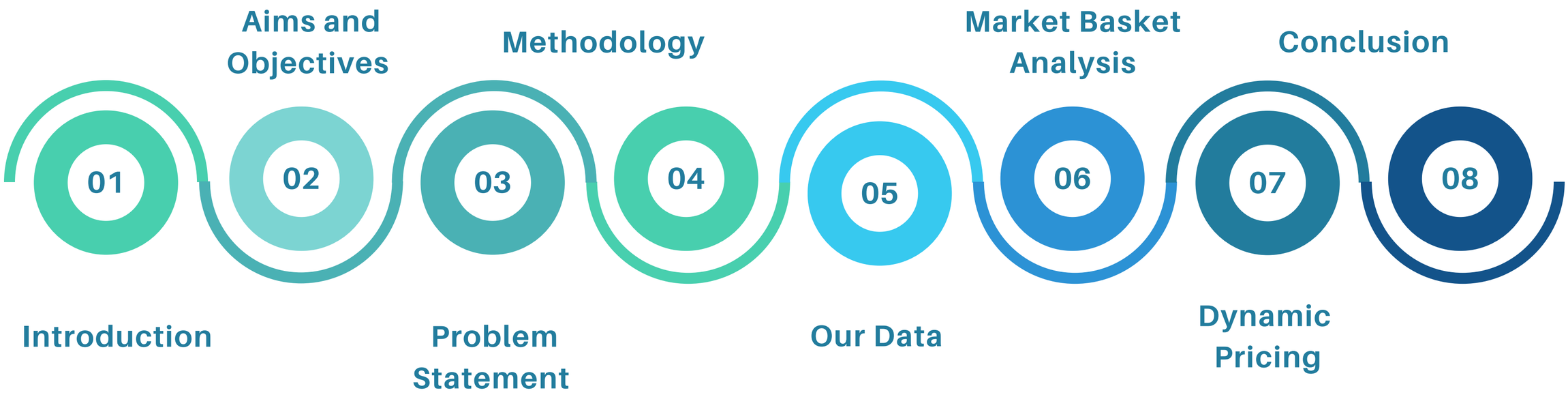
AVAILA

E-commerce Optimization Tool

Presented By

Reham Rashad El-Bayoumi
Mariam Tarek Mohammed
Mariam Ahmed Gharib
Dina Alaa Helmy

AGENDA


01
Introduction
02
Aims and Objectives
03
Problem Statement
04
Methodology
05
Our Data
06
Market Basket Analysis
07
Dynamic Pricing
08
Conclusion

INTRODUCTION

WHAT IS AVAILA?

It is an E-commerce Optimization Tool, designed to bridge the gap between seller profitability and customer satisfaction.

PROBLEM STATEMENT

Problem Statement

- Displaying goods in a random and disorganized way.
- Difficulty in locating products efficiently.
- Customers find it difficult to locate the products, which lowers demand for the business owner.
- Maximizing the demand and profit.

AIMS AND OBJECTIVES

Aims and Objectives

- Data Preparation
- Develop association rule algorithms from transaction data.
- Optimization the models to determine the most effective pricing strategies.
- Apply Apriori Algorithm.
- Test and validate effectiveness of the tool.
- Conduct simulations to evaluate the impact of dynamic pricing strategies.
- Apply ML model to predict product prices according to dynamic pricing strategies

METHODOLOGY

Methodology:

Market Basket Analysis

- A data mining technique that uncovers patterns and relationships between items purchased together, helping businesses understand customer behaviors and optimize product placement for better sales and inventory management.

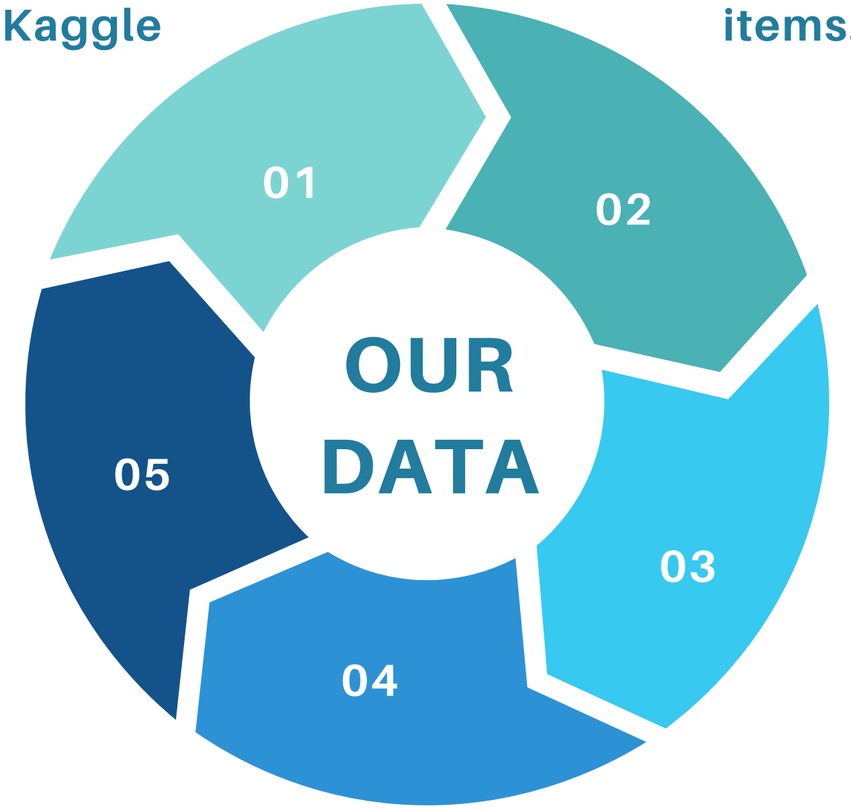
Dynamic Pricing

- A pricing strategy that aims to maximize profitability by setting optimal prices that respond to market conditions while maintaining customer satisfaction.

OUR DATA

Turkish
e-commerce website
From Kaggle

27K Turkish
supermarket
items.



Heterogeneous

81 stores
20K real Turkish names
customer, address.
Females 54.6%
Males 45.4%

1,000,000 entries
27 columns

DATA PREPROCESSING

Handling Missing Values

Handling Duplicate Data

Correcting the calculation for (TOTALPRICE) column

Change the values in USERGENDER from categorical to numerical in Binary format

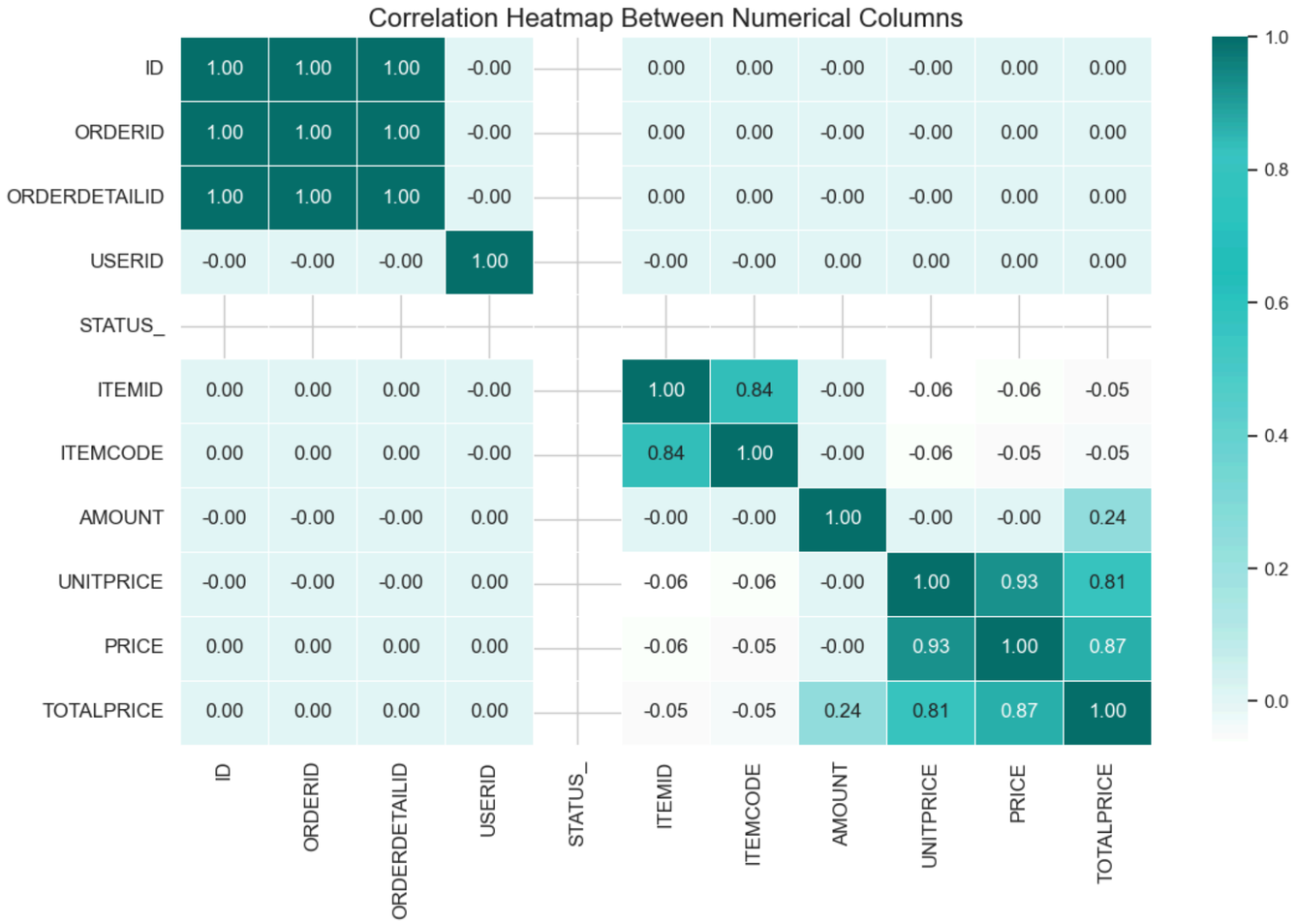
Drop useless columns

Handling the outliers

Data Visualization I

Correlation Heatmap Between Numerical Columns

To show us how much the columns depend on each other

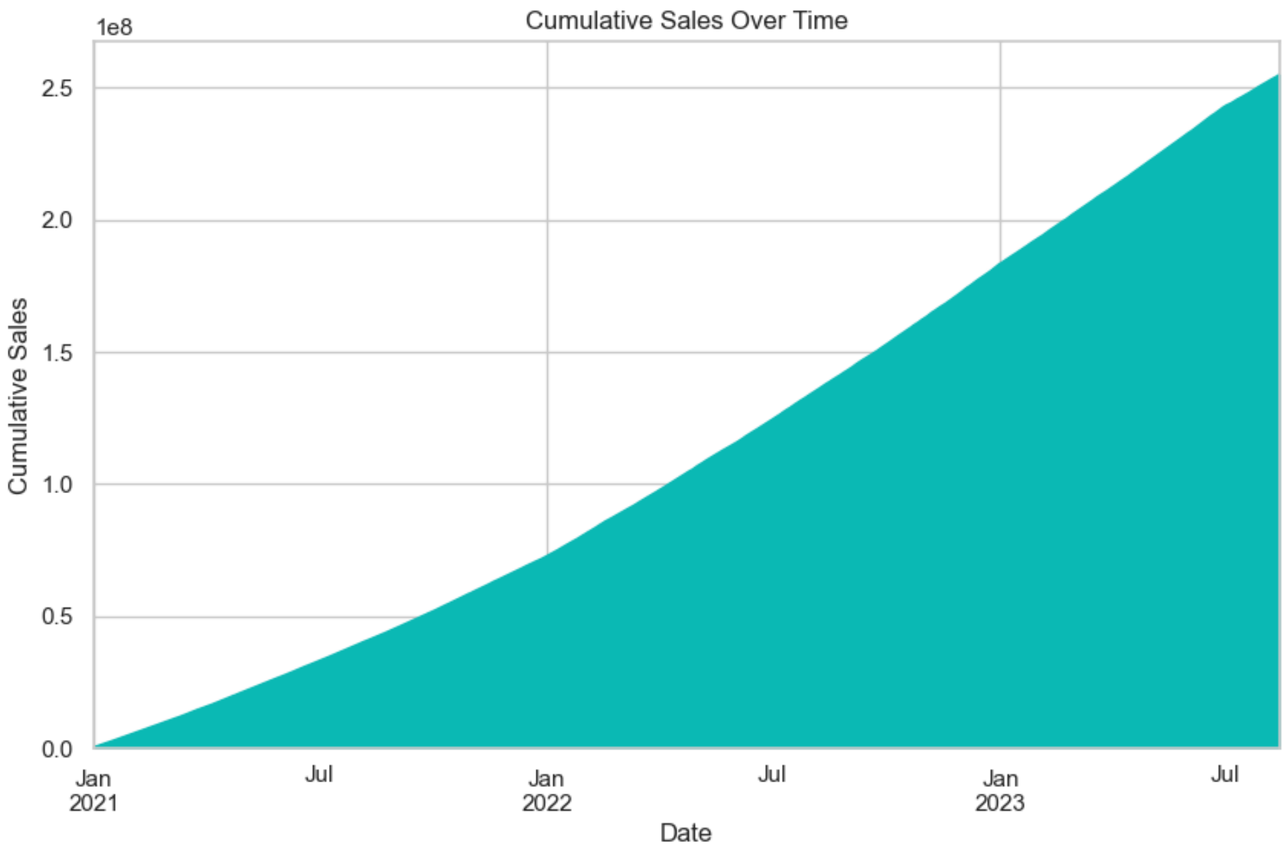


Data Visualization II

Cumulative Sales Over Time

To show the cumulative sales behavior over time

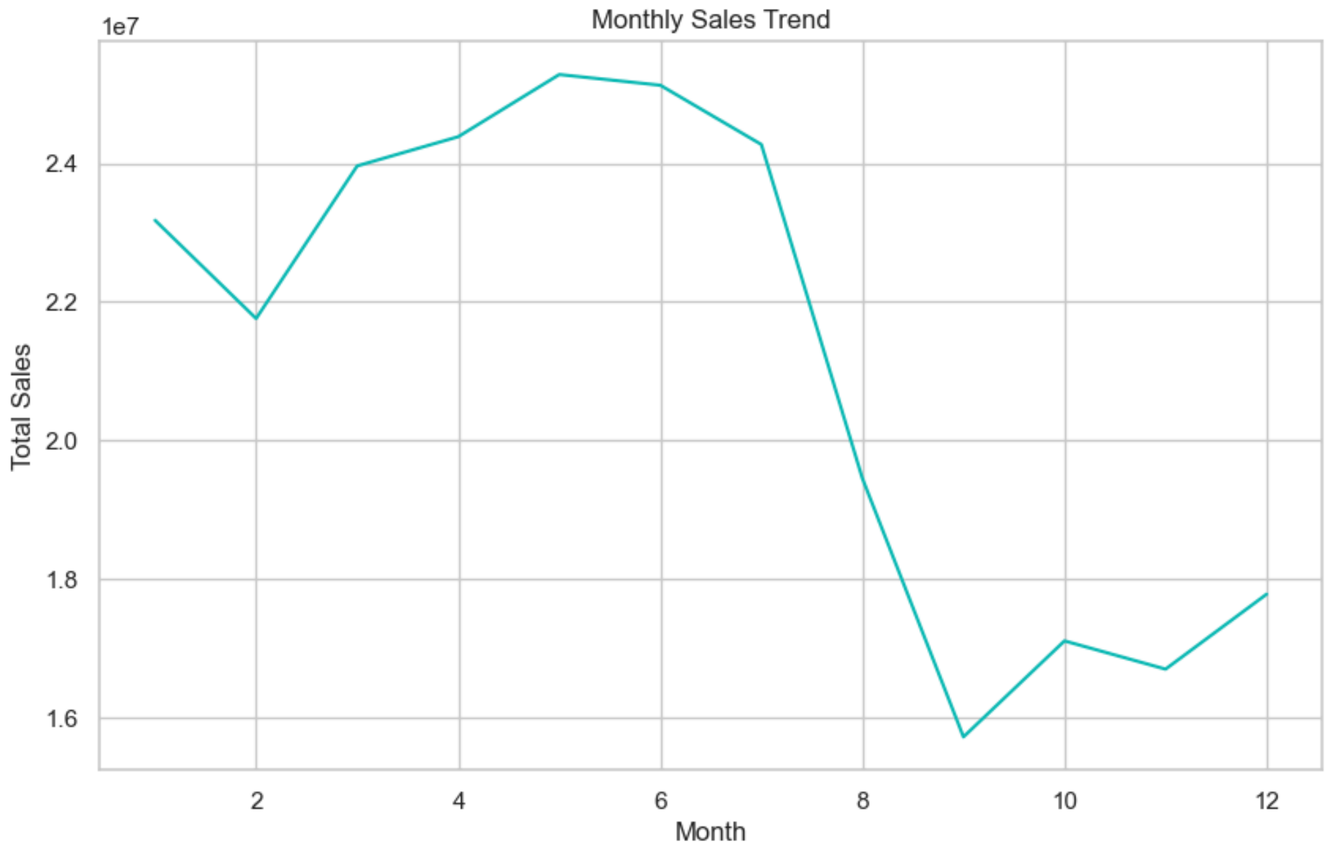
- steadily increasing over time, indicating consistent sales growth.



Data Visualization |||

Monthly Sales Trend

Explore price trends from month to month and know the selling behavior at the month level



Data Visualization |V

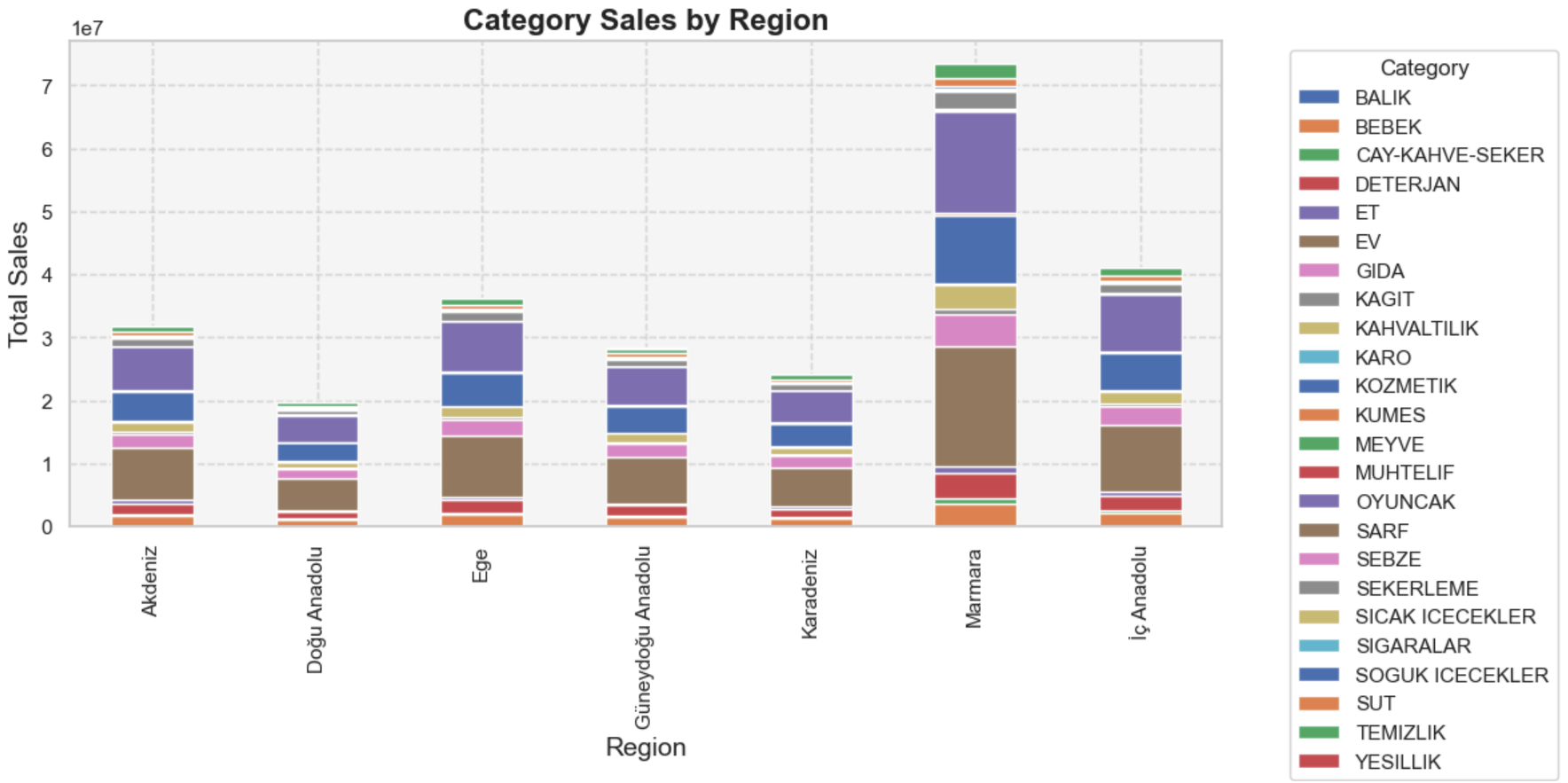
Price vs. Amount

Explore the impact of pricing on the level of customer demand for a product across our data

Data Visualization V

Category Sales by Region

- EV(Home) and Kagit (Paper) contribute to total sales over all regions.
- ET (meat) contributes especially in Marmara.
- Balik(Fish) contributes to Marmara and Akdeniz. Marmara and Akdeniz.
- Ege: has high total sales over all regions.
- ic Anadole: has balanced sales over all categories.

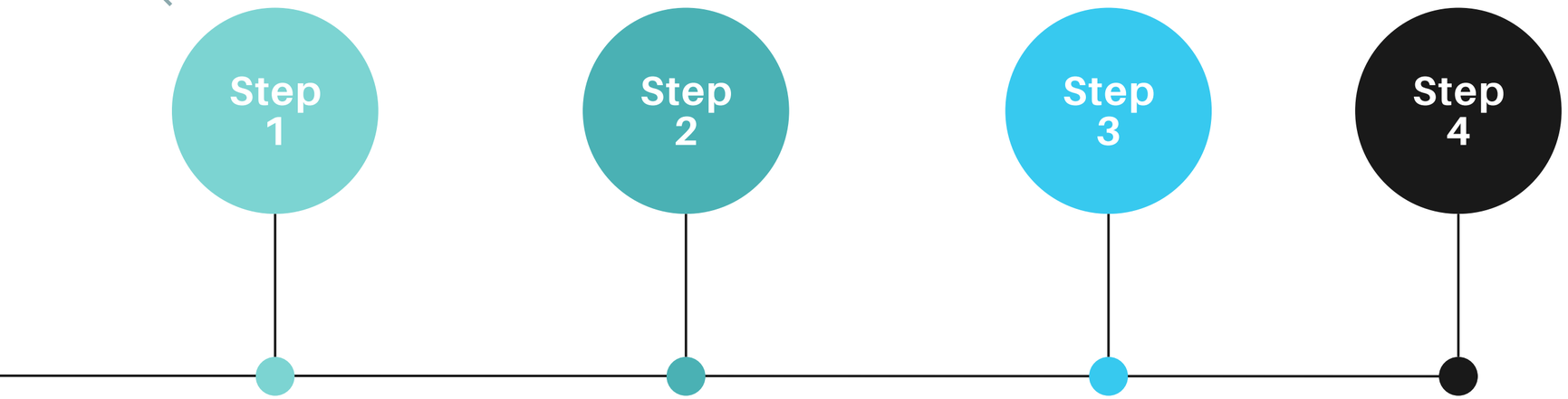


MARKET BASKET ANALYSIS

WHAT IS MARKET BASKET ANALYSIS?

Studies of customer behavior in shopping and their choices in purchasing help us understand which products go well with which. This insight leads to better product organization, which consequently improves the general customer experience.

STEPS



Encode the Transactions

Apply the Apriori Algorithm

Computing association rules

Cross-selling recommendations

Step 1

Encode the Transactions

perform one-hot encoding on a list of transactions. One-hot encoding is a process that converts categorical data into a format that can be provided to machine learning algorithm

```
orderid
1    [EV, TEMIZLIK, SEKERLEME, DETERJAN, GIDA]
2               [EV, SEKERLEME, KOZMETIK, GIDA]
3                                   [SEKERLEME]
4                             [EV, KAGIT, GIDA]
5                                     [OYUNCAK]
Name: category1, dtype: object
```

	DETERJAN	EV	GIDA	KAGIT	KOZMETIK	OYUNCAK	SEKERLEME	TEMIZLIK
0	1	1	1	0	0	0	1	1
1	0	1	1	0	1	0	1	0
2	0	0	0	0	0	0	1	0
3	0	1	1	1	0	0	0	0
4	0	0	0	0	0	1	0	0

Step 2

Apply the Apriori Algorithm

- Apriori is a popular algorithm for mining frequent itemsets and generating association rules.
- applied Apriori to transaction data encoded in a one-hot format.

Steps applied:

1. Identify all frequent itemsets of size 1 by scanning the dataset once and calculating their support.
2. Generate candidate itemsets of size 2 by joining pairs of frequent itemsets found in step 1.
3. Prune candidate itemsets that do not meet the minimum support threshold

Step 3

Computing association rules

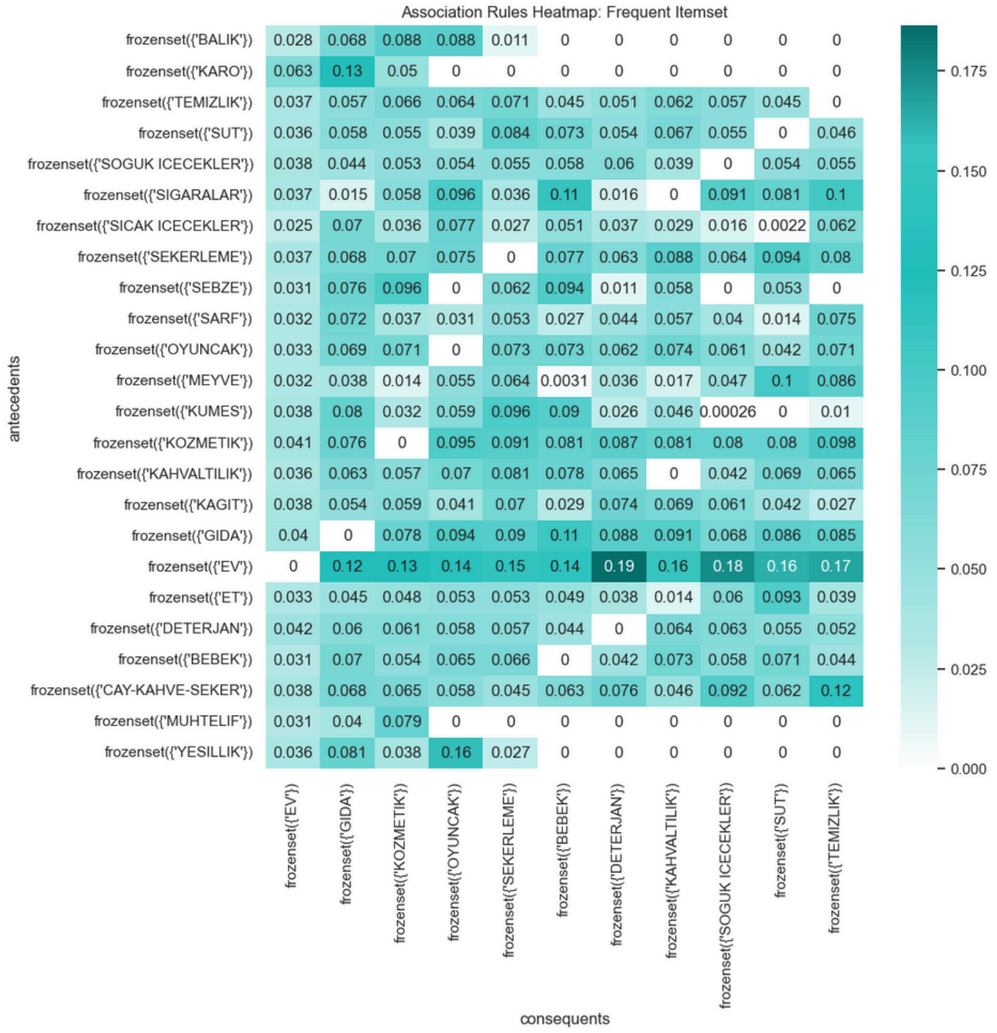
- Association rules are patterns or relationships discovered in large datasets that describe how items tend to occur together.

Steps applied:

1. applying association rules on the frequent itemsets that was extracted from Apriori.
2. filtering them based on some metrices values.

this plot explain the association between frequent itemsets based on zhangs_metric
The plot shows that there is a strong relationship between EV (home) and DETERJAN (DETERGENT) with

Association Rules Heatmap: Frequent Itemset



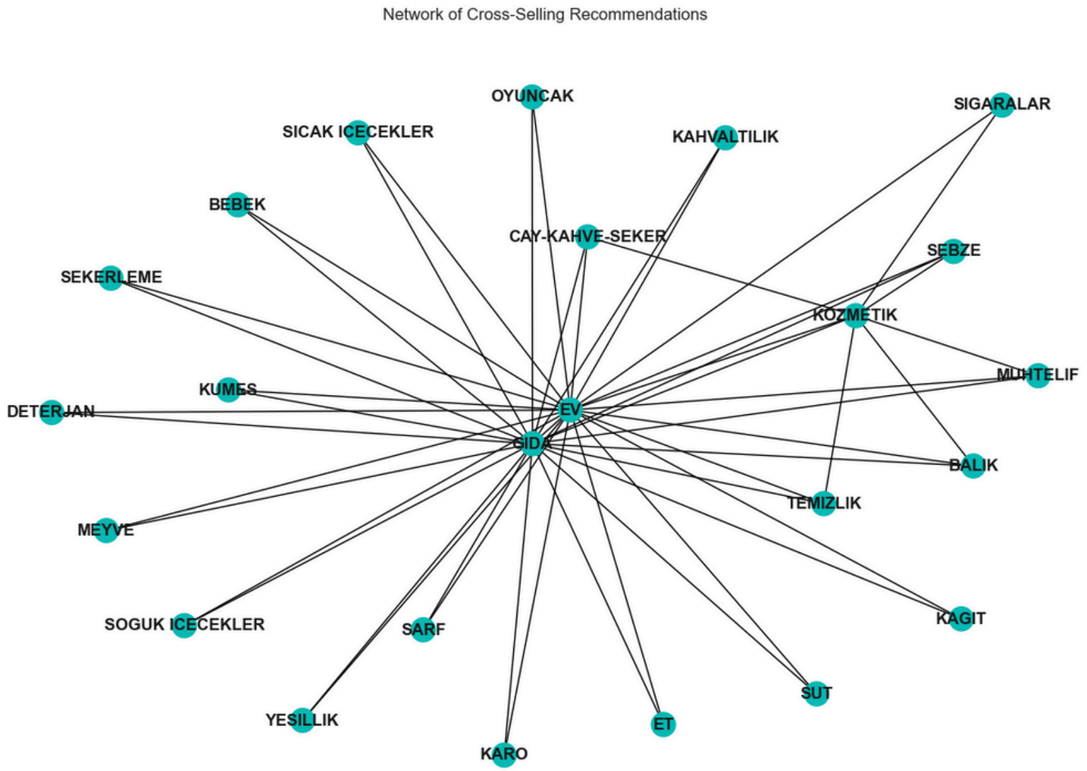
Step 4

Cross-selling Recommendations

after applying Apriori algorithm and Association rules, the cross-selling recommendation applied using the frequent item sets extracted from Apriori and use metrices of association rules to sort recommendation based on them

- Products like "KOZMETIK", "EV", and "GIDA" are central, indicating they are often bought with multiple other products.
- Products on the outer edges of the network have fewer connections, they are less commonly recommended for cross-selling

Network diagram for cross-selling recommendations



RECOMMENDATIONS

Recommend the categories that sold together

EX:

Customers who bought 'KARO' (Tile) also bought 'EV'(Home).
Customers who bought 'DETERJAN'(DETERGENT) also bought 'EV'(Home).
Customers who bought 'TEMIZLIK' (CLEANING)also bought 'EV'.
Customers who bought 'SEKERLEME' (candy)also bought 'GIDA'(food).
Customers who bought 'SICAK ICECEKLER'(hot drinks) also bought 'GIDA'.

DYNAMIC PRICING

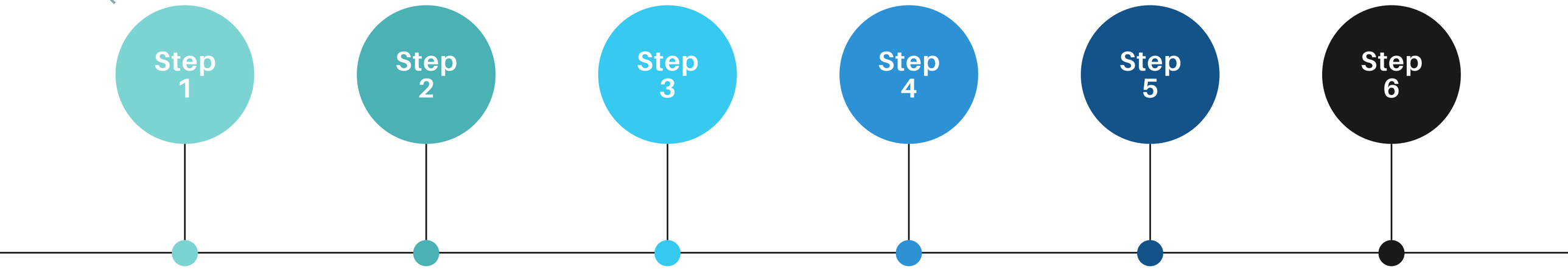
WHAT IS DYNAMIC PRICING?

This deals with flexible pricing strategies that are sensitive to market and consumer behaviors while trying to maximize both sellers' revenue and customers' satisfaction.

Types of Dynamic Pricing

- Segmented pricing
- **Time-based Pricing**
- **Bulk pricing**
- Peak Pricing
- Penetration Pricing
- Competitive pricing

STEPS



Choose the best models to train

Splitting the data
Features vs Target
Training VS Testing

Training each model

Testing each model

compare between models

Apply the best model

Choosen models to train and test I

Linear Regression:

- A good baseline model provides sensible predictions if the relationship between unit price, amount, demand, and new price is approximately linear.

Bayesian Linear Regression:

- Provides point estimates and confidence intervals, essential for understanding uncertainty and making informed pricing decisions.

Decision Tree:

- Captures complex relationships and interactions between unit price, amount, and demand.

Choosen models to train and test II

Random Forests:

- Offers robust predictions by capturing complex relationships and interactions.

XGBoost:

- Highly accurate for price predictions, especially with large and complex datasets.

CatBoost:

- Avoids overfitting with mechanisms like ordered boosting and gradient-based feature selection, ensuring better generalization for unseen data.

Splitting the data

Training vs Testing

- we use strategy 70/30, where 70% of the data is for training the models, and 30% of the data is for testing the mode

Features VS Target

while we use this model to predict the price based on some features we split the data to:

- Features to train the model: Demand, Amount, and UnitPrice (original price of the product)
- Target: price that we will predict

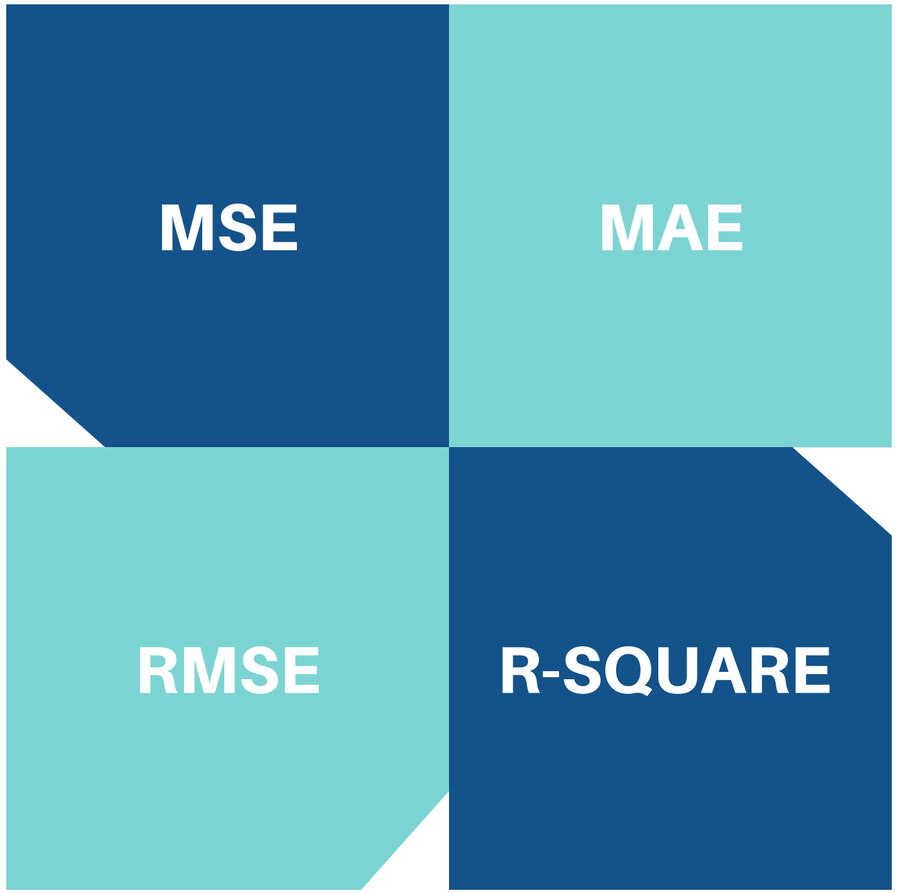
Testing the model's criteria

Mean Square Error

Represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals

Root Mean Squared Error

the square root of Mean Squared error. It measures the .standard deviation of residuals



Mean Absolute Error

Represents the average of the absolute difference between the actual and predicted values in the dataset. It measures the average of the residuals in the dataset.

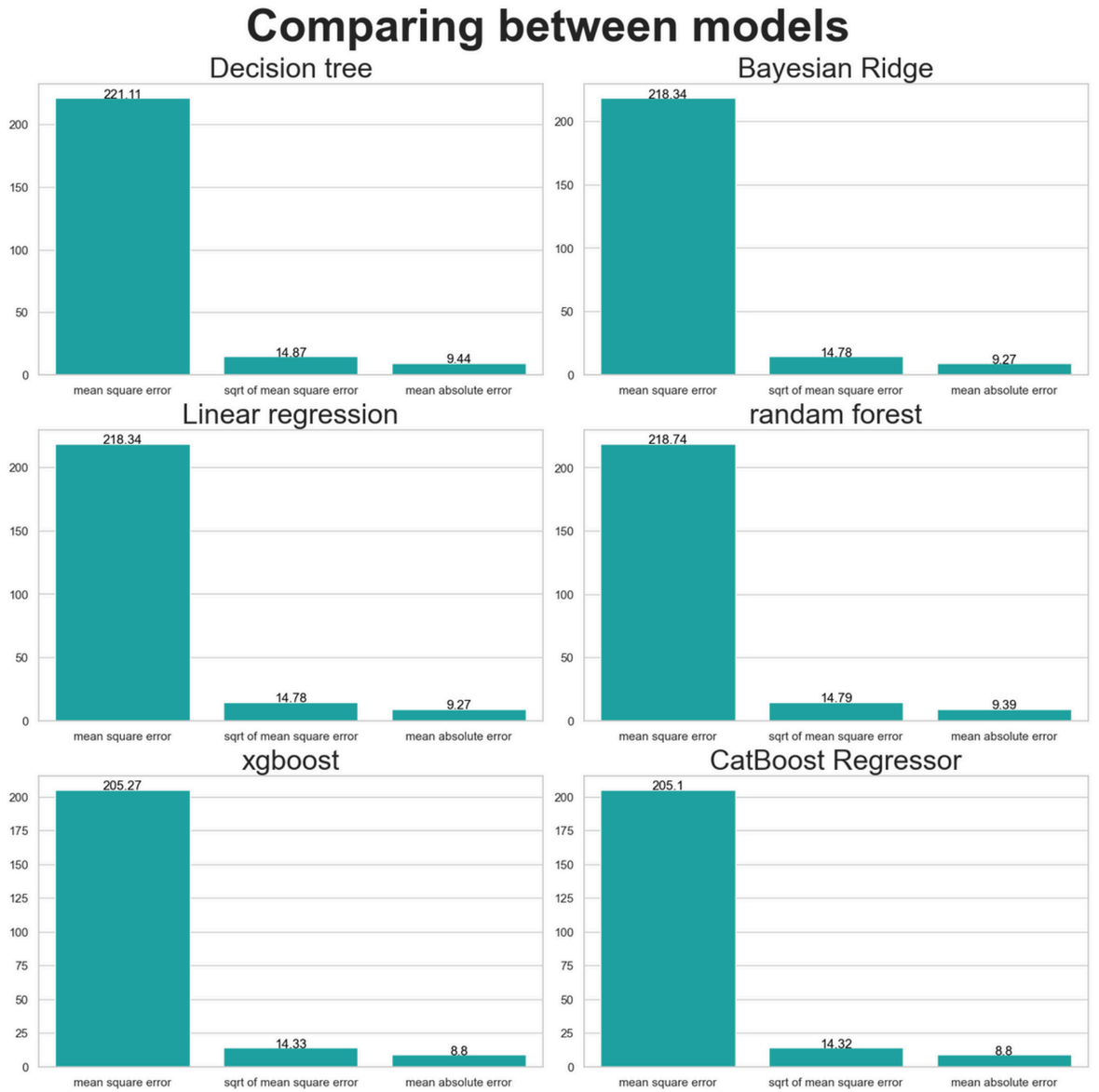
R-SQQUARE

represents the proportion of the variance in the dependent variable which is explained by the linear regression model. It is a scale-free score i.e. irrespective of the values being small or large, the value of R square will be less than one.

Comparing between the models

R ^ 2

CatBoost Model	0.812599
XGB Model	0.812449
Bayesian Ridge	0.800502
Linear Regression	0.800502
Random Forest	0.800140
Decision Tree	0.797974



Apply the best Model (CatBoost Regressor)

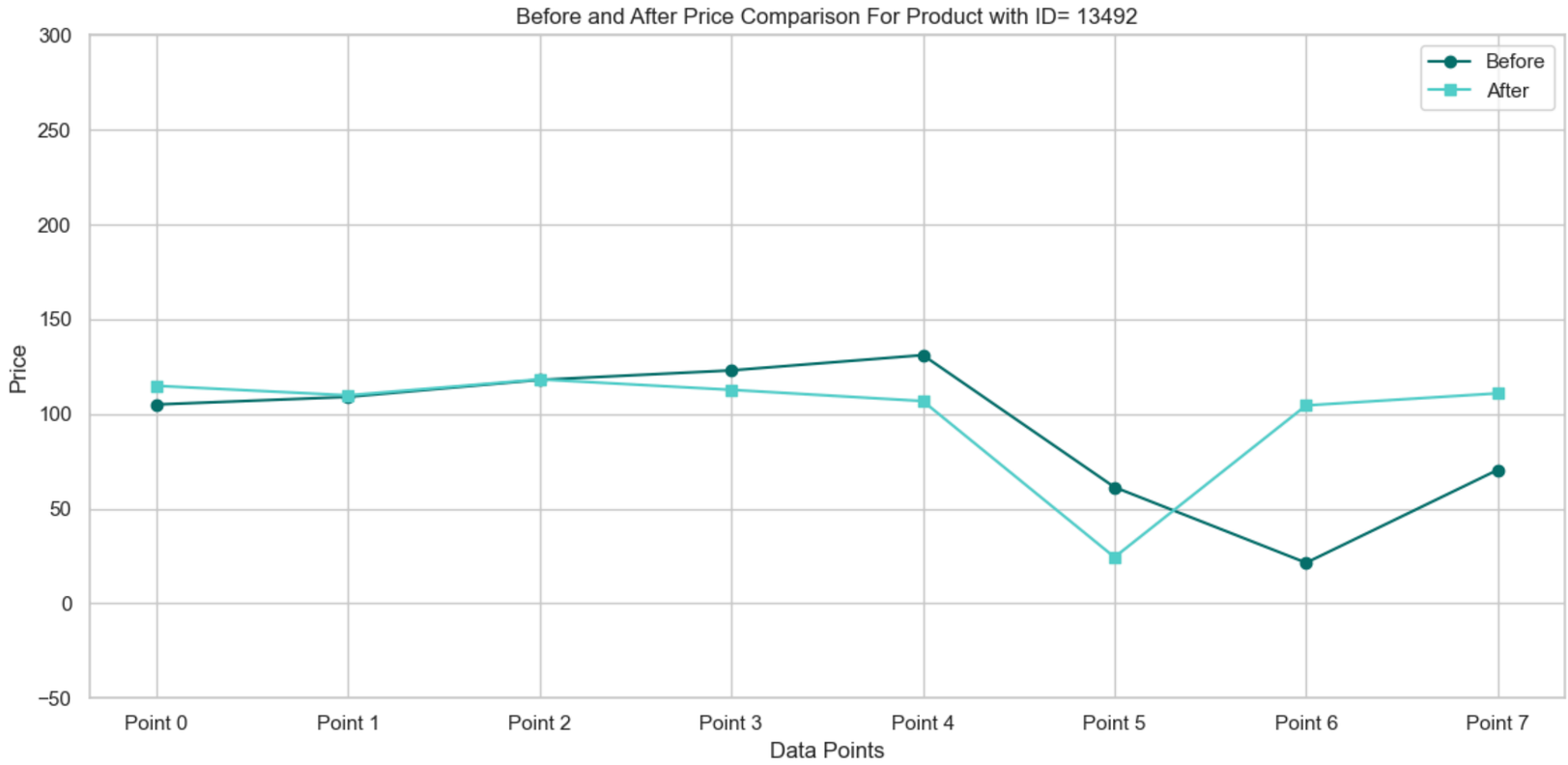
have the best values in testing criteria with a maximum R-square and a minimum for others:

- MSE= 205.1027475948813
- RMSE= 14.321408715447.
- MAE= 8.798734691965825.
- R-square= 0.8125990653982991.
- The score is 81.3%.

Fastest, more accurate model because it has in-built mechanisms to avoid overfitting and can work well in our case because it can handle complex relationships between featues and targets.

Example: Before and After Price Comparison For product ID = 13492

Total difference in price is **62.94**



The Link Between MBA and DP results I

Identifying Complementary Products

- **MBA Insight**: MBA will be able to pick up pairs or sets of products that go together in sales.
- **Dynamic Pricing Application**: This information is useful to retailers in developing bundled pricing strategies.

Optimizing Cross-Selling Opportunities

- **MBA Insight**: MBA provides insight into which pair of products are usually bought together.
- **Dynamic Pricing Application**: While the customer is making the purchase, the system can dynamically alter prices or may even indicate discounts to the relevant ones identified through MBA, so that the opportunity for cross-selling is enhanced more.

The Link Between MBA and DP results II

For example: using product ID = 7438 from category EV and product ID = 18171 from category KARO, the original prices for both are 32.9 and 50.0 if the first product will be sold 5 items in 2021-04-19 the price of the product will be 20.5, and in the same day the second one price will be 43.5, we have two ways to recommend products in that case:

- Buy both items in a bundle with a price of 20.5+43.5 = 46 and the original price was 32.9+50=83, so he can buy it in a bundle with a discount of 23%
- Recommend the second one to him with a discount of 13% and a price of 43.5

Conclusion

- The E-commerce Optimization Tool bridges the gap between seller profitability and customer satisfaction.
- Through dynamic pricing and market basket analysis, our tool empowers businesses with data-driven insights to optimize their e-commerce strategies.

THANK YOU!