Amazon Dynamic Pricing Recommendation Analysis

Pooja Shreni Addisherla 100002917
Divya Vinodbhai Koriya 100002007

CONTENTS

01 | Business to Data Understanding

02 | Data Preparation

03 | Modeling Approaches

04 | Results & Insights

05 | Business Impact & Dashboard

Predicting Amazon Recommendations

A Machine Learning Project Following CRISP-DM Methodology to predict customer likelihood to recommend Amazon based on demographics, behavior, and pricing perception.

Goal
Predict recommendation likelihood.

Data
5,000-record survey dataset.

Method
CRISP-DM Methodology.

Whether a customer is likely to recommend Amazon based on their perception of Amazon's dynamic pricing strategy.

Dataset Structure Overview

11 Columns
Customer_ID, Age, Gender, Location, Annual_Income...

Target Variable
Likely_to_Recommend_Amazon_Based_on_Pricing

	Customer_ID	Age	Gender	Location	Annual_Income	Browsing_Time_per_Week_Hours	Purchase_Frequency_Per_Month	Impact_of_Dynamic_Pricing_on_Purchase	Perception_of_Amazon_Revenue_Growth_due_to_Dynamic_Pricing	Perception_of_Competition_in_Amazon_Marketplace	Likely_to_Recommend_Amazon_Based_on_Pricing
0	1	69	Other	Australia	82187	13.61	8	High Impact	Negative Impact	Decreased Competition	Highly Likely
1	2	57	Other	Europe	101939	10.83	12	Moderate Impact	Significant Growth	No Change	Likely
2	3	60	Female	North America	79316	20.47	3	Moderate Impact	No Growth	Increased Competition	Highly Likely
3	4	67	Other	Australia	26415	4.64	8	Moderate Impact	No Growth	Decreased Competition	Highly Unlikely
4	5	43	Female	Australia	145038	10.49	9	High Impact	Moderate Growth	Increased Competition	Unlikely

Target Variable Distribution

- The target variable is **well balanced**, with each class representing about **24–26%** of customers.

- Customer sentiment toward Amazon’s dynamic pricing is **evenly split** between positive and negative recommendations.
- **“Highly Likely”** is the largest class, but only slightly, showing no strong dominance.

- This balanced distribution makes the target variable **ideal for multi-class classification** and reliable model training.

```
============================================================
TARGET VARIABLE DISTRIBUTION
============================================================
Likely_to_Recommend_Amazon_Based_on_Pricing
Highly Likely      1310
Highly Unlikely    1232
Unlikely           1232
Likely             1226
Name: count, dtype: int64

Distribution Percentage:
Likely_to_Recommend_Amazon_Based_on_Pricing
Highly Likely      26.20
Highly Unlikely    24.64
Unlikely           24.64
Likely             24.52
Name: count, dtype: float64
```

Data Preparation (EDA)

- Feature engineering was performed to **transform raw numerical variables into meaningful categories**.
- **Income_Category** captures spending power differences among customers.
- **Age_Group** helps model behavioral variations across life stages.
- **Browsing_Category** and **Purchase_Freq_Category** summarize engagement and shopping intensity, improving model
- interpretability and performance.

```
Number of duplicate rows: 0

============================================================
DATA TYPES
============================================================
Customer_ID                                                 int64
Age                                                         int64
Gender                                                     object
Location                                                   object
Annual_Income                                               int64
Browsing_Time_per_Week_Hours                              float64
Purchase_Frequency_Per_Month                                int64
Impact_of_Dynamic_Pricing_on_Purchase                      object
Perception_of_Amazon_Revenue_Growth_due_to_Dynamic_Pricing object
Perception_of_Competition_in_Amazon_Marketplace            object
Likely_to_Recommend_Amazon_Based_on_Pricing                object
dtype: object
```

```
============================================================
CATEGORICAL VARIABLES - UNIQUE VALUES
============================================================

Gender:
Gender
Female    1690
Other     1674
Male      1636
Name: count, dtype: int64

Location:
Location
South America    1028
Australia        1001
Asia             1000
Europe            994
North America     977
Name: count, dtype: int64

Impact_of_Dynamic_Pricing_on_Purchase:
Impact_of_Dynamic_Pricing_on_Purchase
No Impact          1272
Moderate Impact    1264
High Impact        1234
Low Impact         1230
Name: count, dtype: int64
```

Train-Test Split Strategy

5,000
Total Samples

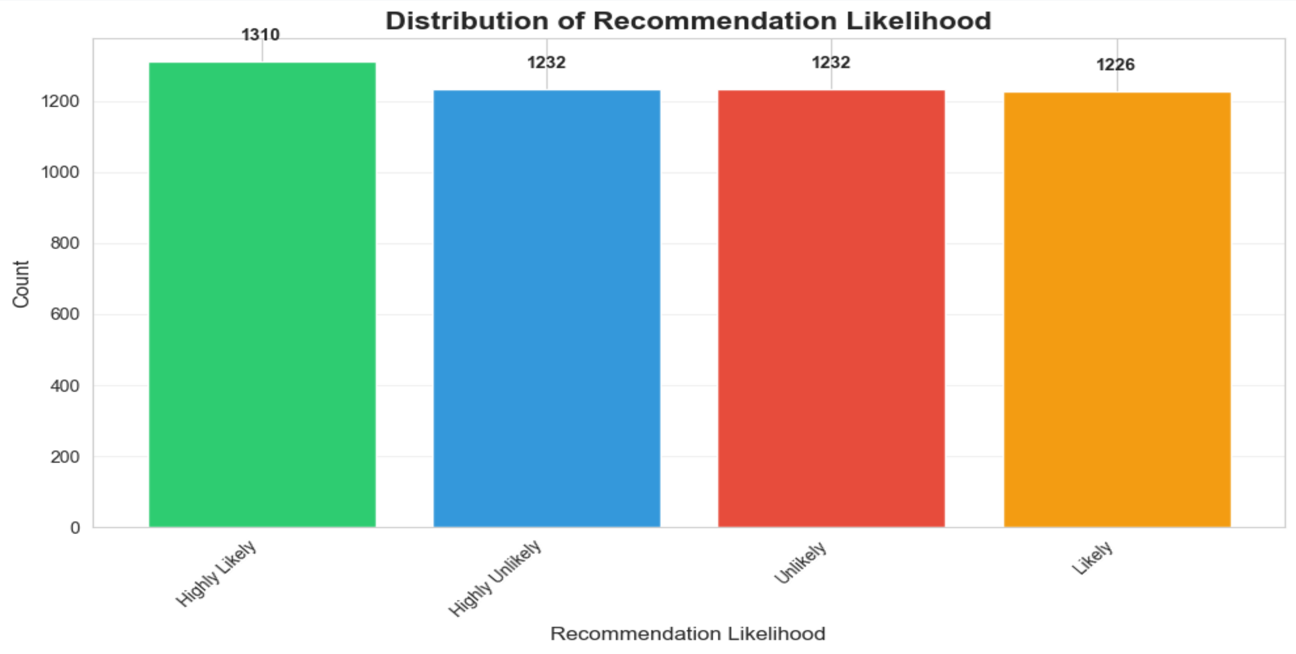
→

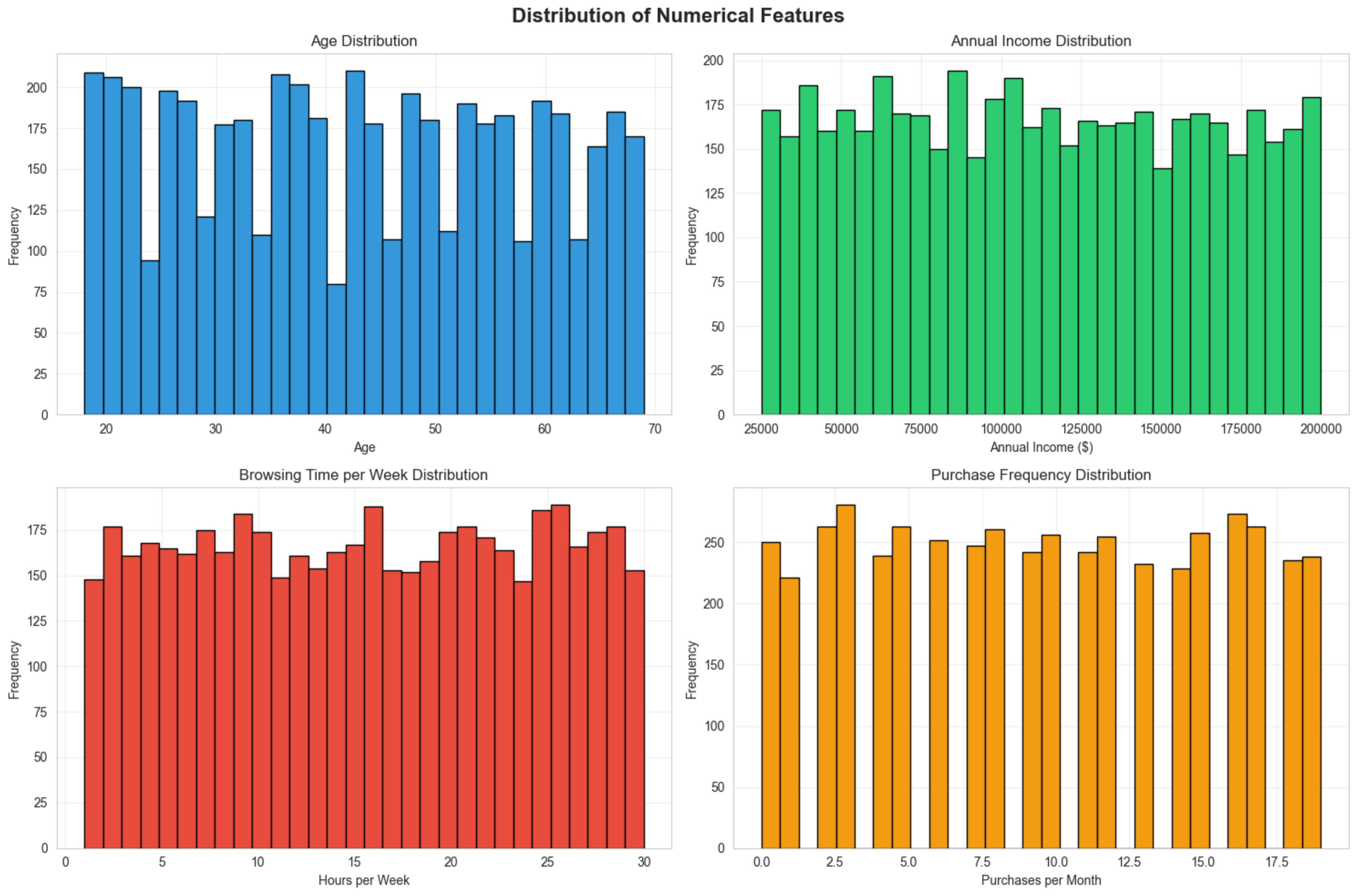
4,000
Training (80%)

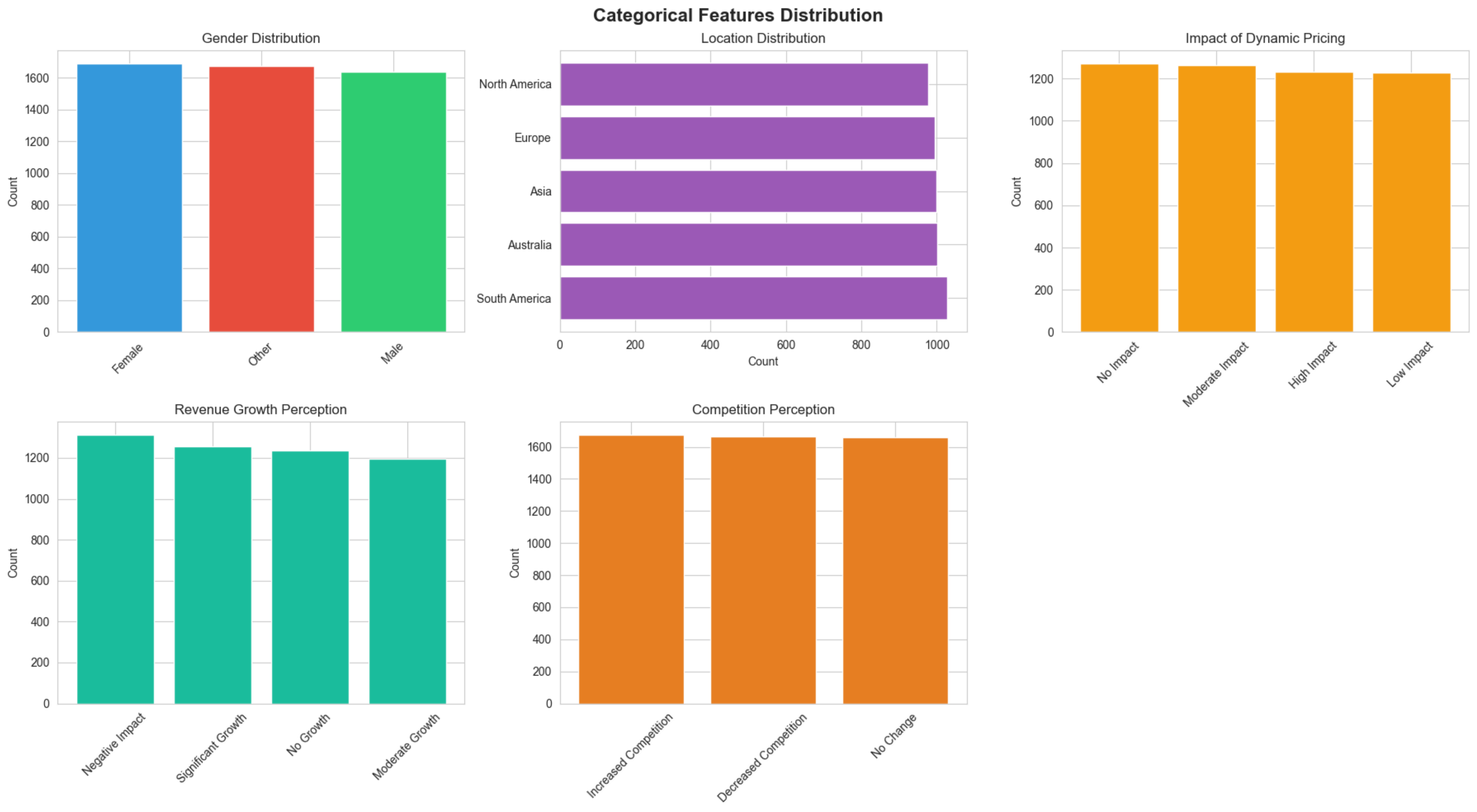
1,000
Test (20%)

Used stratified sampling to maintain class distribution.

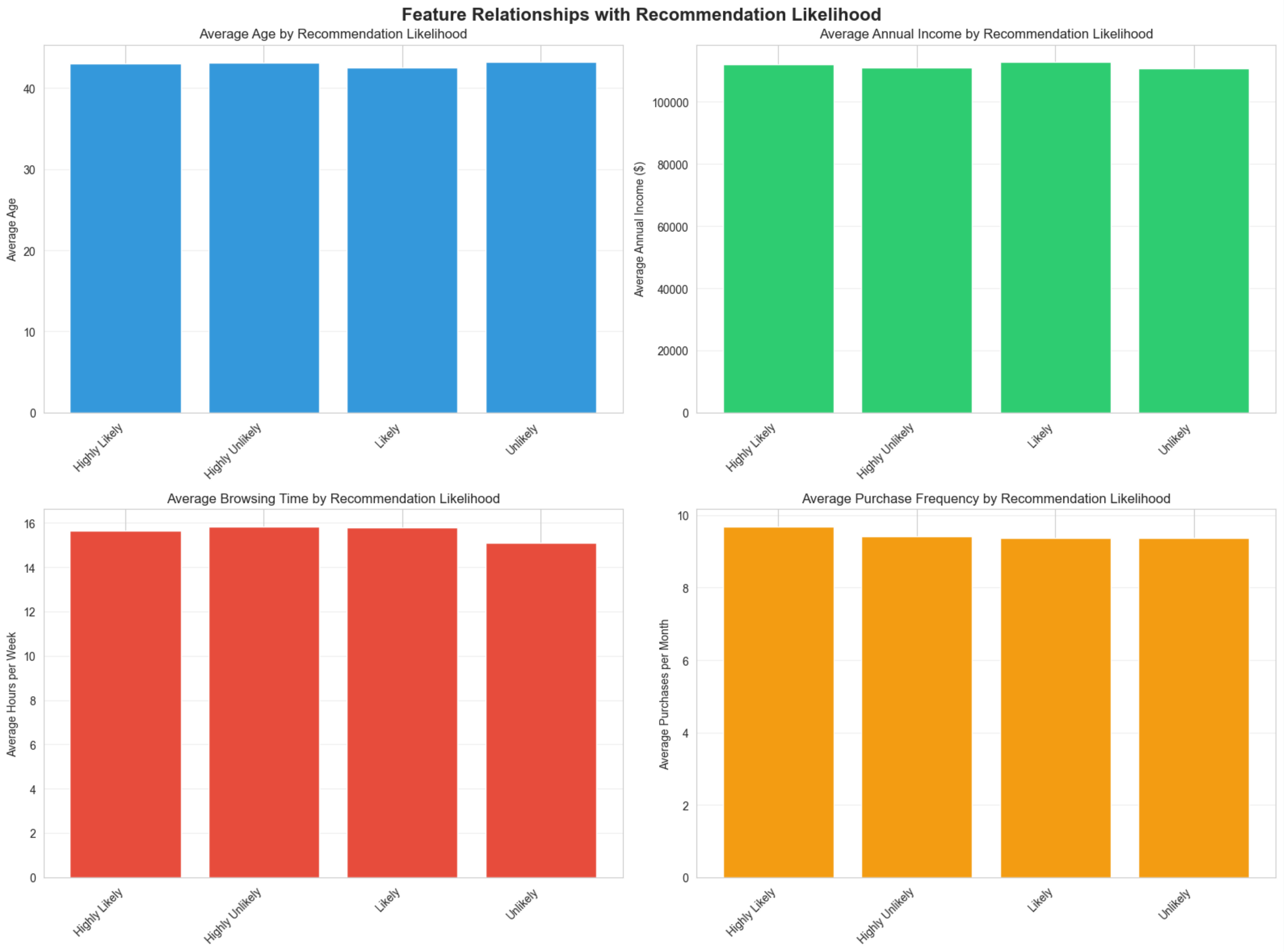
- **Distribution of Recommendation Likelihood**
- The recommendation likelihood is **fairly balanced** across all four categories: **Highly Likely, Likely, Unlikely, Highly Unlikely.**
- **Highly Likely** has a **slightly higher count**, indicating a marginally positive overall customer sentiment.
- This balance suggests the dataset is **well-distributed**, making it suitable for training unbiased recommendation or pricing models.




Distribution of Numerical Features
Age Distribution
Frequency
Age
Annual Income Distribution
Annual Income ($)
Browsing Time per Week Distribution
Hours per Week
Purchase Frequency Distribution
Purchases per Month

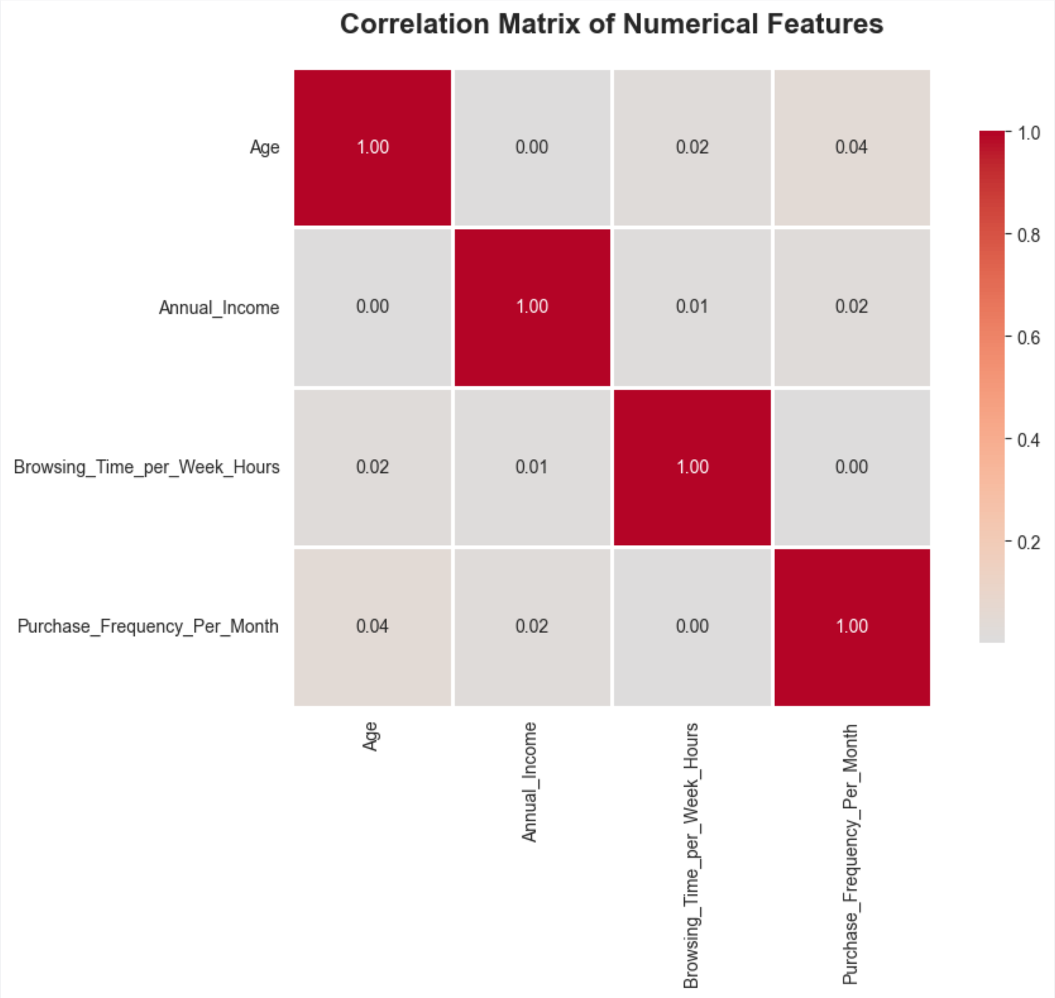


Categorical Features Distribution
Gender Distribution
Count
Female
Other
Male
Location Distribution
North America
Europe
Asia
Australia
South America
Count
Impact of Dynamic Pricing
No Impact
Moderate Impact
High Impact
Low Impact
Revenue Growth Perception
Negative Impact
Significant Growth
No Growth
Moderate Growth
Competition Perception
Increased Competition
Decreased Competition
No Change


Feature Relationships with Recommendation Likelihood
Average Age by Recommendation Likelihood
Average Age
0
10
20
30
40
Highly Likely
Highly Unlikely
Likely
Unlikely
Average Annual Income by Recommendation Likelihood
Average Annual Income ($)
0
20000
40000
60000
80000
100000
Highly Likely
Highly Unlikely
Likely
Unlikely
Average Browsing Time by Recommendation Likelihood
Average Hours per Week
0
2
4
6
8
10
12
14
16
Highly Likely
Highly Unlikely
Likely
Unlikely
Average Purchase Frequency by Recommendation Likelihood
Average Purchases per Month
0
2
4
6
8
10
Highly Likely
Highly Unlikely
Likely
Unlikely

Correlation Heatmap

- All numerical features show **very low correlations** with each other (values close to 0).
- **Age** has almost no relationship with income, browsing time, or purchase frequency.
- **Annual income** is weakly related to browsing and purchasing behavior.
- **Browsing time** and **purchase frequency** are nearly independent.
- Indicates **low multicollinearity**, which is ideal for machine learning models.
- Each feature contributes **independent information** to dynamic pricing and recommendation systems.



Modeling Approach

Algorithm Selection

Random Forest

Ensemble method, robust to overfitting.

Gradient Boosting

Sequential learning, high accuracy.

Decision Tree

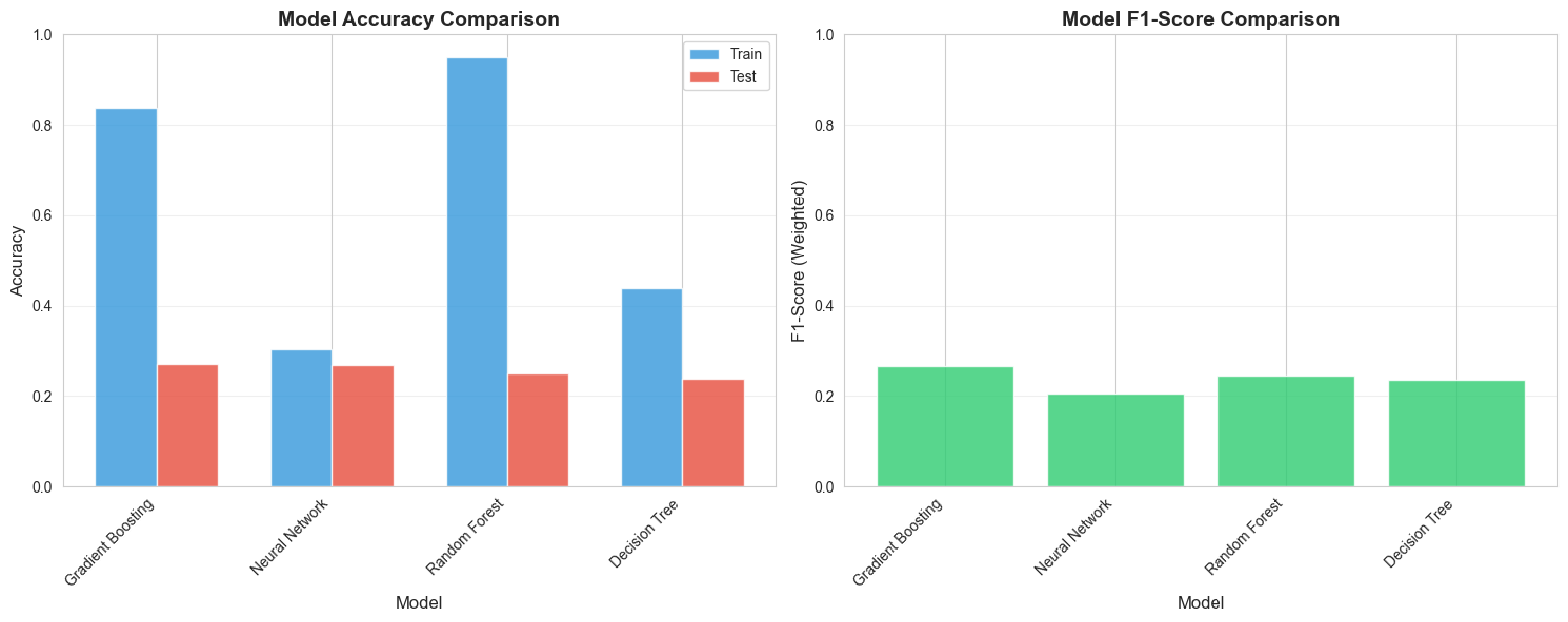
Interpretable, simple baseline.

MLP Classifier

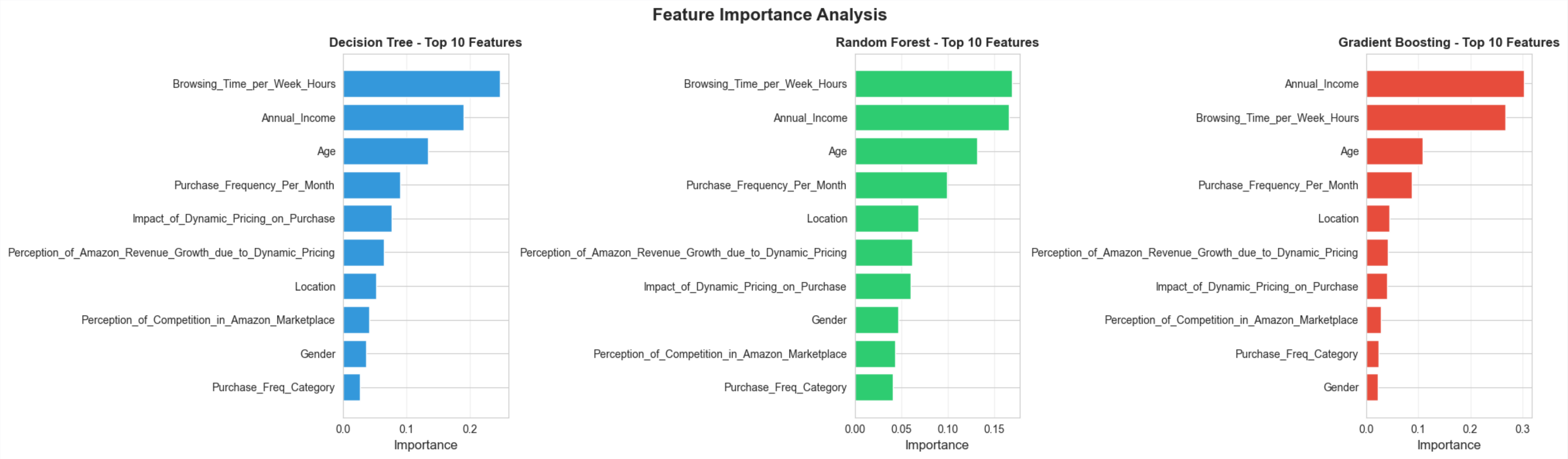
Neural network, complex patterns.

==
MODEL COMPARISON
==

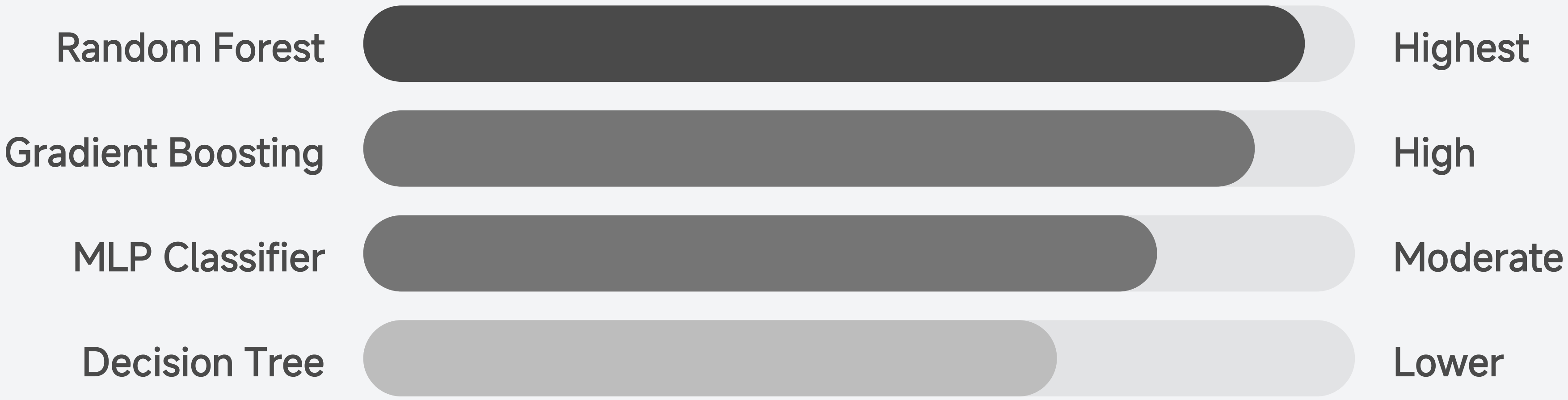
Model	Train Accuracy	Test Accuracy	Test F1-Score
Gradient Boosting	0.83800	0.270	0.267112
Neural Network	0.30225	0.269	0.204855
Random Forest	0.95050	0.249	0.245222
Decision Tree	0.43800	0.238	0.236400



- Browsing time and purchase behaviour are the strongest predictors of recommendation likelihood.
- Pricing perception features matter more than demographic details like gender or location.
- Ensemble models (Random Forest, Gradient Boosting) capture feature importance more reliably than Decision Tree



Cross-Validation Results



Performance indicates potential overfitting for Decision Tree.

Results & Insights

Model Performance Comparison

Random Forest

Best Performance
(Highest Accuracy & F1-Score)

Gradient Boosting

Second Best

MLP Classifier

Competitive

Decision Tree

Suffered from Overfitting

Key Business Findings

- **Perception is Paramount:** Pricing perception outweighs demographics in predicting recommendations.
- **Strongest Predictor:** The perceived impact of dynamic pricing is the most critical factor.
- **Strategic Focus:** Prioritize transparent communication about pricing strategies.

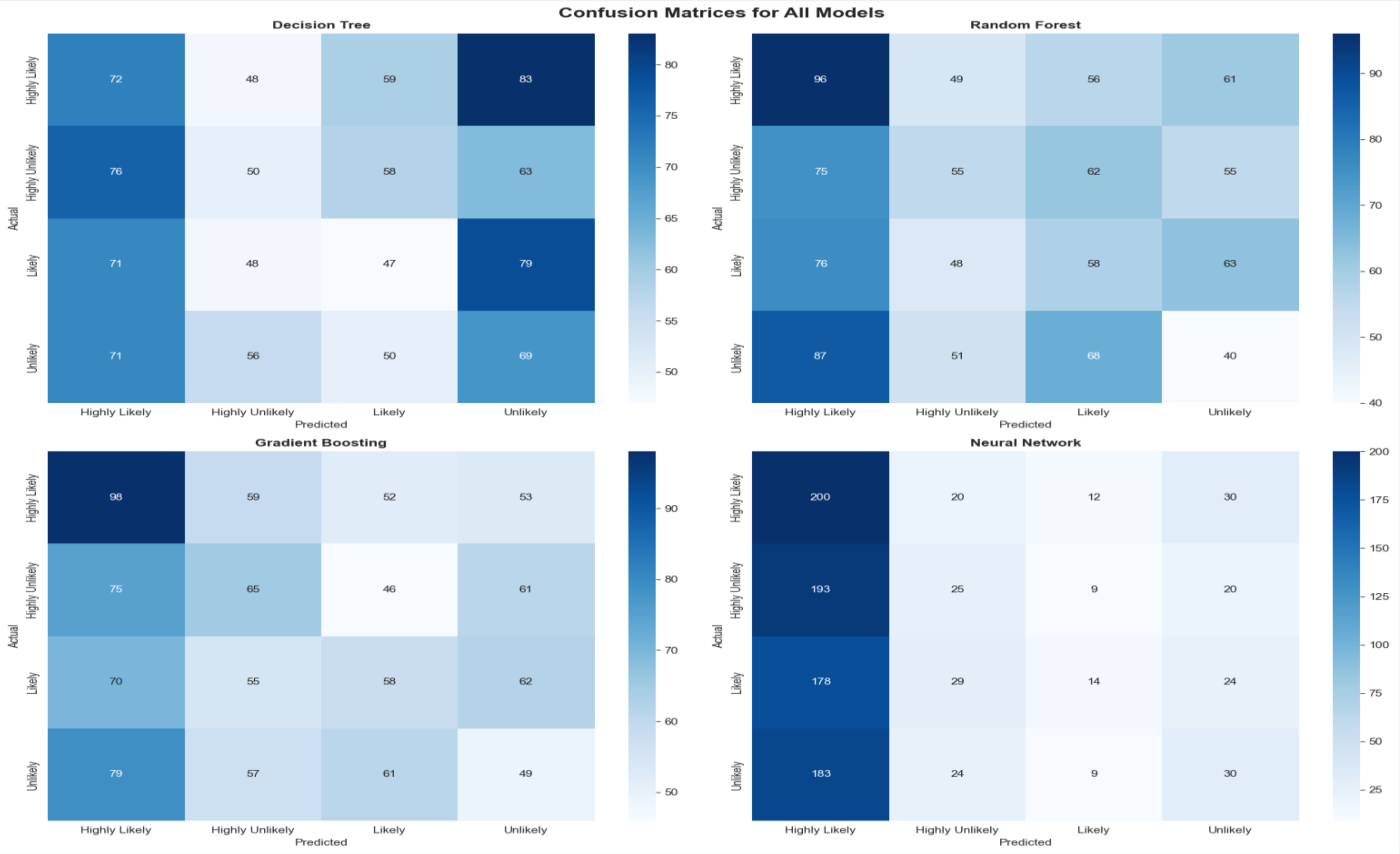
Confusion Matrix Insights

Strong Diagonal Performance

High accuracy for each class.

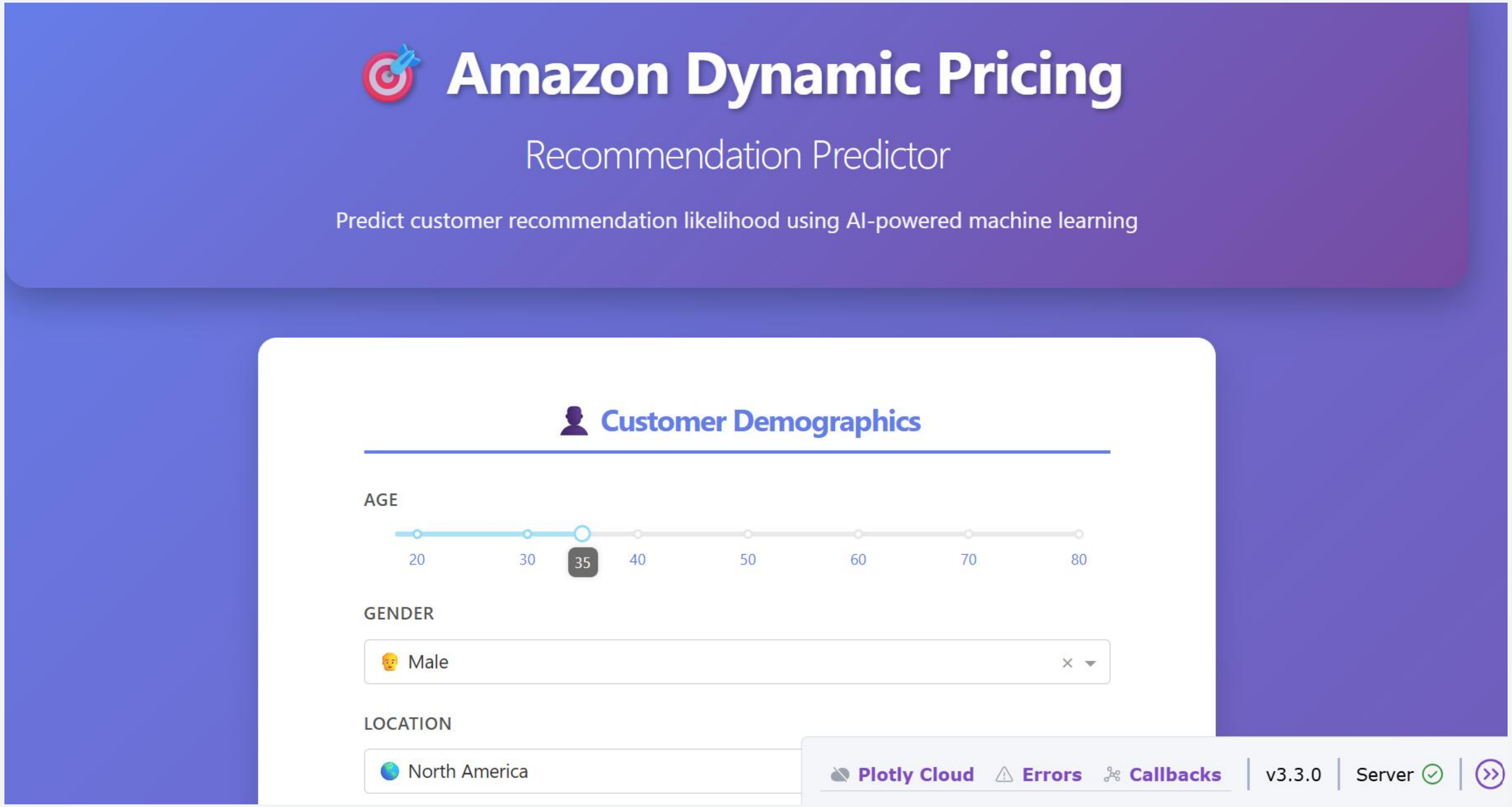
Minimal Adjacent Misclassification

Effectively distinguishes sentiment.

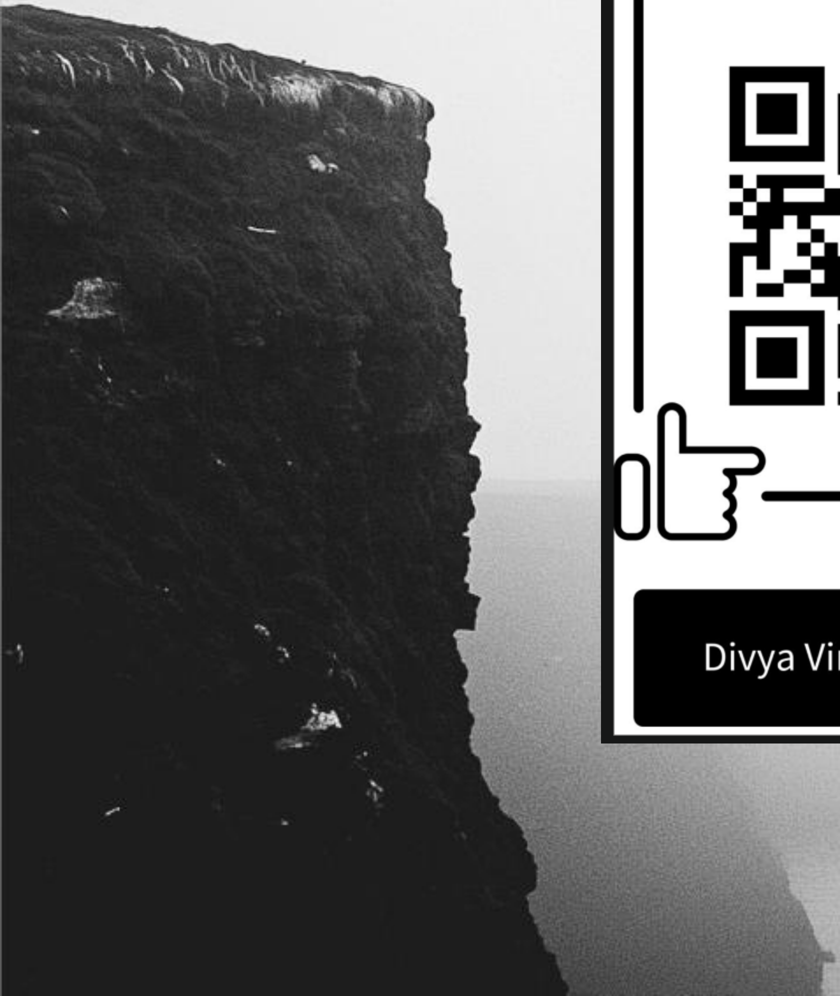


Interactive Prediction Dashboard

- A user-friendly web interface built with Dash for real-time prediction, providing instant probability scores and visual feedback.
- Real-time prediction
- Interactive input sliders
- Visual probability chart



THANK YOU



Divya Vinodbhai Koriya

Pooja Shreni Addisherla