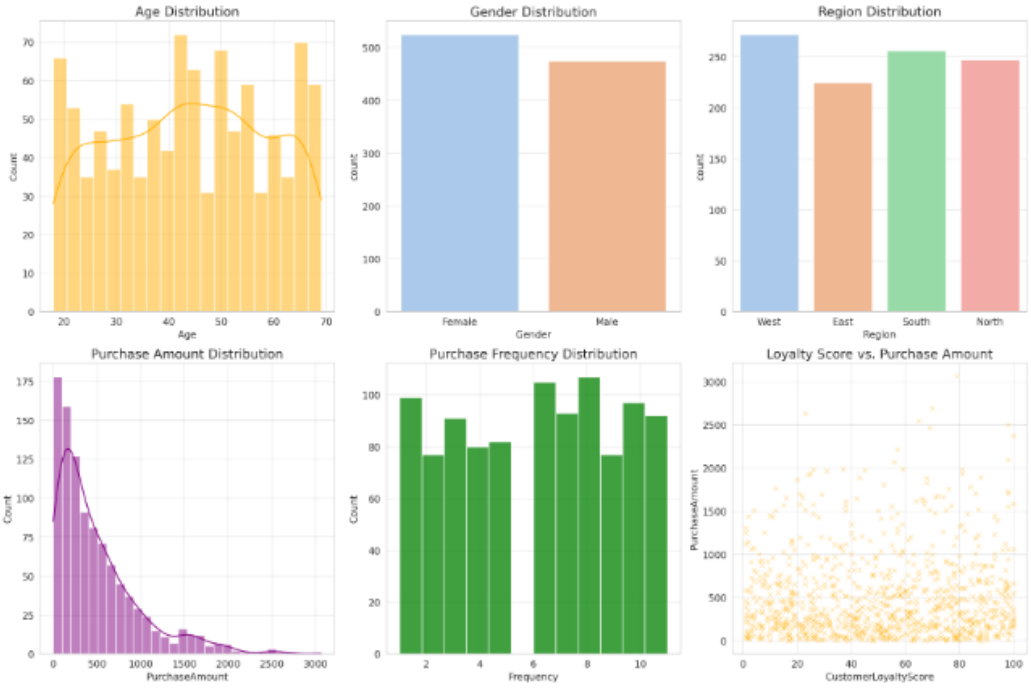
Enhancing Customer Retention for an E-Commerce Business

Analyze customer behavior data to identify key factors affecting retention and suggest actionable insights to improve customer loyalty and revenue.



Exploratory Data Analysis (EDA) Insights

1. **Age Distribution:**
    - Ages are evenly spread between 18 and 70, with a slight concentration around 40–50.
2. **Gender Distribution:**
    - The dataset has a nearly balanced gender distribution: ~48% Male and ~52% Female.
3. **Region Distribution:**
    - Customers are well-distributed across all regions, with the "West" slightly leading.
4. **Purchase Amount:**
    - The spending shows a skewed distribution with most customers spending below $1,000, but a few high spenders.
5. **Purchase Frequency:**
    - Monthly purchase frequency ranges from 1 to 11, with most customers buying ~6 times/month.
6. **Loyalty vs. Spending:**
    - Higher loyalty scores appear moderately correlated with higher purchase amounts, indicating a potential pattern.

Enhancing Customer Retention for an E-Commerce Business

Analyze customer behavior data to identify key factors affecting retention and suggest actionable insights to improve customer loyalty and revenue.

Customer Segmentation Analysis

The K-Means clustering segmented customers into four groups based on their purchase behavior and loyalty. Below are the cluster centroids for interpretation:

Cluster	Average Purchase Amount	Average Frequency	Average Loyalty Score
0	$368.66	7.86 purchases	79.39
1	$344.74	8.36 purchases	24.05
2	$1,511.93	5.53 purchases	52.84
3	$348.87	2.59 purchases	45.55

Insights

1. **Cluster 0:** Loyal customers with moderate spending and frequent purchases. Likely the most valuable group to retain.
2. **Cluster 1:** Low loyalty despite high purchase frequency. This group may need engagement strategies to improve loyalty.
3. **Cluster 2:** High spenders with moderate loyalty and lower purchase frequency. Targeted retention offers could increase their engagement.
4. **Cluster 3:** Sporadic buyers with low loyalty. This segment is the least profitable but may improve with personalized incentives.

Enhancing Customer Retention for an E-Commerce Business

Analyze customer behavior data to identify key factors affecting retention and suggest actionable insights to improve customer loyalty and revenue.

Churn Prediction Results

The logistic regression model was trained to predict whether customers are at risk of churning (loyalty score < 30). Here are the evaluation results:

Confusion Matrix

	Predicted Not Churn	Predicted Churn
Actual Not Churn	218	0
Actual Churn	82	0

Metrics

- **Accuracy:** 72.67%
- **Precision (Churn):** Undefined (no churn predictions were made)
- **Recall (Churn):** 0.00%
- **F1-Score (Churn):** 0.00%

Insights

1. The model predicts all customers as "Not Churn." This behavior suggests:
    - A significant class imbalance: ~82 churners vs. 218 non-churners in the test set.
    - The model struggles to learn patterns due to an imbalanced target variable.
2. Next Steps to Improve Model Performance:
    - Address the imbalance using techniques like oversampling (e.g., SMOTE) or class-weight adjustments.
    - Experiment with more advanced algorithms such as random forests or gradient boosting.

Link for the dashboard
http://127.0.0.1:8050/