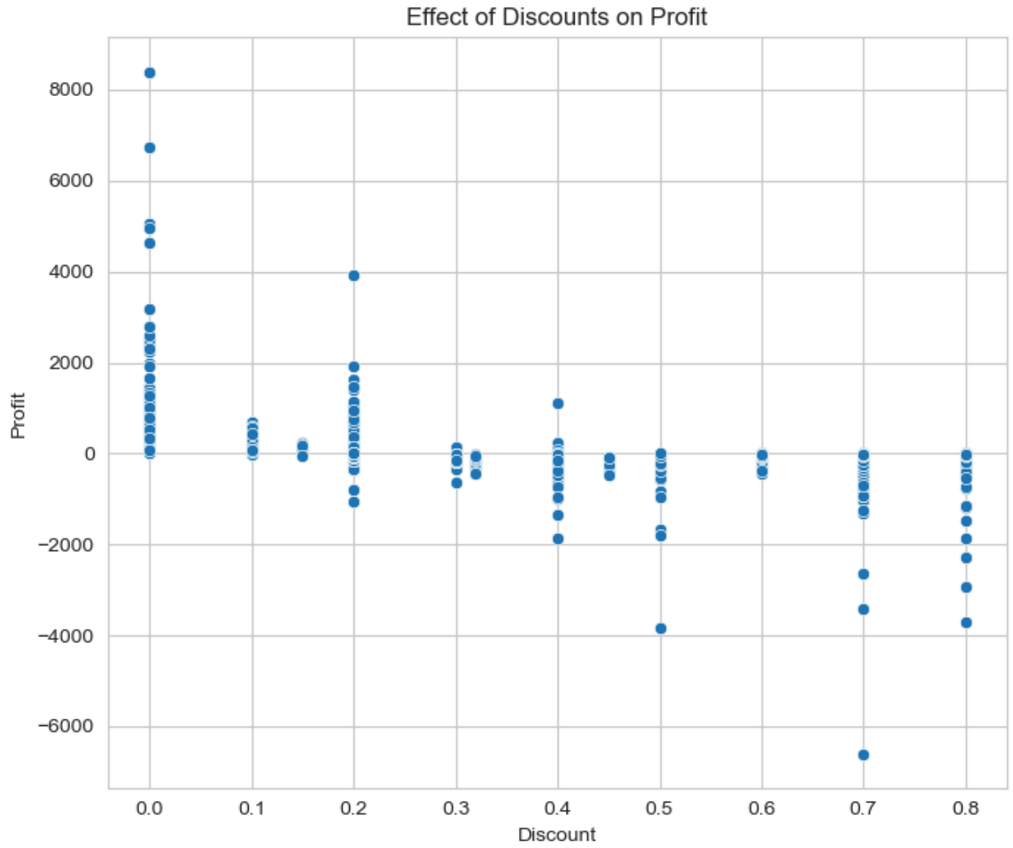
Effect of Sales on Profit

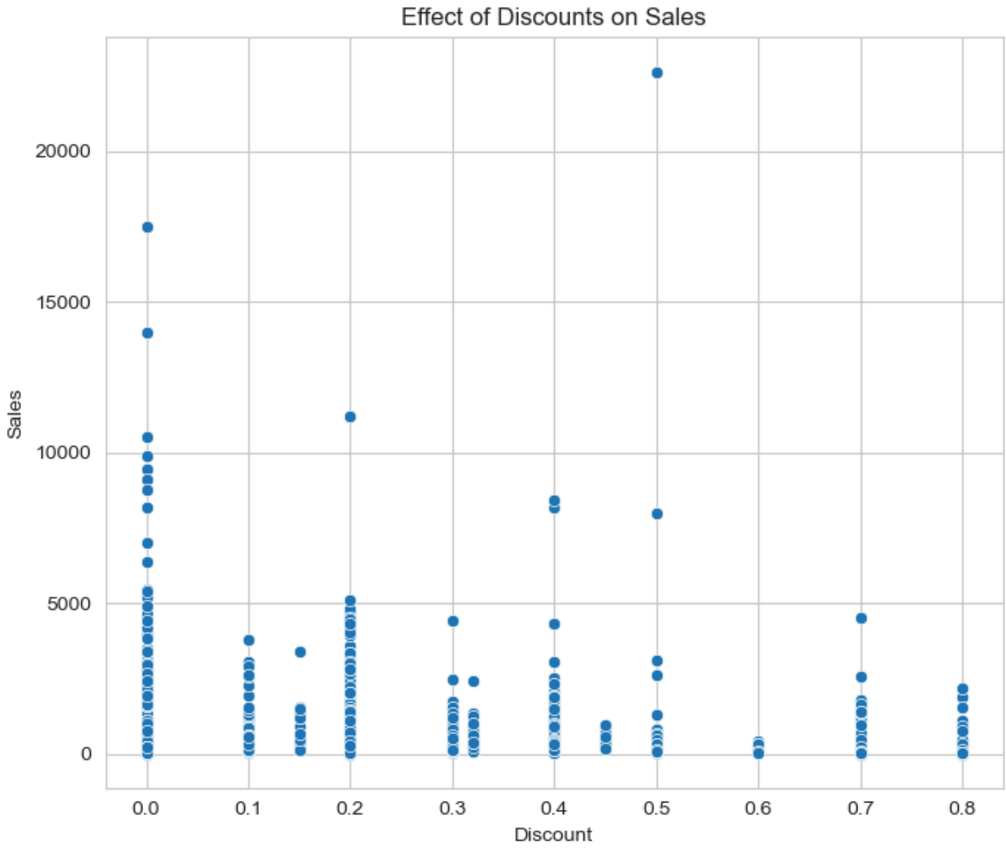


Total Sales and Profit Over Time



Effect of Discounts on Profit and Sales

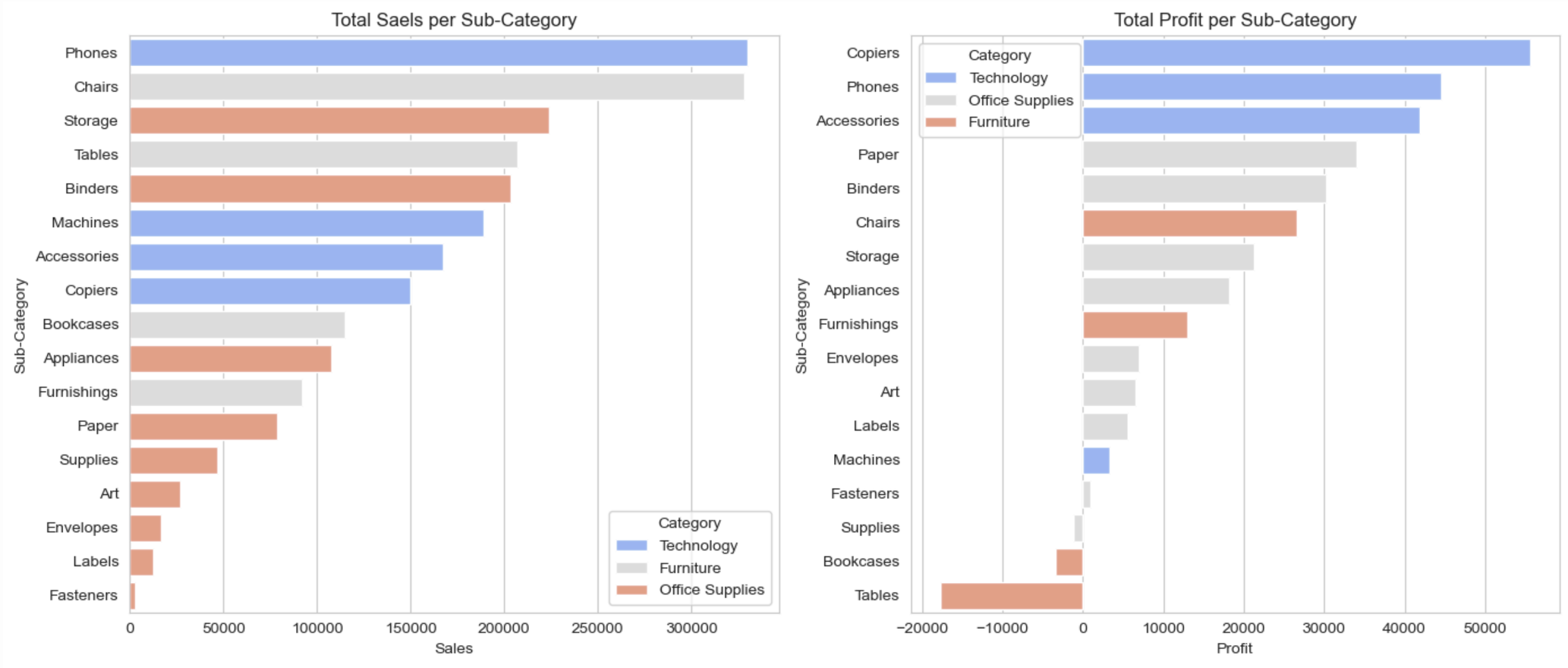


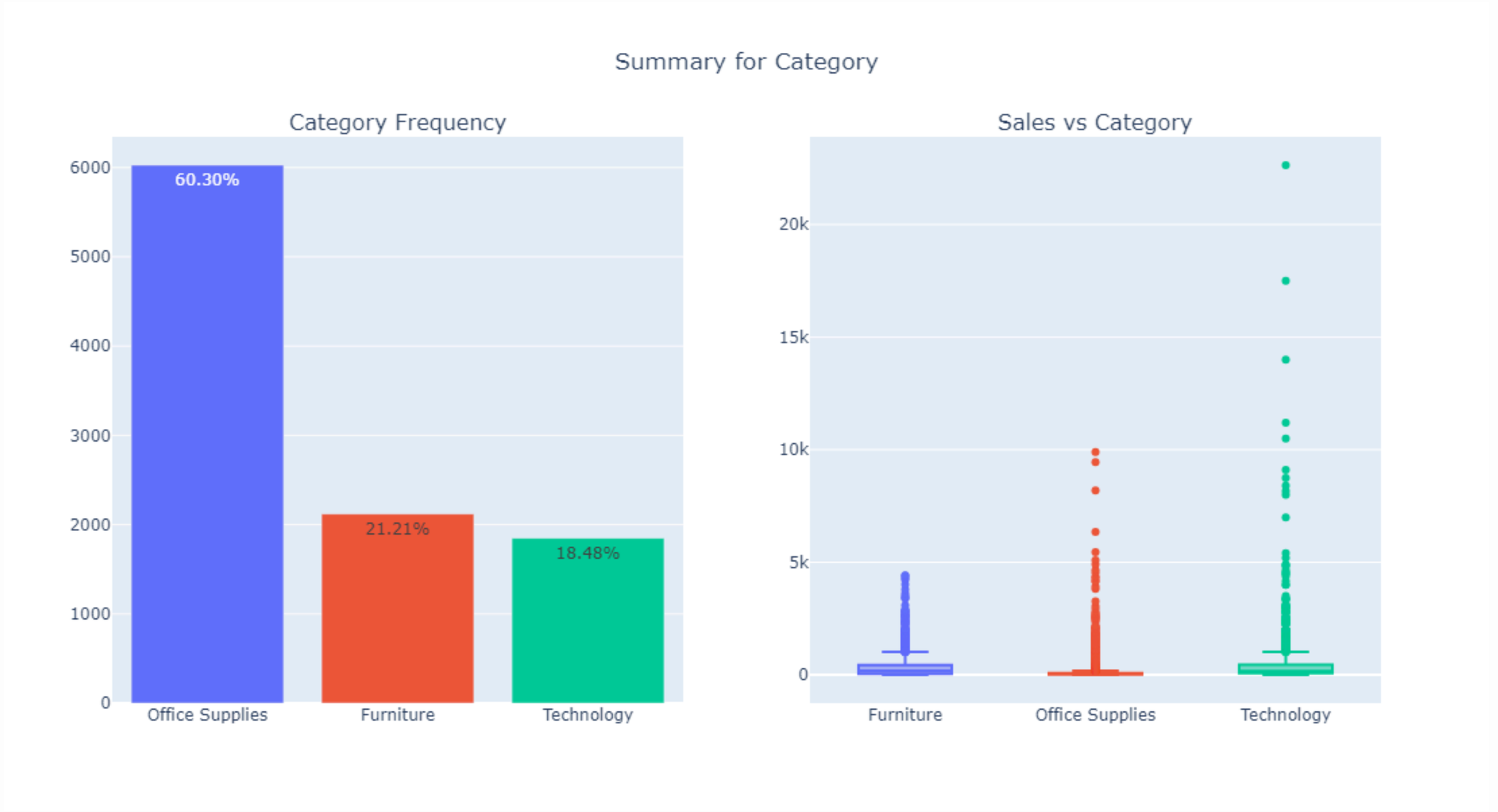


SAMSUNG

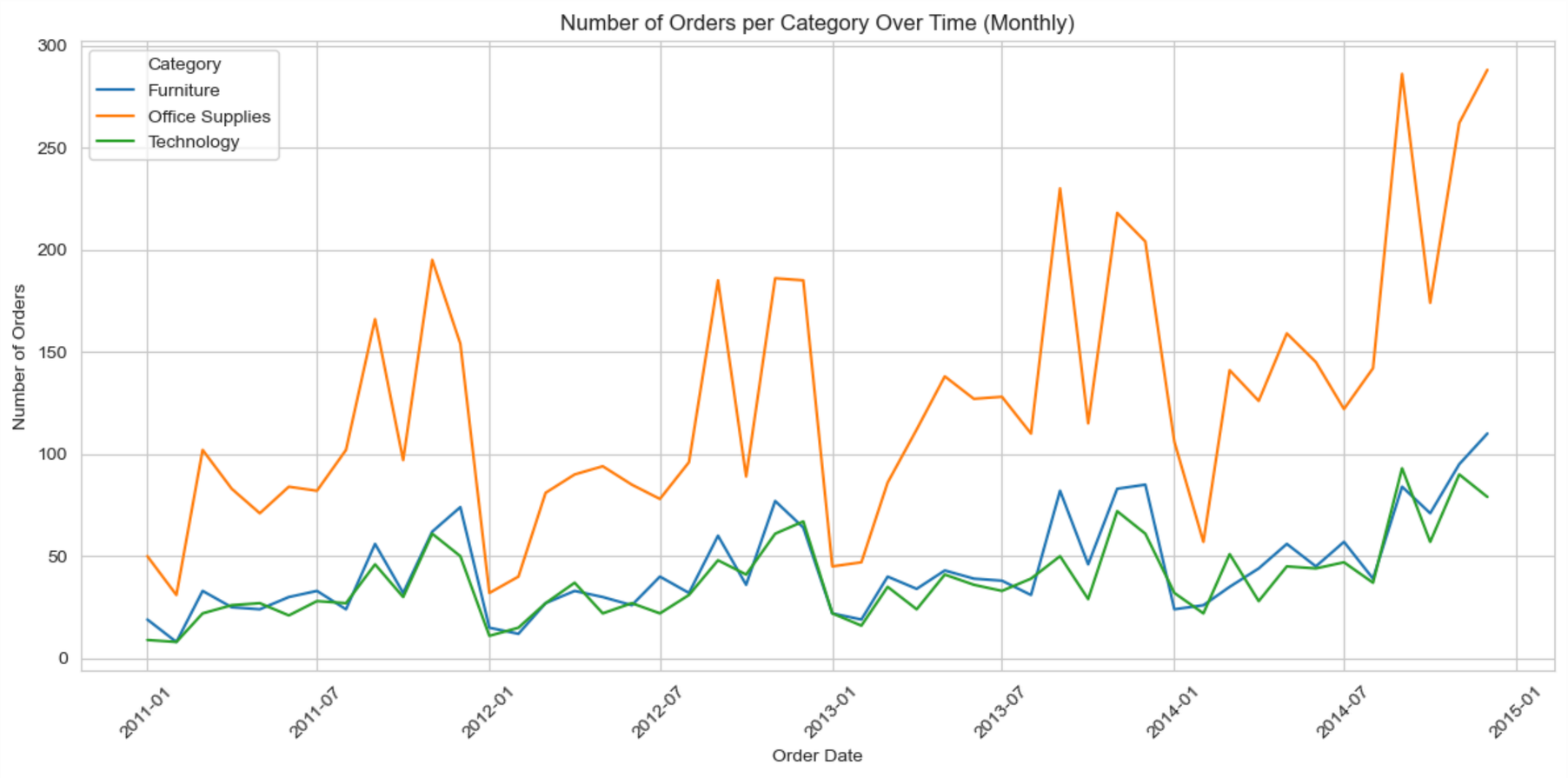
PRODUCT CATEGORIES

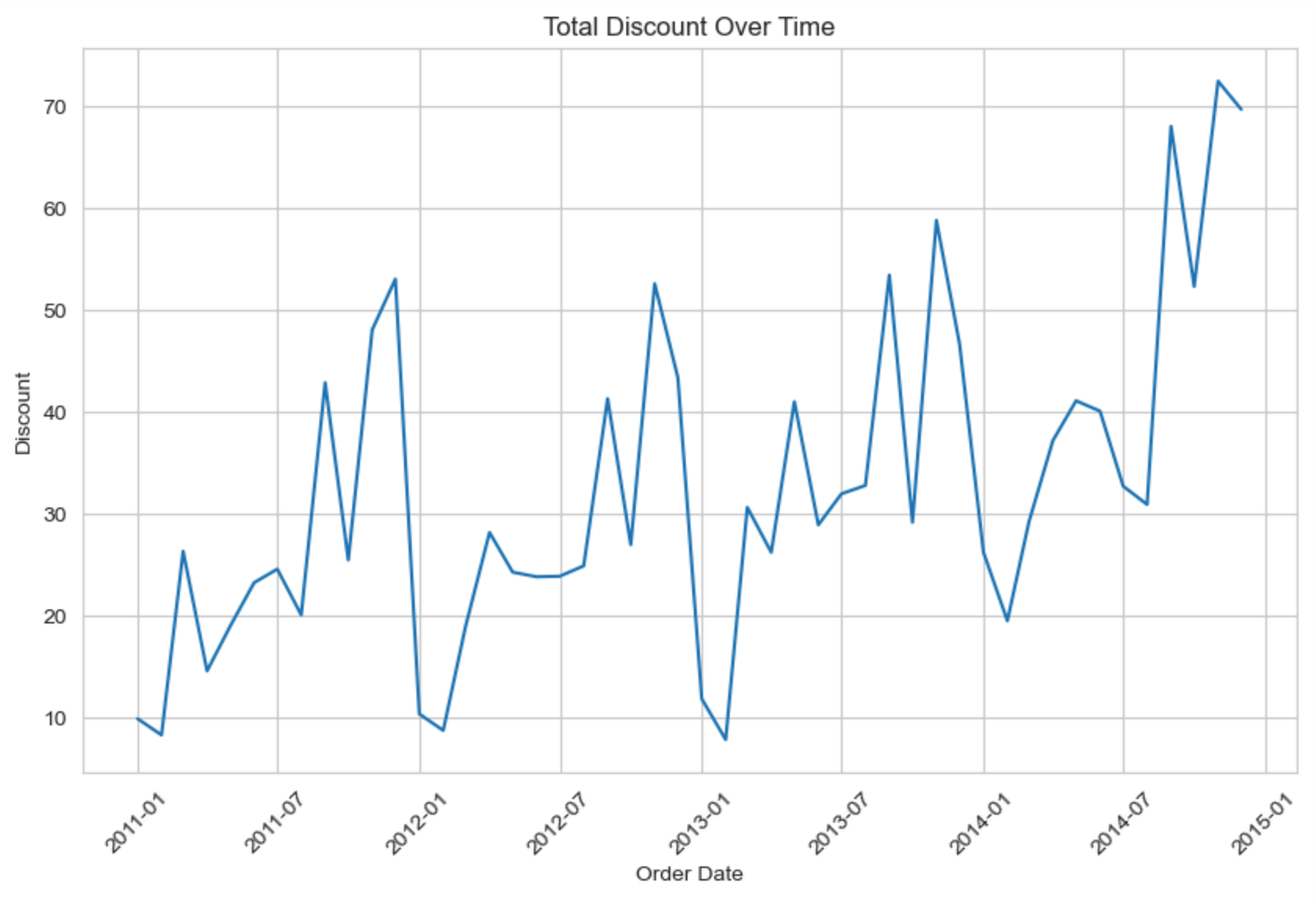
Product Categories




Summary for Category
Category Frequency
60.30%
21.21%
18.48%
6000
5000
4000
3000
2000
1000
0
Office Supplies
Furniture
Technology
Sales vs Category
20k
15k
10k
5k
0
Furniture
Office Supplies
Technology

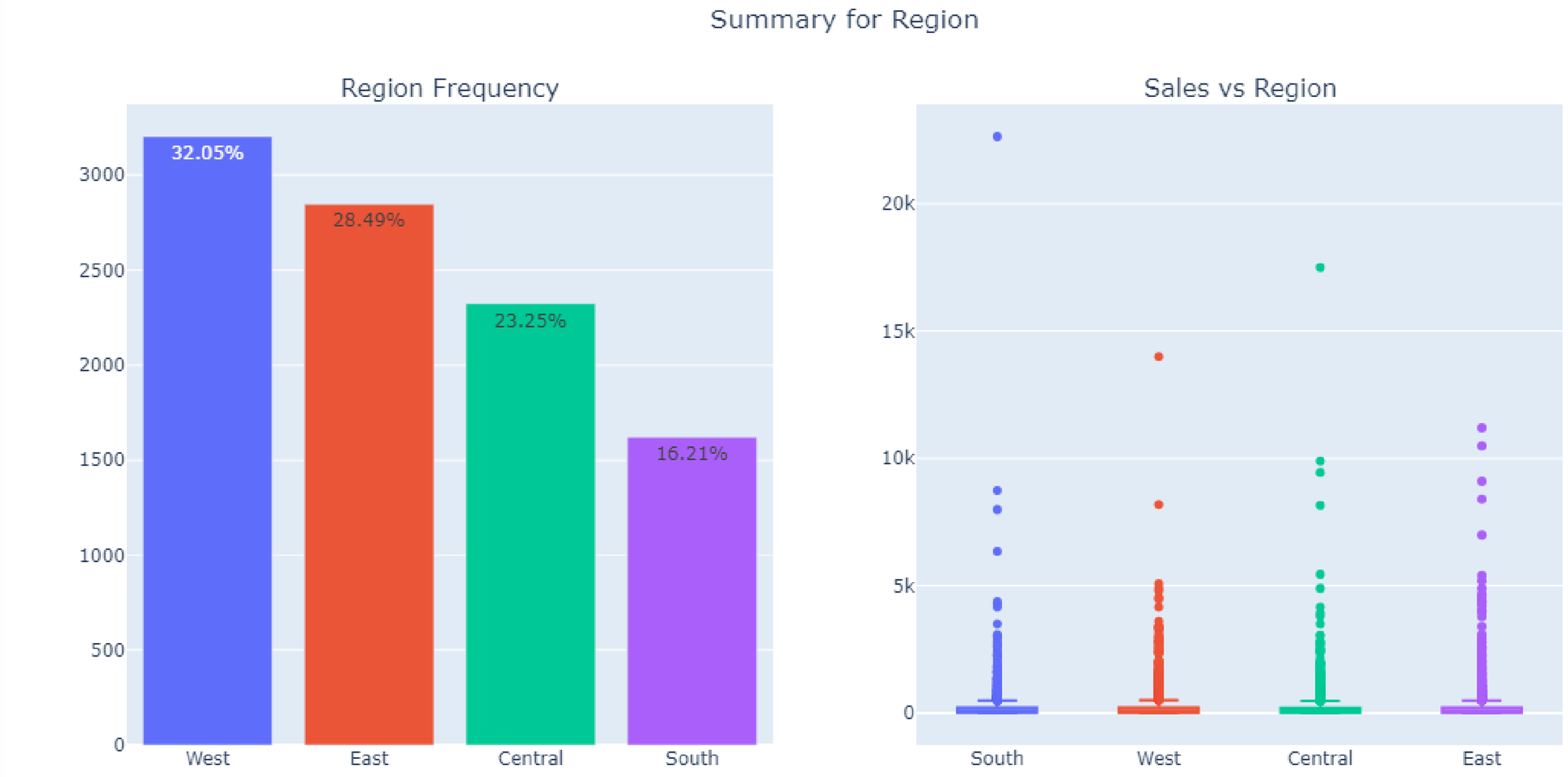
Number of Orders per Category Over Time

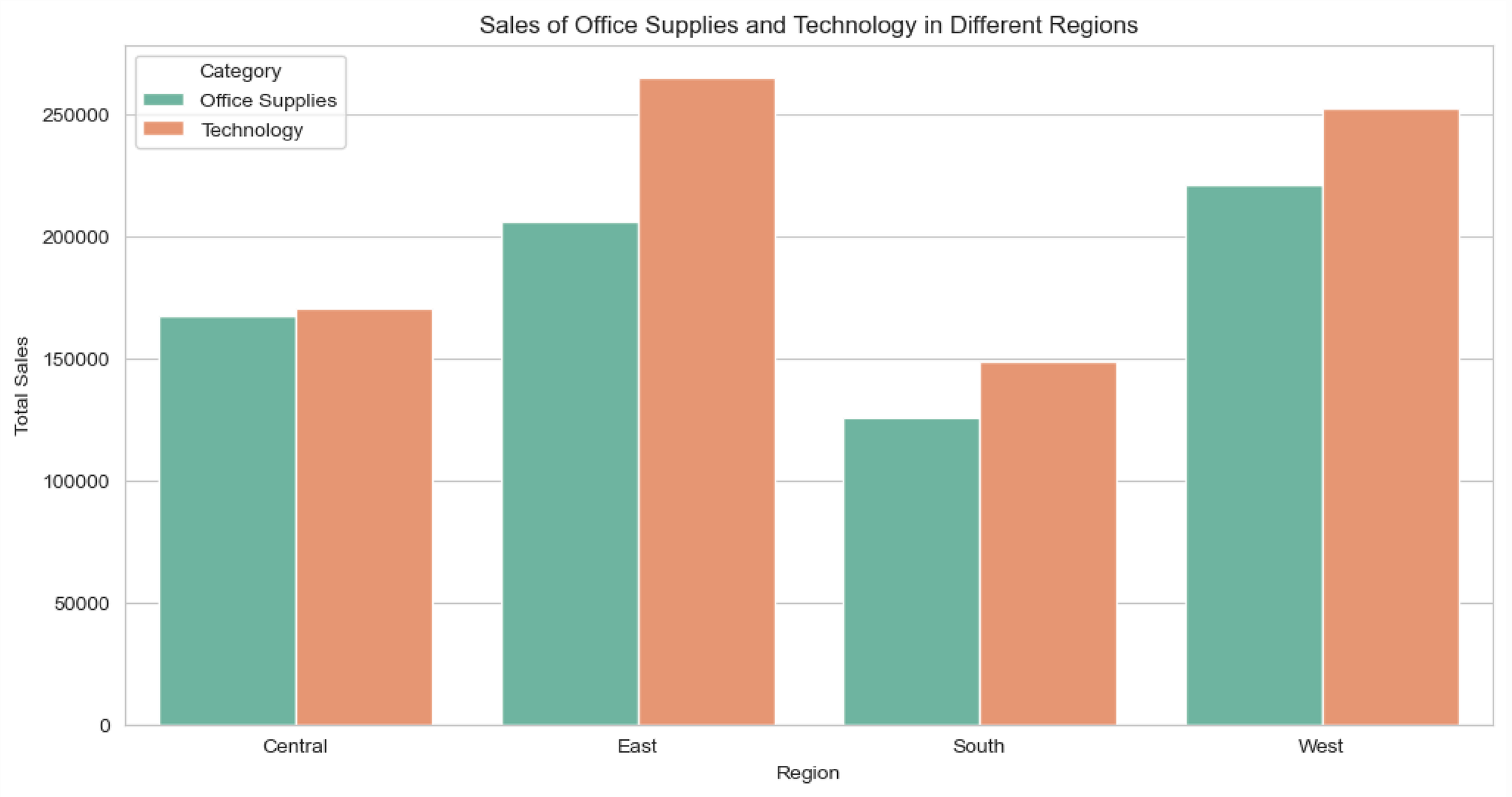



Total Discount Over Time
Discount
70
60
50
40
30
20
10
2011-01
2011-07
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01
Order Date

SAMSUNG

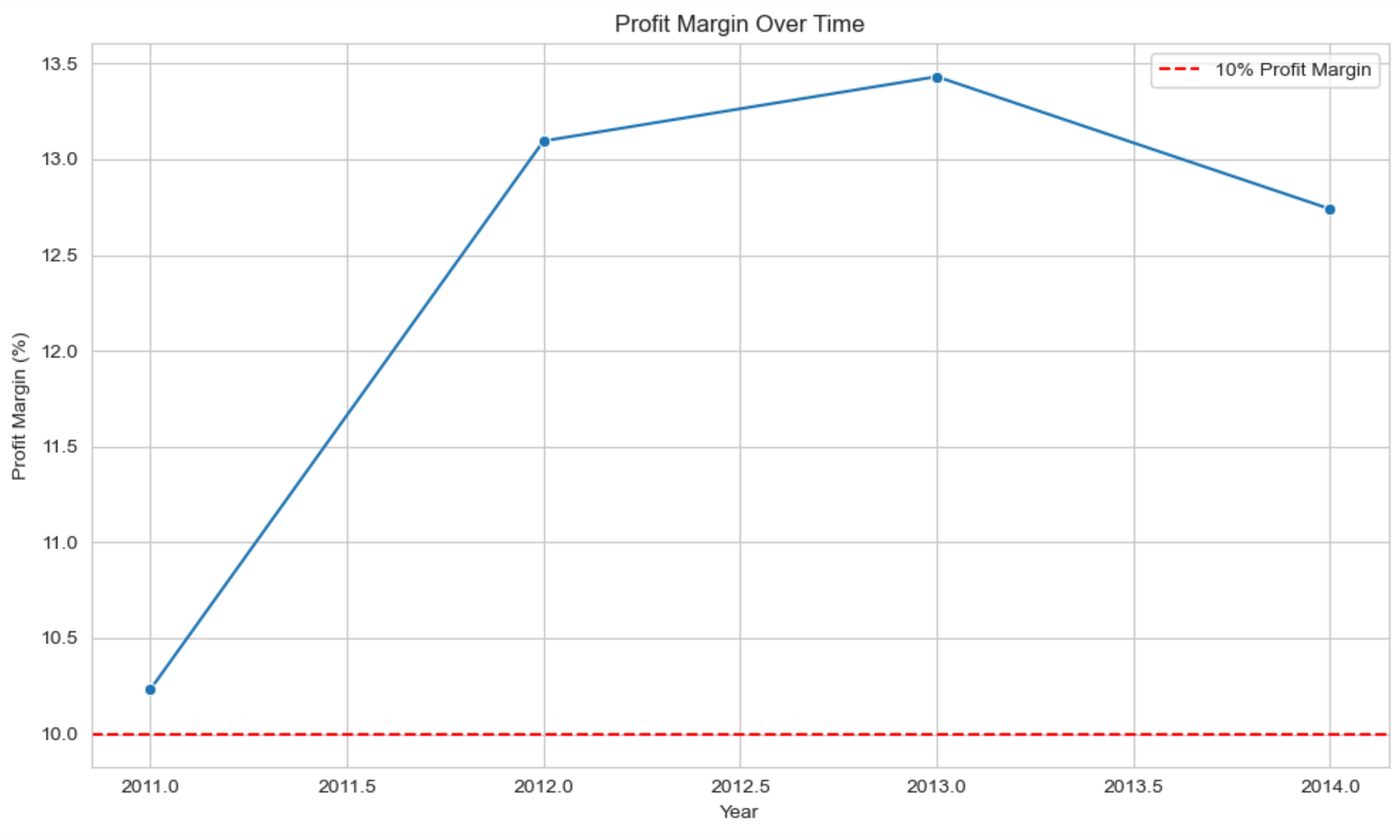
GEOGRAPHICAL INSIGHTS

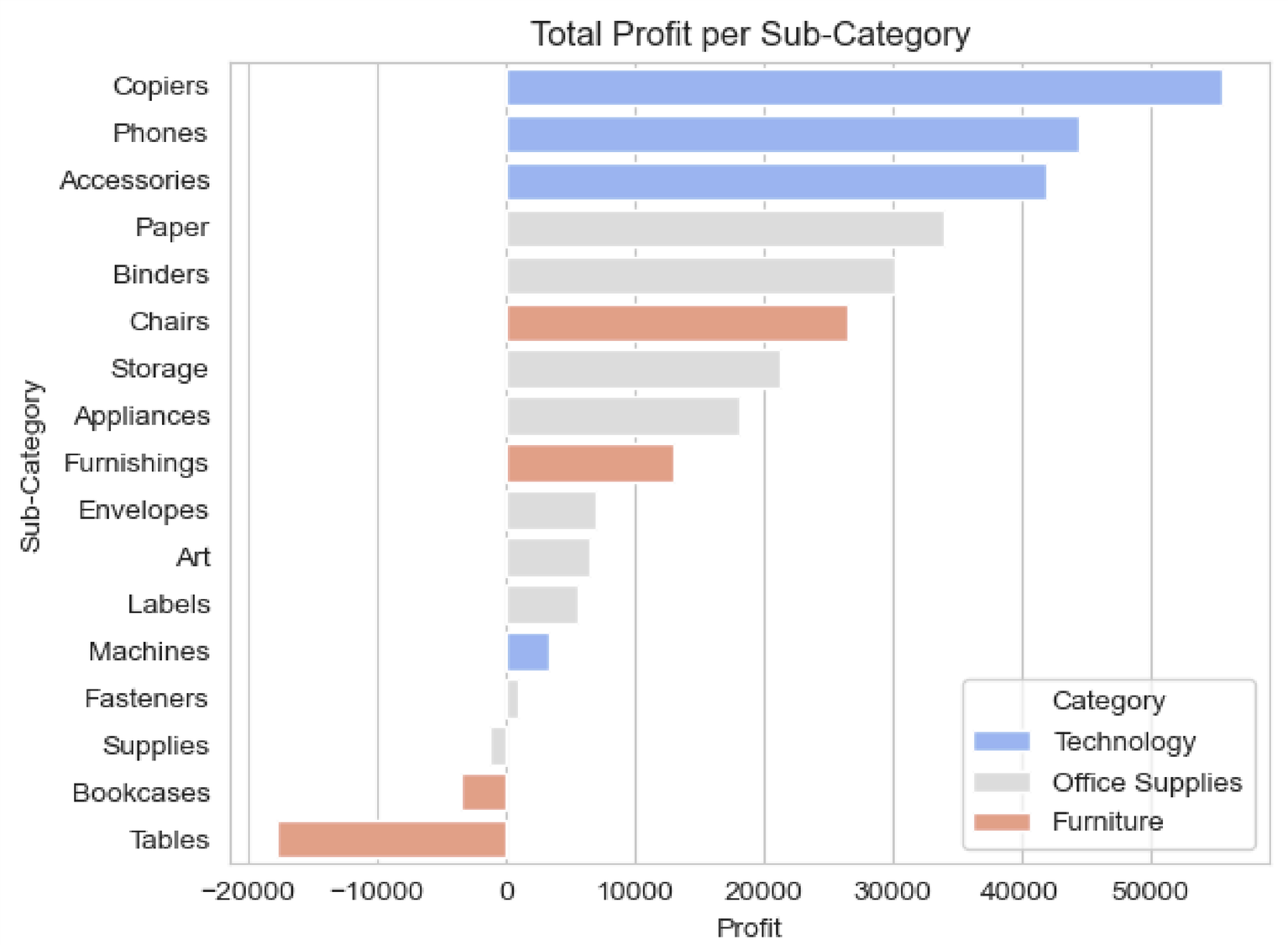

Summary for Region
Region Frequency
32.05%
28.49%
23.25%
16.21%
3000
2500
2000
1500
1000
500
0
West
East
Central
South
Sales vs Region
20k
15k
10k
5k
0
South
West
Central
East

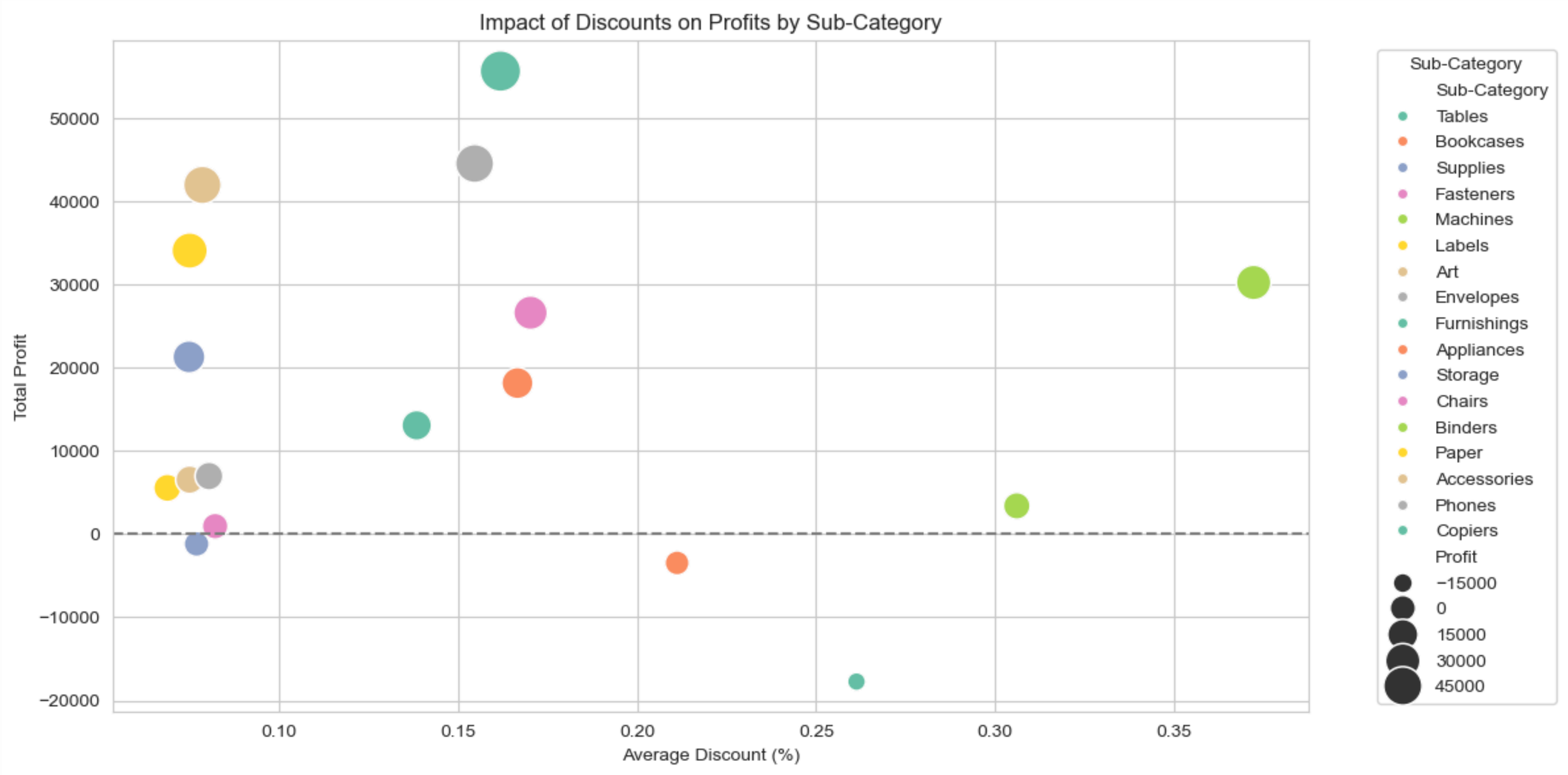

Sales of Office Supplies and Technology in Different Regions
Category
Office Supplies
Technology
250000
200000
150000
100000
50000
0
Total Sales
Central
East
South
West
Region

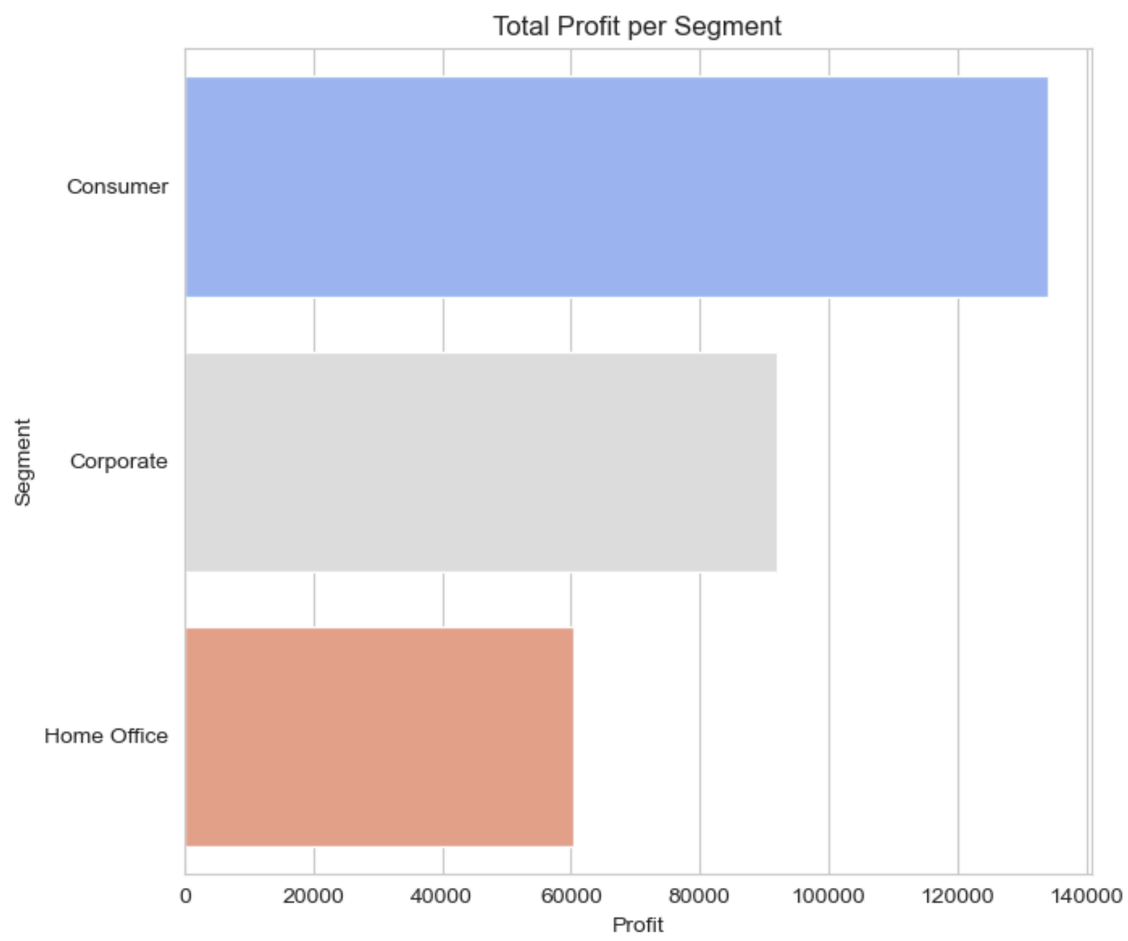
SAMSUNG

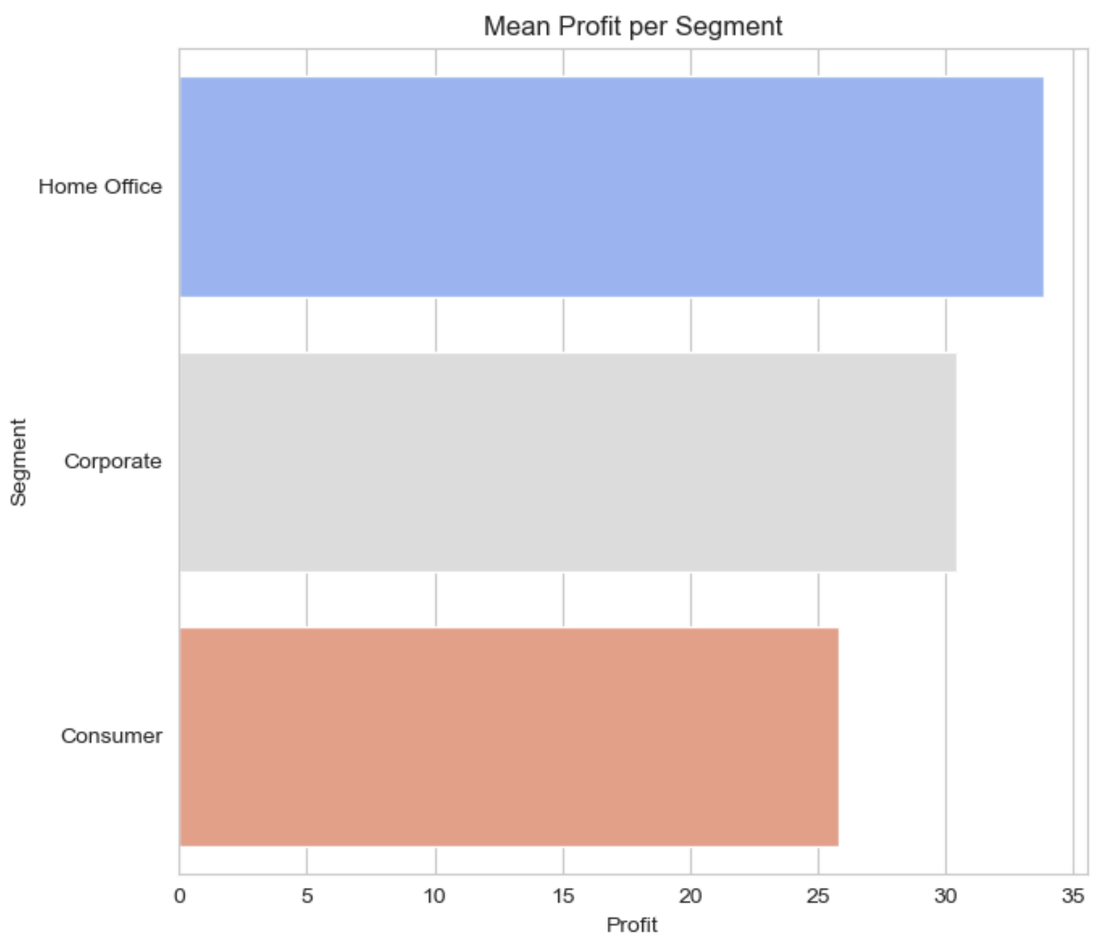
PROFITABILITY


Profit Margin Over Time
10% Profit Margin
Profit Margin (%)
13.5
13.0
12.5
12.0
11.5
11.0
10.5
10.0
2011.0
2011.5
2012.0
2012.5
2013.0
2013.5
2014.0
Year


Total Profit per Sub-Category
Copiers
Phones
Accessories
Paper
Binders
Chairs
Storage
Appliances
Furnishings
Envelopes
Art
Labels
Machines
Fasteners
Supplies
Bookcases
Tables
Sub-Category
−20000
−10000
0
10000
20000
30000
40000
50000
Profit
Category
Technology
Office Supplies
Furniture


Impact of Discounts on Profits by Sub-Category
Total Profit
50000
40000
30000
20000
10000
0
−10000
−20000
0.10
0.15
0.20
0.25
0.30
0.35
Average Discount (%)
Sub-Category
Sub-Category
Tables
Bookcases
Supplies
Fasteners
Machines
Labels
Art
Envelopes
Furnishings
Appliances
Storage
Chairs
Binders
Paper
Accessories
Phones
Copiers
Profit
−15000
0
15000
30000
45000


Total Profit per Segment
Consumer
Corporate
Home Office
Segment
0
20000
40000
60000
80000
100000
120000
140000
Profit


Mean Profit per Segment
Home Office
Corporate
Consumer
Segment
0
5
10
15
20
25
30
35
Profit

SAMSUNG

BUSINESS INSGHITS

Problem

- High Discount Dependency Leading to Low Profit Margins

Solution

- Redefine Discount Strategies

Action: Implement data-driven discount strategies that vary by product category and customer segment.

Problem

- Inconsistent Sales Across Product Categories

Solution

- Boost Sales in Underperforming Product Categories

Action: Launch targeted marketing campaigns promoting underperforming categories like Furniture and Technology.

Problem

- Uneven Regional Sales Distribution

Solution

- Region-Specific Sales Strategies:

Action: Use sales data to identify regional demand patterns and preferences.

Problem

- Limited Insight into Customer Behavior

Solution

- Develop Customer Loyalty and Retention Programs:

Action: Use purchase history to personalize offers and encourage long-term customer retention.

SAMSUNG

Data Preprocessing

Encoding

Label encoder

Feature Engineering

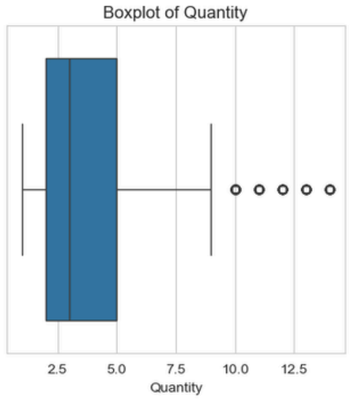
01 Profit Margin

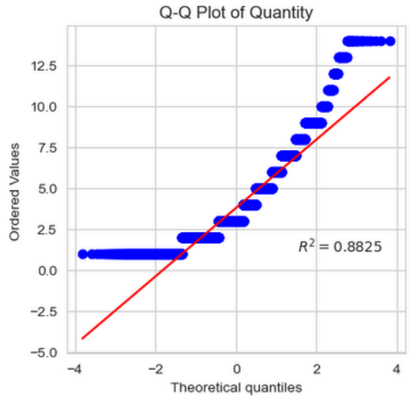
02 Discounted Profit

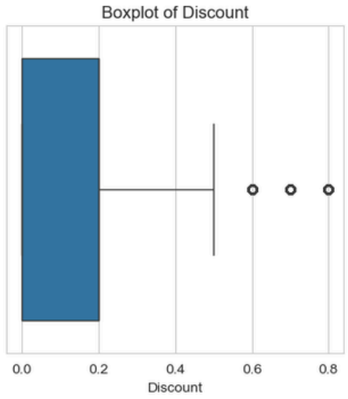
03 Discount Percentage

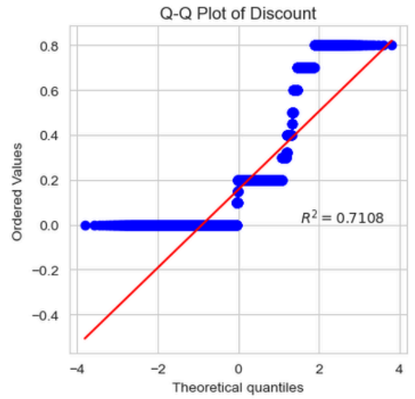
SAMSUNG

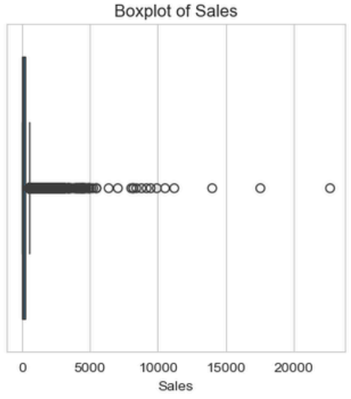
SKEWENESS HANDLING

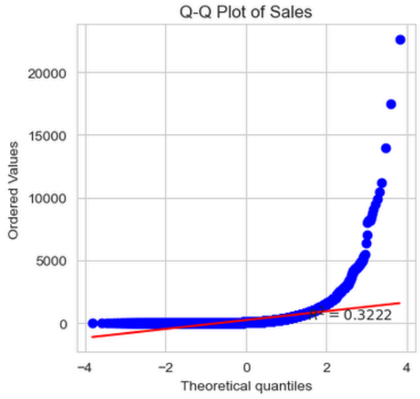

Boxplot of Quantity
Quantity

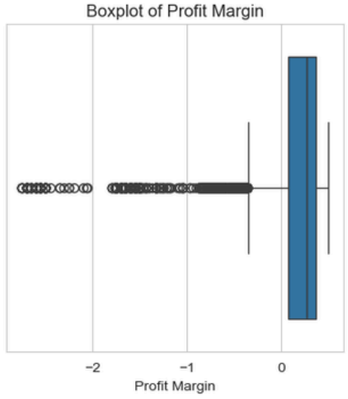

Q-Q Plot of Quantity
Ordered Values
Theoretical quantiles
$R^2 = 0.8825$

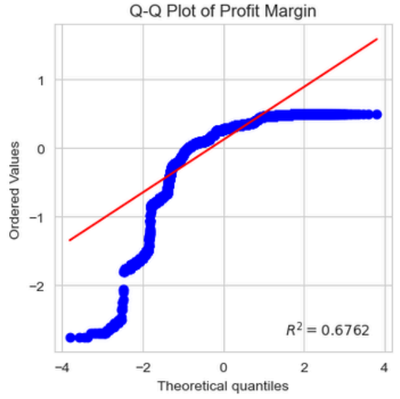

Boxplot of Discount
Discount

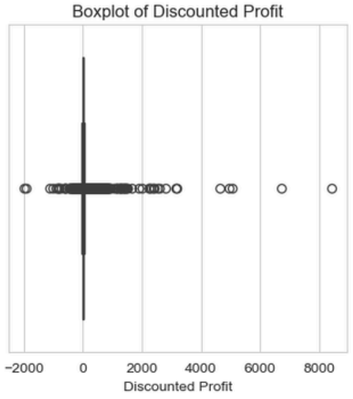

Q-Q Plot of Discount
Ordered Values
Theoretical quantiles
$R^2 = 0.7108$

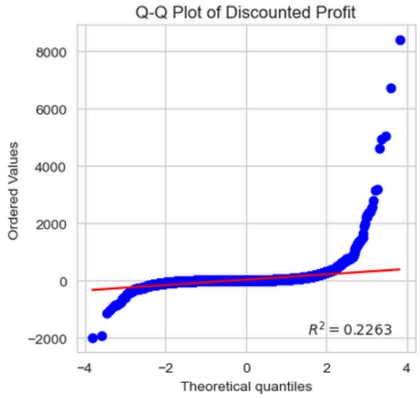

Boxplot of Sales
Sales

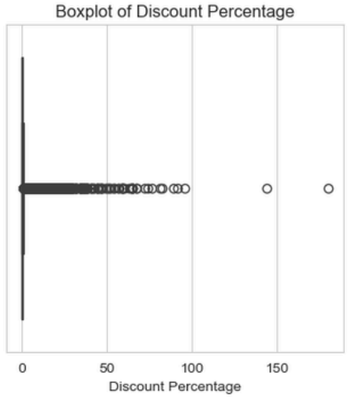

Q-Q Plot of Sales
Ordered Values
Theoretical quantiles
$R^2 = 0.3222$

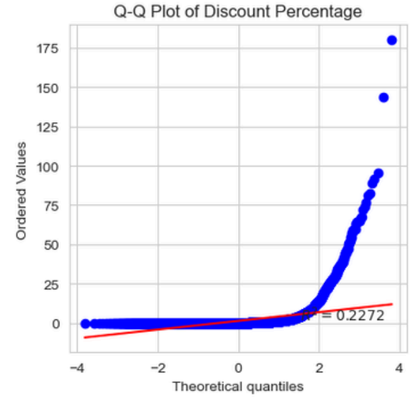

Boxplot of Profit Margin
Profit Margin


Q-Q Plot of Profit Margin
Ordered Values
Theoretical quantiles
$R^2 = 0.6762$


Boxplot of Discounted Profit
Discounted Profit


Q-Q Plot of Discounted Profit
Ordered Values
Theoretical quantiles
$R^2 = 0.2263$


Boxplot of Discount Percentage
Discount Percentage


Q-Q Plot of Discount Percentage
Ordered Values
Theoretical quantiles
$R^2 = 0.2272$

Feature Transformation

01

log_transform

Quantity
Discounted Profit

02

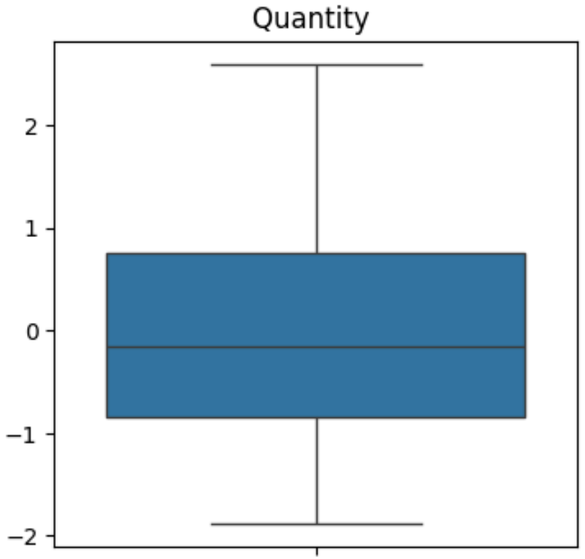
power_transform

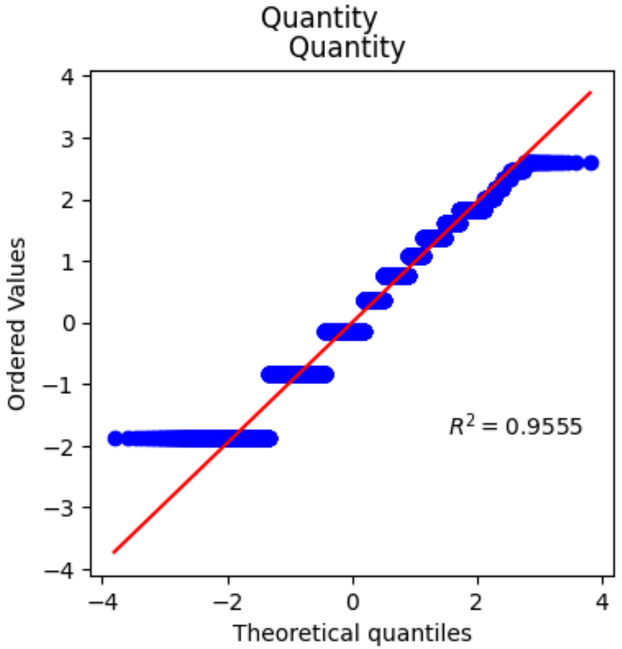
Profit Margin
Sales
Discount Percentage

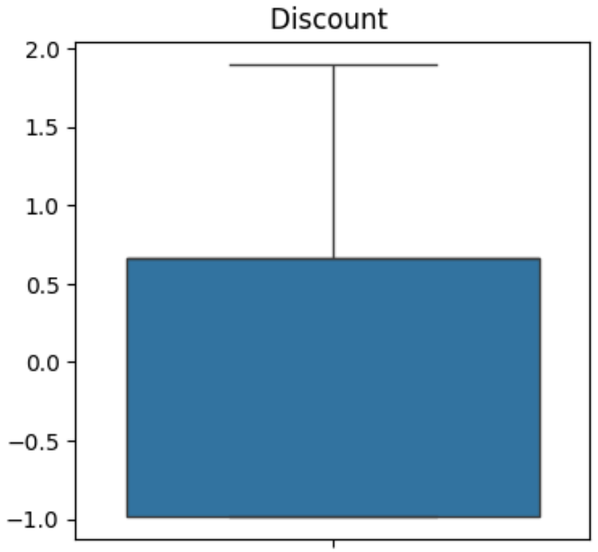
03

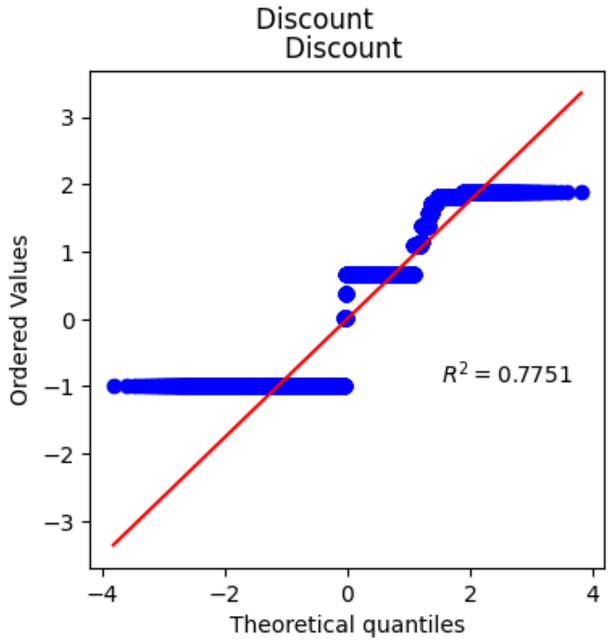
sqrt_transform

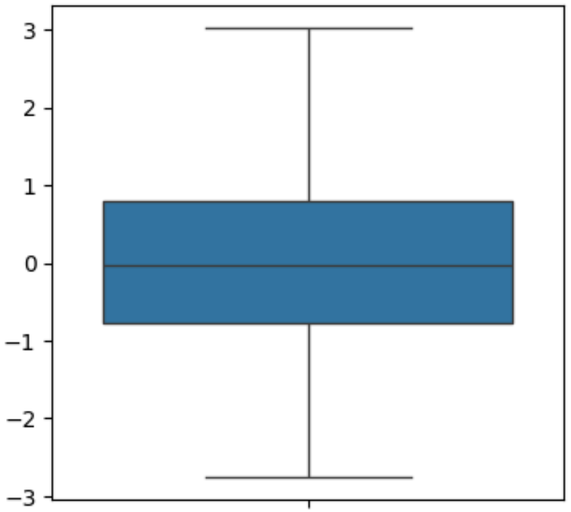
Discount

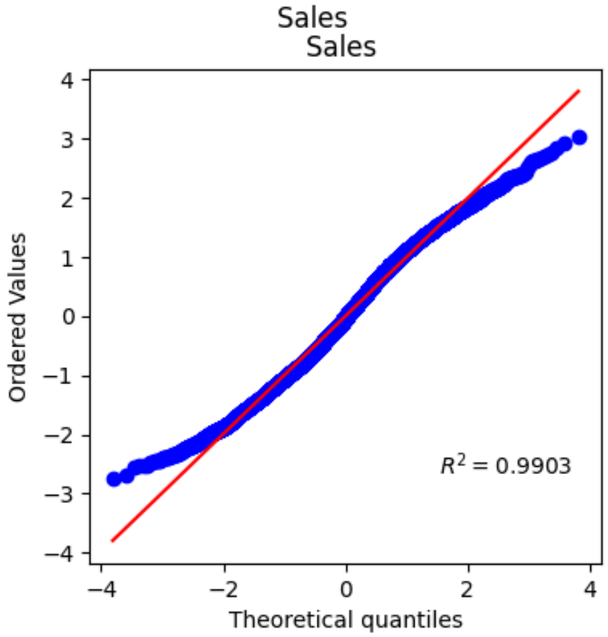

Quantity
2
1
0
−1
−2

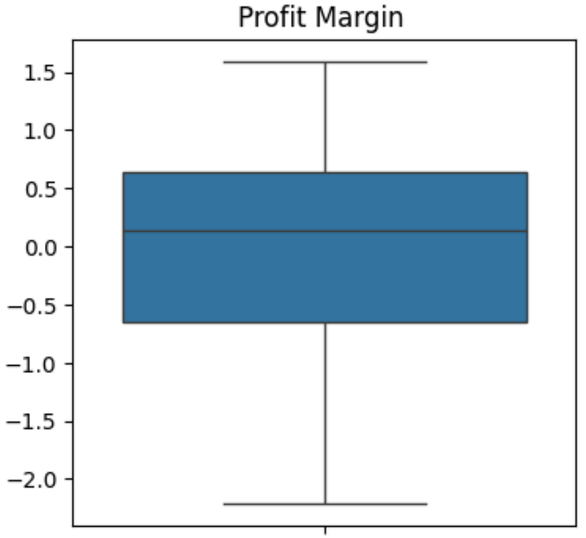

Quantity
Quantity
Ordered Values
Theoretical quantiles
$R^2 = 0.9555$
4
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4

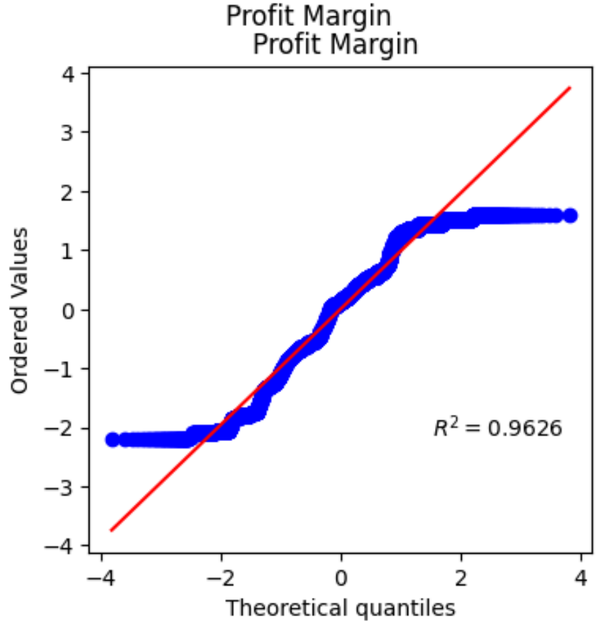

Discount
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0

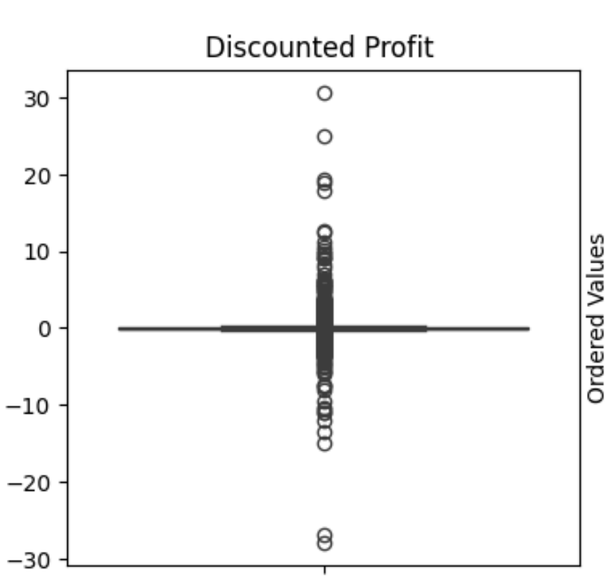

Discount
Discount
Ordered Values
Theoretical quantiles
$R^2 = 0.7751$
3
2
1
0
−1
−2
−3
−4
−2
0
2
4

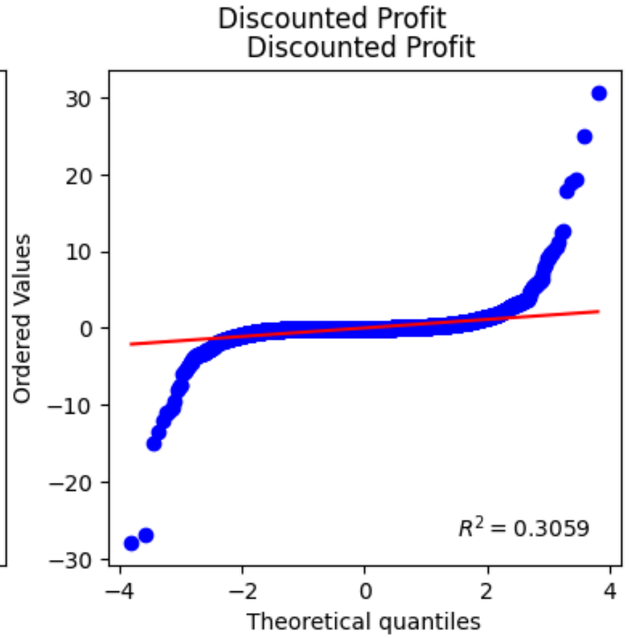

Sales
3
2
1
0
−1
−2
−3

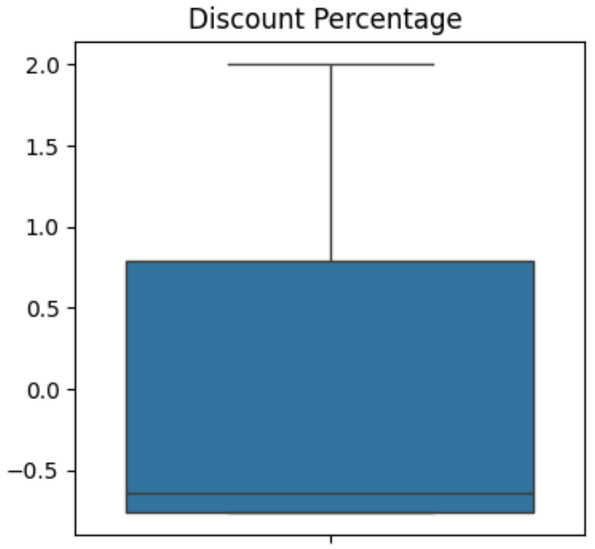

Sales
Sales
Ordered Values
Theoretical quantiles
$R^2 = 0.9903$
4
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4

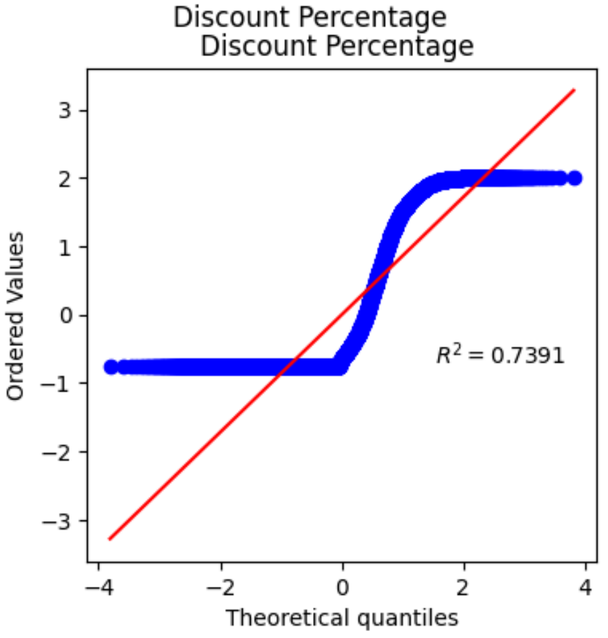

Profit Margin
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0


Profit Margin
Profit Margin
Ordered Values
Theoretical quantiles
$R^2 = 0.9626$
4
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4


Discounted Profit
30
20
10
0
−10
−20
−30


Discounted Profit
Discounted Profit
Ordered Values
Theoretical quantiles
$R^2 = 0.3059$
30
20
10
0
−10
−20
−30
−4
−2
0
2
4


Discount Percentage
2.0
1.5
1.0
0.5
0.0
−0.5


Discount Percentage
Discount Percentage
Ordered Values
Theoretical quantiles
$R^2 = 0.7391$
3
2
1
0
−1
−2
−3
−4
−2
0
2
4

SAMSUNG

Model Overview

Model Selection

We built multiple regression models to predict key metrics in sales performance.

The models tested include:

- K-Nearest Neighbors (KNN)
- Random Forest
- XGBoost
- AdaBoost

To ensure the accuracy and relevance of predictions, we used hyperparameter tuning and scaling techniques to optimize each model's performance.

Scaling Techniques

Scaling techniques are crucial in machine learning, especially when models rely on distance measurements or when features vary in magnitude.

We tested three scaling methods:

- RobustScaler: Handles outliers by using the interquartile range.
- MinMaxScaler: Scales values between 0 and 1.
- StandardScaler: Scales data to have zero mean and unit variance.

The performance of models varies based on how the input features are scaled.

Hyperparameter Tuning

Hyperparameter tuning was done using GridSearchCV, which tests different parameter combinations to find the most effective configuration.

Key hyperparameters:

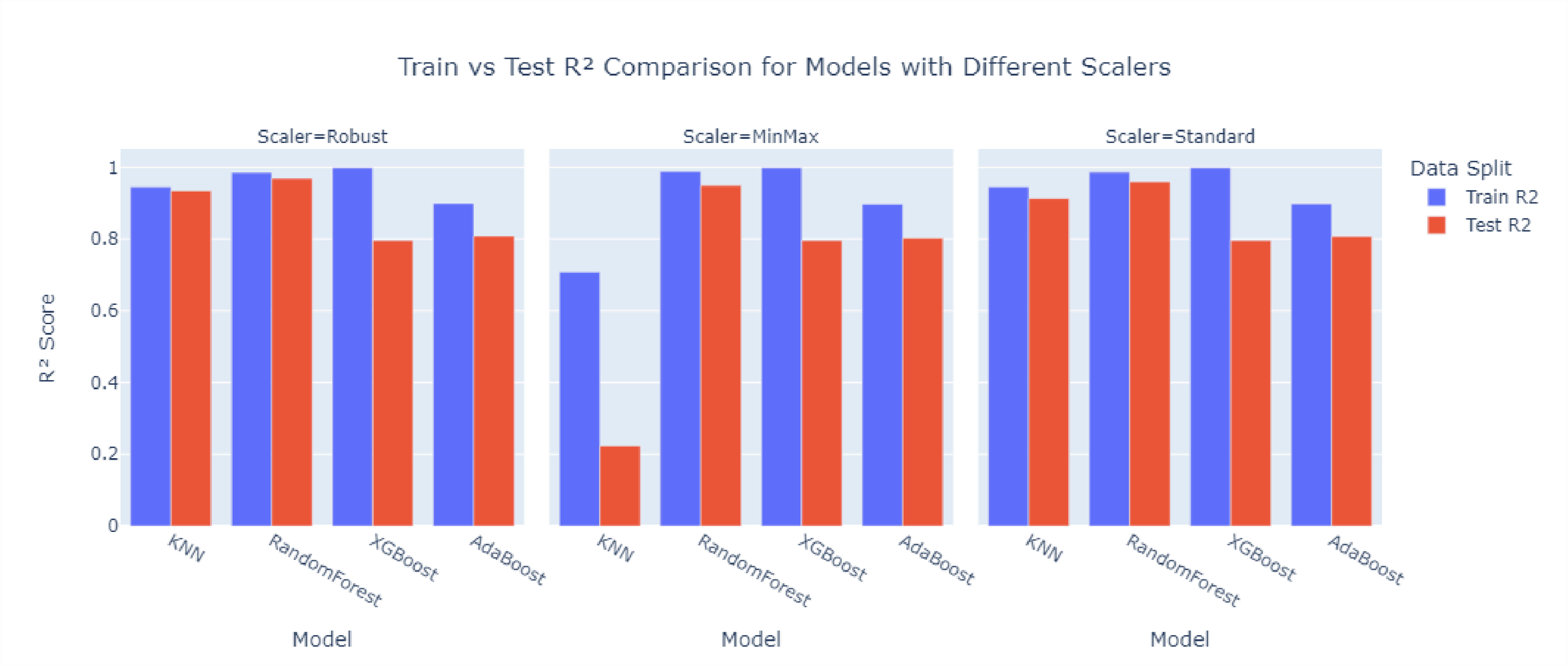
- KNN: n_neighbors (3, 5, 7)
- RandomForest: n_estimators, max_depth
- XGBoost: n_estimators, max_depth, learning_rate
- AdaBoost: n_estimators, learning_rate

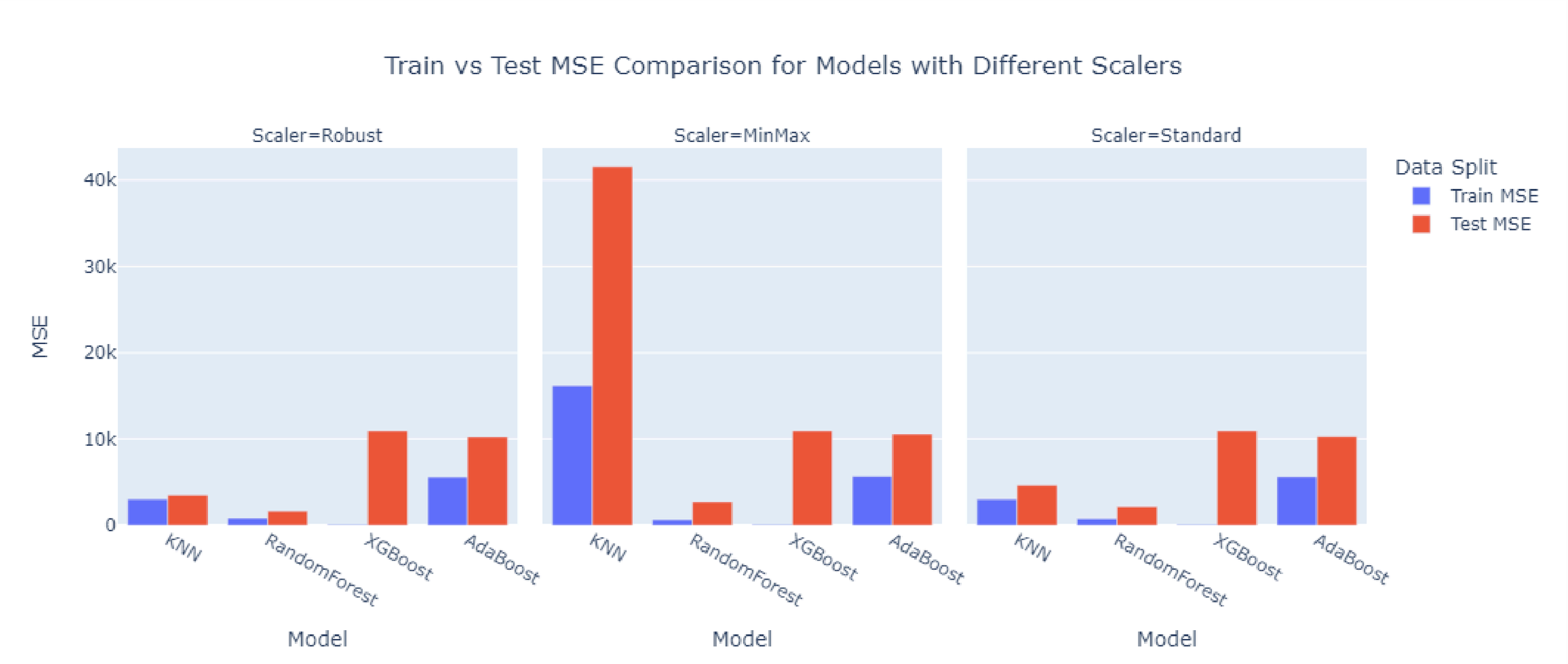
Tuning improves model accuracy, making predictions more reliable for decision-making.

SAMSUNG

MODEL EVALUATION

	Scaler	Model	Train R2	Test R2	Train MSE	Test MSE	Train RMSE	Test RMSE	Train MAE	Test MAE
0	Robust	KNN	0.944956	0.934284	3052.301997	3516.813055	55.247642	59.302724	6.040234	9.144991
1	MinMax	KNN	0.707839	0.223556	16200.985060	41551.509818	127.283090	203.841875	34.181396	57.830215
2	Standard	KNN	0.945054	0.912716	3046.896187	4671.023586	55.198697	68.344887	15.868940	24.689630
3	Robust	RandomForest	0.984758	0.968853	845.192388	1666.824503	29.072193	40.826762	1.559644	3.945505
4	MinMax	RandomForest	0.987976	0.948960	666.759451	2731.406999	25.821686	52.262864	1.427619	4.407531
5	Standard	RandomForest	0.985799	0.959706	787.455433	2156.334119	28.061636	46.436345	2.806085	5.368107
6	Robust	XGBoost	0.998386	0.795704	89.478879	10932.943725	9.459328	104.560718	3.469071	11.984223
7	MinMax	XGBoost	0.998386	0.795704	89.478879	10932.943725	9.459328	104.560718	3.469071	11.984223
8	Standard	XGBoost	0.998386	0.795704	89.478879	10932.943725	9.459328	104.560718	3.469071	11.984223
9	Robust	AdaBoost	0.898779	0.808420	5612.952096	10252.450796	74.919638	101.254387	41.215511	44.825566
10	MinMax	AdaBoost	0.897264	0.802070	5696.924893	10592.269689	75.477976	102.918753	41.451512	45.405819
11	Standard	AdaBoost	0.898097	0.807508	5650.727679	10301.236380	75.171322	101.495007	39.797622	43.257175


Train vs Test R² Comparison for Models with Different Scalers
Scaler=Robust
Scaler=MinMax
Scaler=Standard
R² Score
1
0.8
0.6
0.4
0.2
0
KNN
RandomForest
XGBoost
AdaBoost
Model
Data Split
Train R2
Test R2


Train vs Test MSE Comparison for Models with Different Scalers
Scaler=Robust
Scaler=MinMax
Scaler=Standard
MSE
40k
30k
20k
10k
0
KNN
RandomForest
XGBoost
AdaBoost
Model
Data Split
Train MSE
Test MSE

Best Models

Random Forest with Robust Scaler

Train R2: 98.5%

Test R2: 97%

KNN with Robust Scaler

Train R2: 94.5%

Test R2: 93.5%

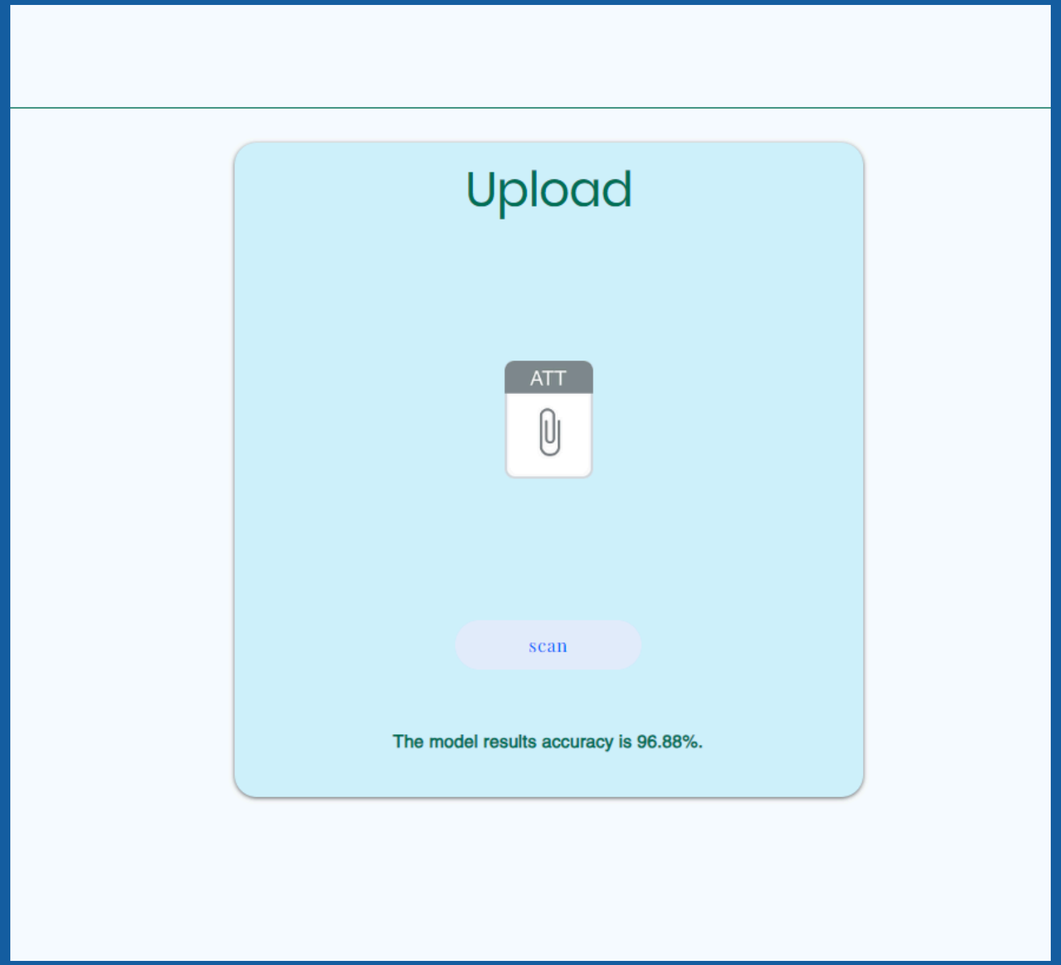
SAMSUNG

Deployment

SAMSUNG

Website

Scan me!



Dashboard


Superstore
1.49M
Sum of Sales
194.80K
Sum of Profit
25K
Sum of Quantity
0.16
Average of Discount
Order Date
11/2/2012
12/31/2014
Sum of Sales by Region
229.33K (15.39%)
480.33K (32.23%)
330.17K (22.16%)
450.38K (30.22%)
Region
West
East
Central
South
ProfitMargin
13.07
0.00
26.14
Sum of Profit by Month
Increase
Decrease
Total
Sum of Profit
200K
150K
100K
50K
0K
January
February
March
April
May
June
July
August
September
October
November
December
Total
Order Date Month
Sum of Sales by Year, Quarter and Month
Sum of Sales
0.1M
0.0M
Jan 2013
Jul 2013
Jan 2014
Jul 2014
Year
Sum of Sales by Category
Technology
Office Supplies
Furniture
Category
0.0M
0.1M
0.2M
0.3M
0.4M
0.5M
0.6M
Sum of Sales
Sum of Sales by Profit
Sum of Sales
15K
10K
5K
0K
5K
0K
5K
10K
Profit
Sum of Sales by Segment and Year
Sum of Sales
350K
300K
250K
200K
150K
100K
50K
0K
2012
2013
2014
Consumer
Corporate
Home Office
Year

SAMSUNG

THANK YOU!

SAMSUNG

Together for Tomorrow!

Enabling People

Education for Future Generations