Predictive mdoeling

📊 Predictive Modeling Summary – Razor Purchase Intent Study

✅ 1. Prepare Modeling Dataset

Objective: Select key variables for predicting razor purchase intent.

- **Target Variable**:
  `Q1. Which statement best describes how likely you would be to buy this razor in the future?`

- **Selected Features** (15 variables):
  - Demographics: `Gender`, `Age`, `NCCS`
  - Behavior: `Shaving product purchase`, `Decision maker`, `Brand used`
  - Emotional: `Feeling after good shave`, `Reaction to cuts`
  - Pack Evaluation: `Pack Design Influence`, `Pack Uniqueness`, `Believability`
  - Innovation & Switching: `Heard of new brands`, `Bluetooth razor interest`, `Switching tendency`, `Razor choice factors`

- **Tool Used**: `Python (pandas)`

- **Output**: Cleaned and structured dataset `df_model`

⚙️ 2. Feature Selection & Encoding

Objective: Convert variables into machine-readable form.

- Used `LabelEncoder` from `sklearn` for all object (categorical) columns
- Target and all features successfully encoded
- Dropped missing values in target column

Output: Final encoded features `X` and target `y`

🧪 3. Split Data: Train/Test

Objective: Validate model accuracy on unseen data.

- Split done using `train_test_split` from `sklearn.model_selection`
- **80% training / 20% testing**, stratified on target variable
- Ensured balanced representation across response classes

Output:

- `X_train`, `X_test`, `y_train`, `y_test`

🌳 4. Train Predictive Model

Objective: Predict future razor purchase intent.

- Model Used: `RandomForestClassifier`
- Parameters: `n_estimators=100`, `random_state=42`
- Trained on training set (`X_train`, `y_train`)

Tool: `sklearn.ensemble`
Output: `rf` – Trained model

📏 5. Evaluate Model Performance

Objective: Measure model accuracy and reliability.

Accuracy Score: ✅ **26.7%**

Classification Report:

Class	Precision	Recall	F1-Score	Support
1	0.00	0.00	0.00	9
2	0.25	0.44	0.32	16
3	0.36	0.47	0.41	19
4	0.00	0.00	0.00	7
5	0.00	0.00	0.00	9

Confusion Matrix:

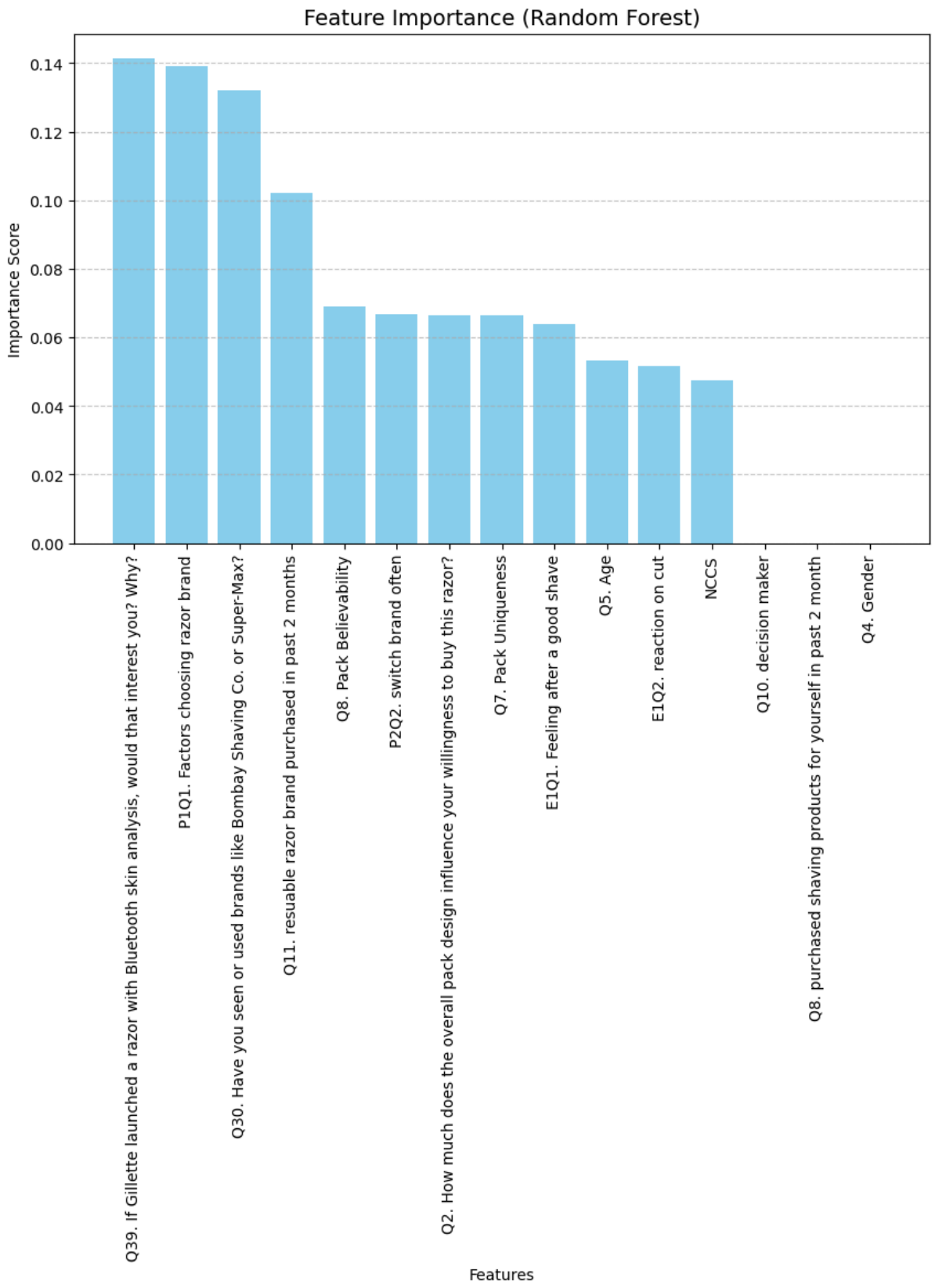
css
CopyEdit

```
[[0 3 6 0 0]
 [3 7 5 1 0]
 [1 7 9 1 1]
 [0 5 2 0 0]
 [0 6 3 0 0]]
```

📊 6. Plot Feature Importance

Objective: Identify top drivers of razor purchase intent.

- Extracted feature importances from Random Forest model
- Visualized using `matplotlib.pyplot`
- **Top Influential Features**:



Output: Visual bar chart of feature importance

✅ Tools Summary:

Step	Tool
Data Cleaning & Feature Selection	`pandas`
Encoding	`LabelEncoder (sklearn)`
Train/Test Split	`train_test_split (sklearn)`
Modeling	`RandomForestClassifier (sklearn)`
Evaluation	`accuracy_score`, `classification_report`, `confusion_matrix`
Visualization	`matplotlib.pyplot`