Customer churn issue from Loblaw - Loblaws Digital

Presented By

Shireesha Thyaranahalli Narayana
Shih-Chieh Ku

Introduction

- In order to develop tactics that would keep its customers loyal, Loblaws Digital is interested in learning which of its clients are likely to attrit. As a data scientist, Loblaws Digital has recruited us to assist in forecasting client attrition and devising retention-boosting tactics.

Data description

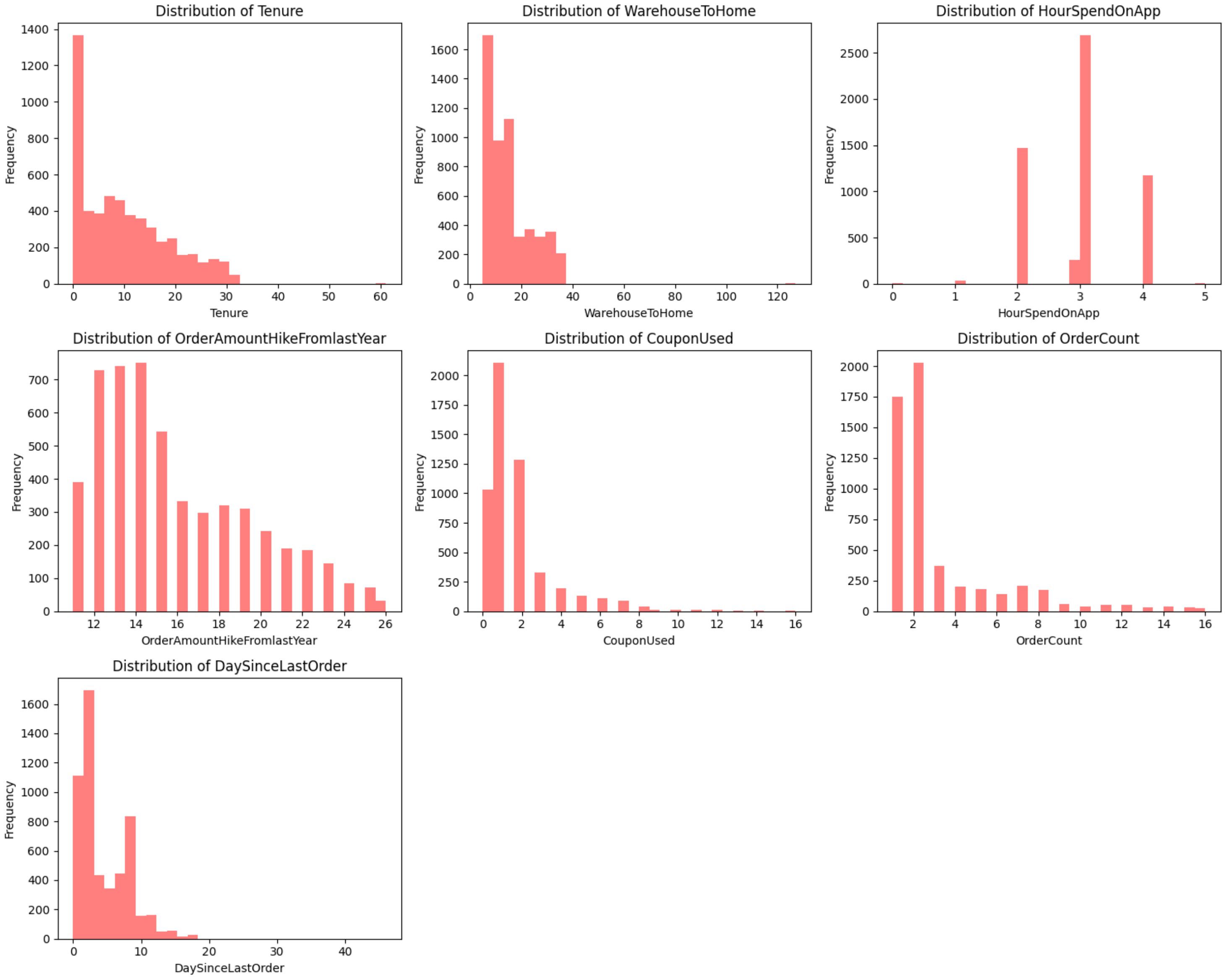
Variable	Variable
CustomerID	Unique customer ID
Churn	Churn Flag
Tenure	Tenure of customer in organization
PreferredLoginDevice	Preferred login device of customer
CityTier	City tier
WarehouseToHome	Distance in between warehouse to home of customer
PreferredPaymentMode	Preferred payment method of customer
Gender	Gender of customer
HourSpendOnApp	Number of hours spend on mobile application or website
NumberOfDeviceRegistered	Total number of deceives is registered on particular customer
PreferedOrderCat	Preferred order category of customer in last month

Data description (continue)

Variable	Variable
SatisfactionScore	Satisfactory score of customer on service
MaritalStatus	Marital status of customer
NumberOfAddress	Total number of added added on particular customer
Complain	Any complaint has been raised in last month
OrderAmountHikeFromlastYear	Percentage increases in order from last year
CouponUsed	Total number of coupon has been used in last month
OrderCount	Total number of orders has been places in last month
DaySinceLastOrder	Day Since last order by customer
CashbackAmount	Average cashback in last month

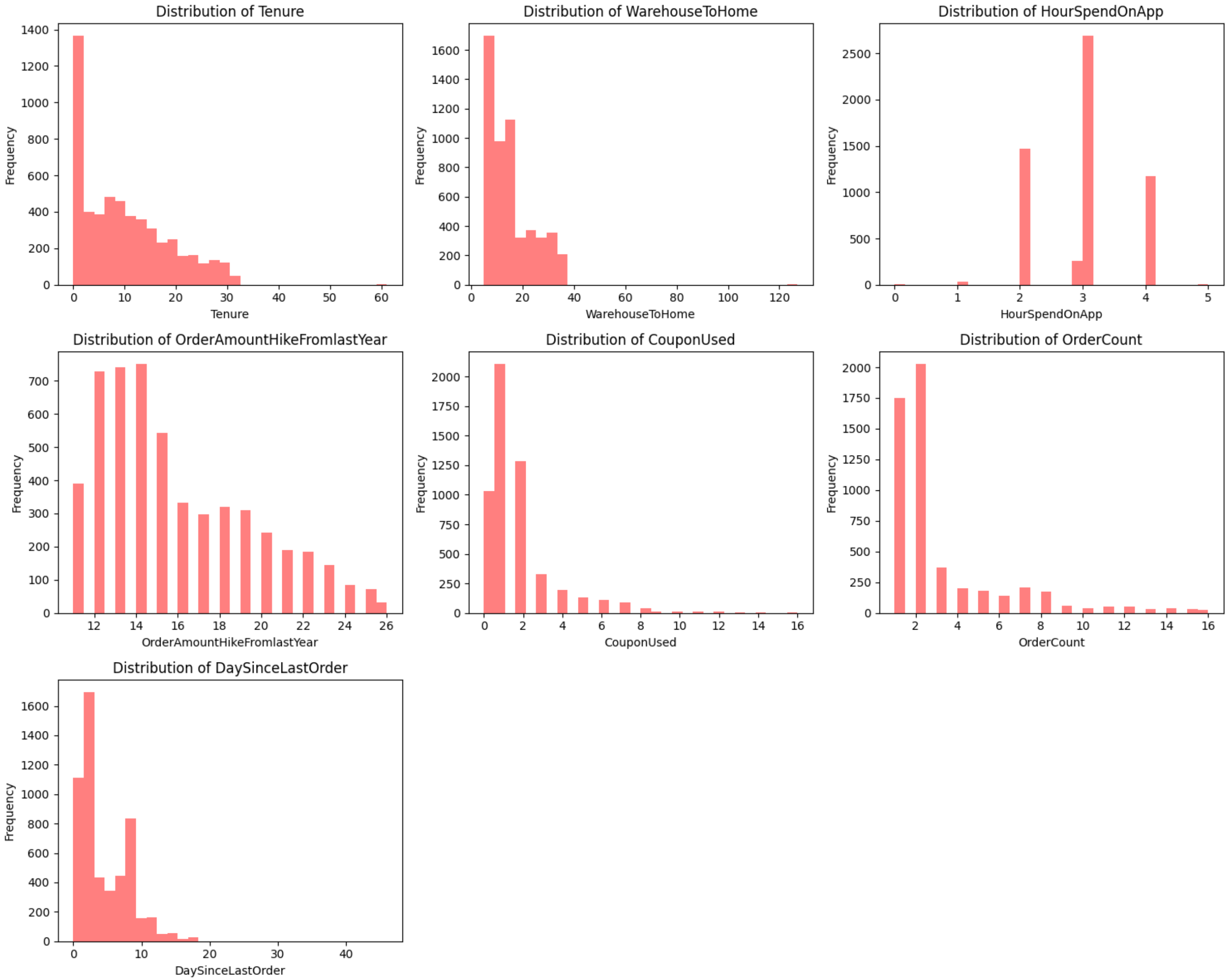
Exploring the data

Distribution



- **Distribution of Tenure**:
  - Most users have a tenure of fewer than 3 years.
- **Distribution of WarehouseToHome**:
  - There are very few users living more than 40 kilometers away.
- **Distribution of DaySinceLastOrder**:
  - Most users have placed an order within the last 10 days.
  - There is a sharp decline in the number of users as the days since the last order increases, with very few users having placed an order after 30 days.

Distribution (continue)



- **Distribution of HourSpendOnApp**:
  - The majority of users spend between 2 and 4 hours on the app.
  - Very few users spend less than 1 hour or more than 5 hours on the app.
- **Distribution of CouponUsed**:
  - Most users have used 0 to 2 coupons, very few users used more than 4 coupons.
- **Distribution of OrderCount**:
  - Most users have placed 2 to 4 orders.
  - The number of users decreases significantly for higher order counts, with very few users placing more than 8 orders.

Overall Thoughts:

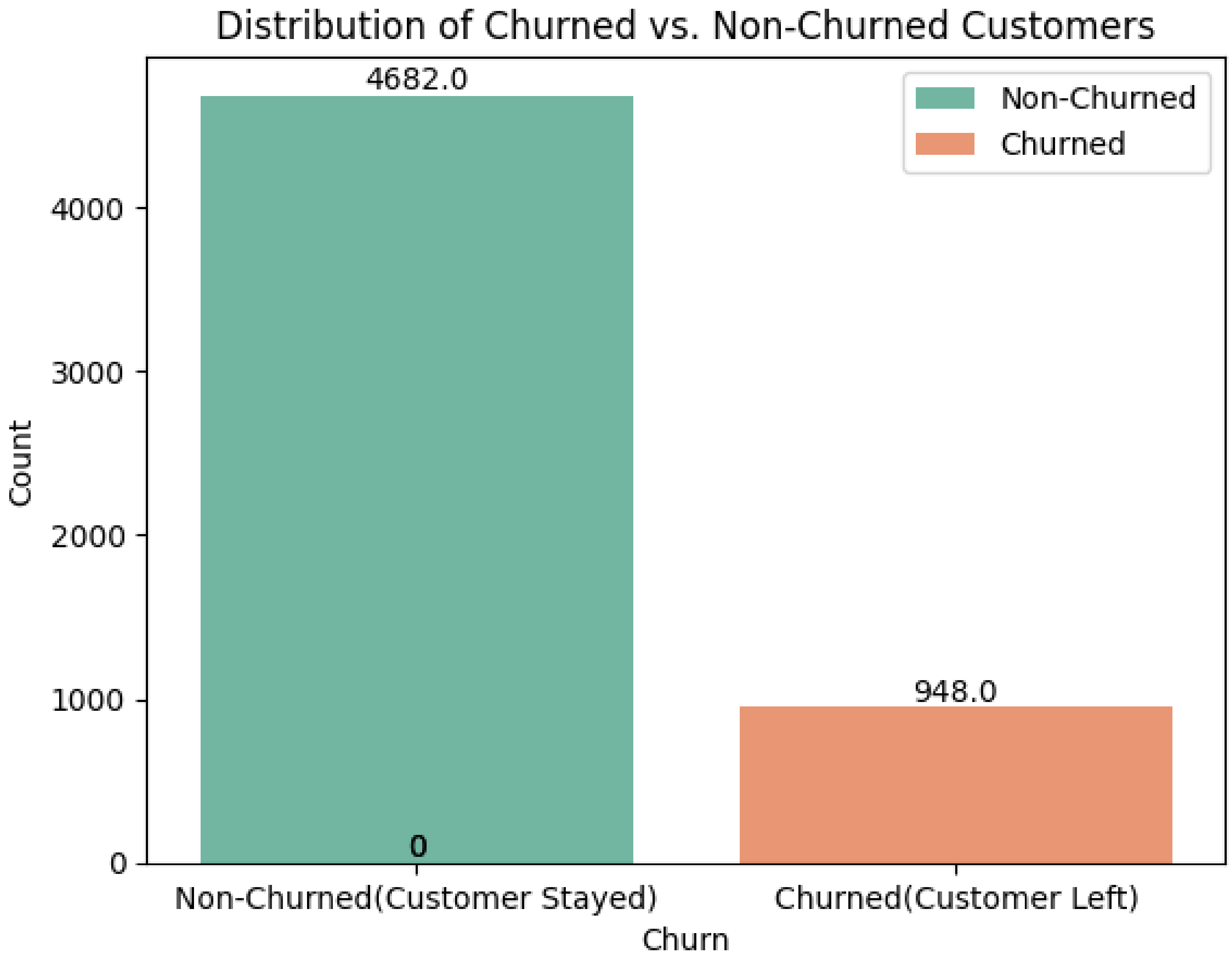
- Tenure:
  - The mean customer tenure (mean: 10.189899) is approximately 10 months.
  - The term ranges from 0 months at the minimum to 61 months at the highest.
- CityTier:
  - It's challenging to immediately evaluate this without further information regarding the city tier classification.
  - The city tier distribution appears to have some spread, as indicated by the standard deviation of 0.915389.
- THourSpendOnApp:
  - Users log on to the app or website for an average of 2.93 hours (mean: 2.931535).
  - The app/website usage time standard deviation of 0.721926 shows a considerable degree of variation.
- SatisfactionScore:
  - 3.06 (mean: 3.066785) is the average satisfaction score. Again, direct interpretation is hampered in the absence of knowledge about the scoring system.
  - The 1.380194 standard deviation indicates a respectable range of satisfaction ratings.

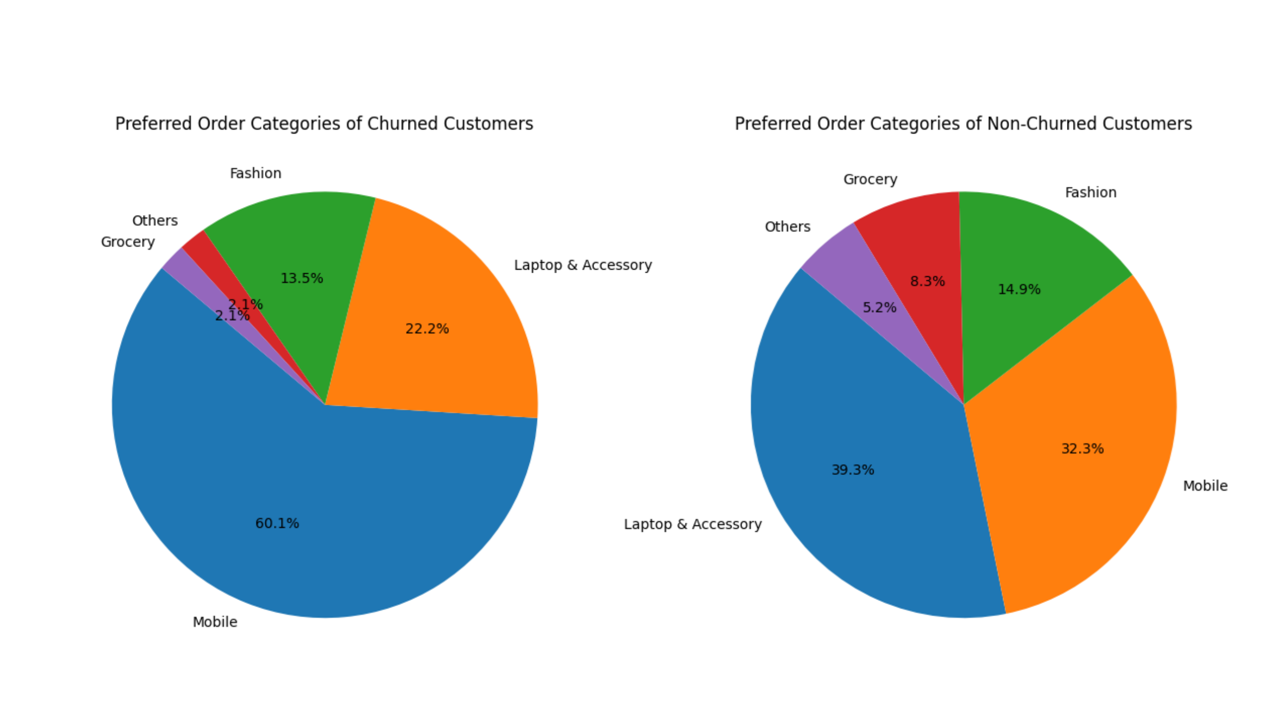
Overall Thoughts (continue):

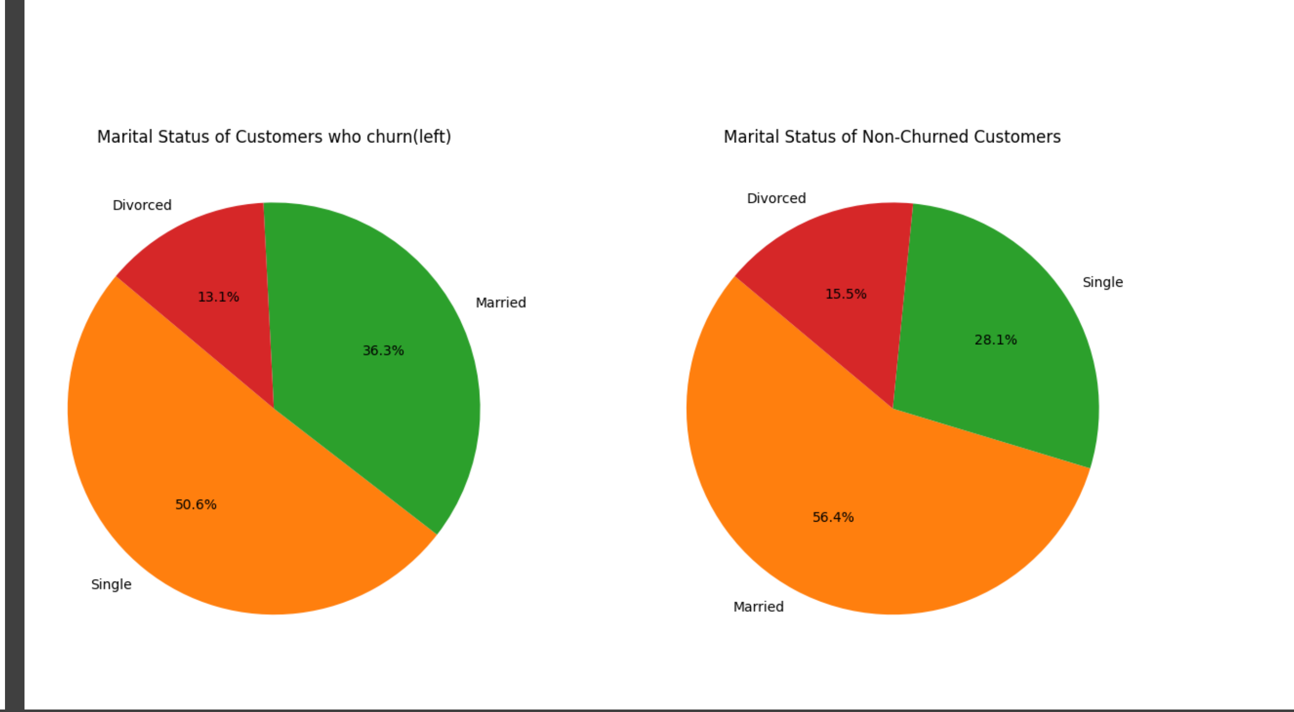
- Complain
  - In the past month, roughly 28.49% of consumers have filed a complaint (mean: 0.284902).
- OrderAmountHikeFromPreviousYear:
  - The average order amount (mean: 15.707922) has risen by roughly 15.71% from the year before.
  - There may be some variance in this growth rate, as indicated by the 3.675485 standard deviation.
- OrderCount:
  - Approximately three orders were placed in the previous month on average (mean: 3.008004).
- DaySinceLastOrder:
  - The mean (4.543491) number of days since the last order is approximately 4.54.
- CashbackAmount:
  - 177.22 units on average (mean: 177.223030) are the cashback that was obtained during the past month.

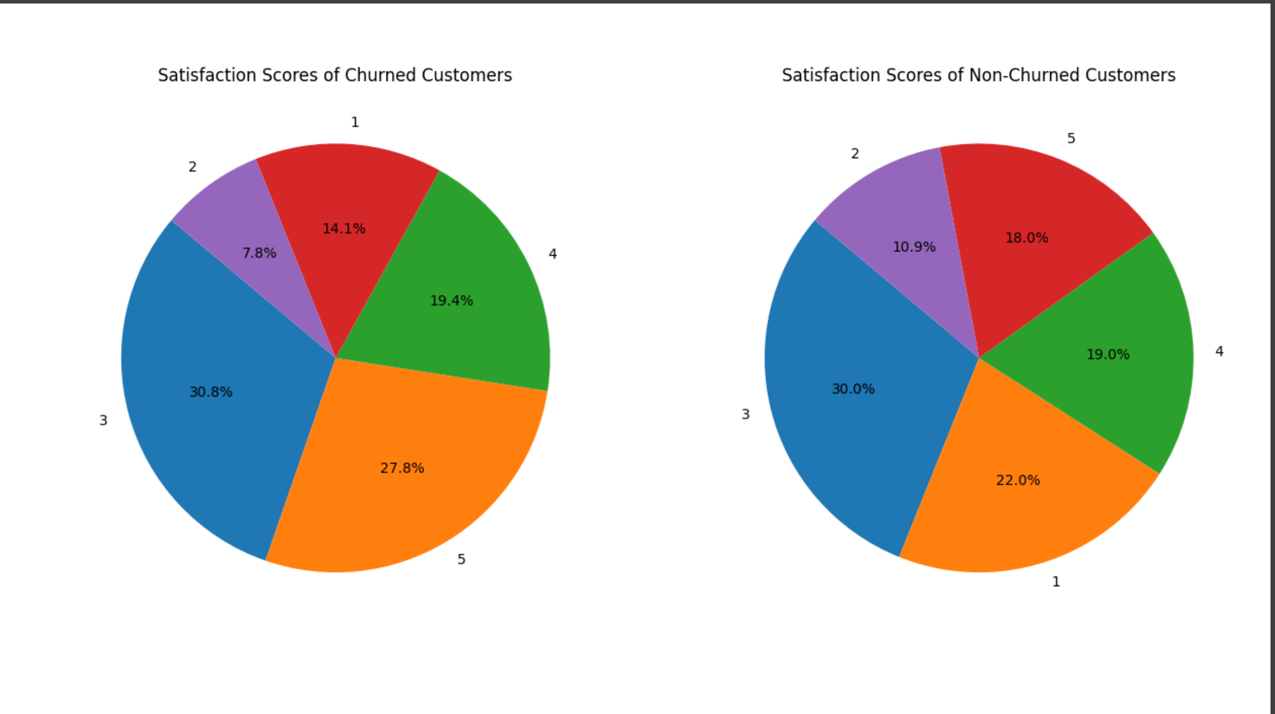
Non-Churned (Customer Stayed): 83.1%

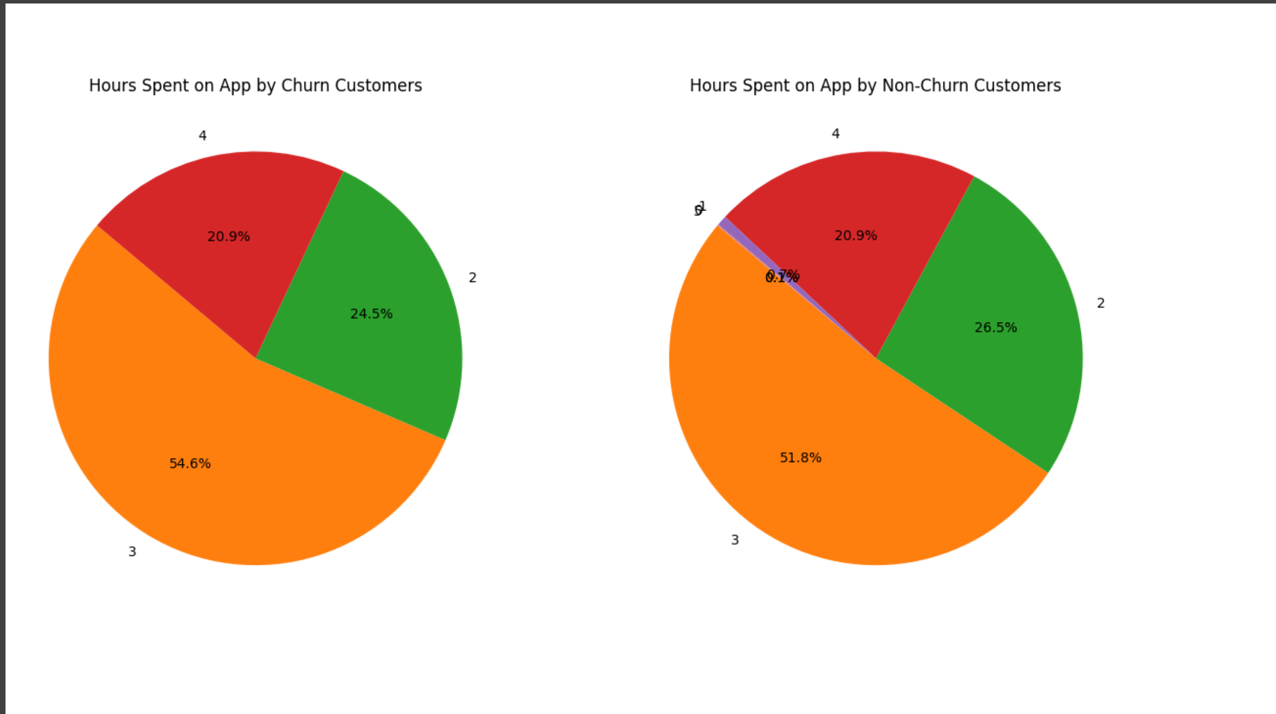
Churned (Customer Left): 16.9%




Preferred Order Categories of Churned Customers
Fashion
Others
Grocery
Laptop & Accessory
Mobile
13.5%
22.2%
2.1%
2.1%
60.1%
Preferred Order Categories of Non-Churned Customers
Grocery
Fashion
Others
8.3%
14.9%
5.2%
39.3%
32.3%
Mobile
Laptop & Accessory


Marital Status of Customers who churn(left)
Divorced
Married
13.1%
36.3%
50.6%
Single
Marital Status of Non-Churned Customers
Divorced
Single
15.5%
28.1%
56.4%
Married

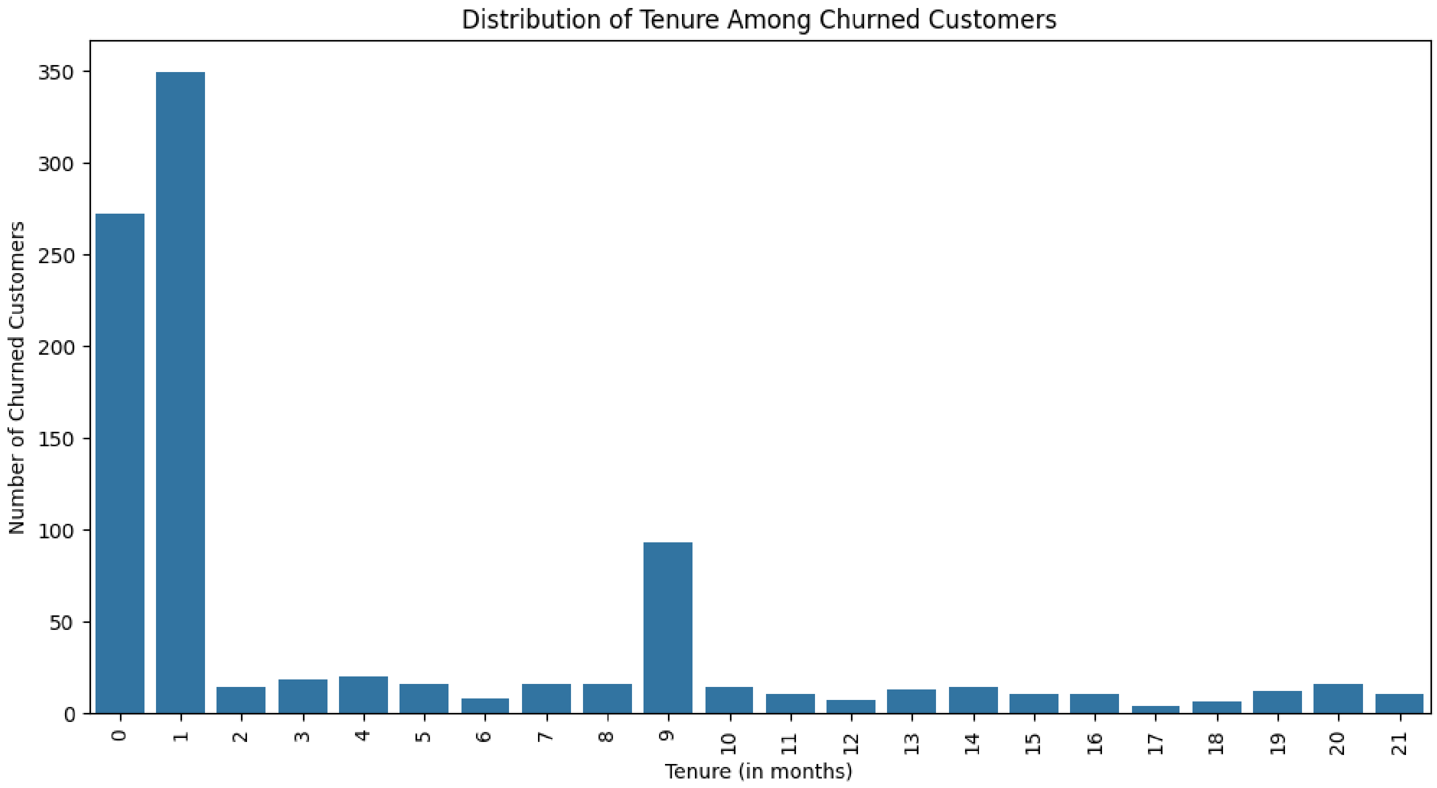

Satisfaction Scores of Churned Customers
1
2
4
14.1%
7.8%
19.4%
30.8%
27.8%
3
5
Satisfaction Scores of Non-Churned Customers
5
2
10.9%
18.0%
19.0%
4
30.0%
22.0%
3
1

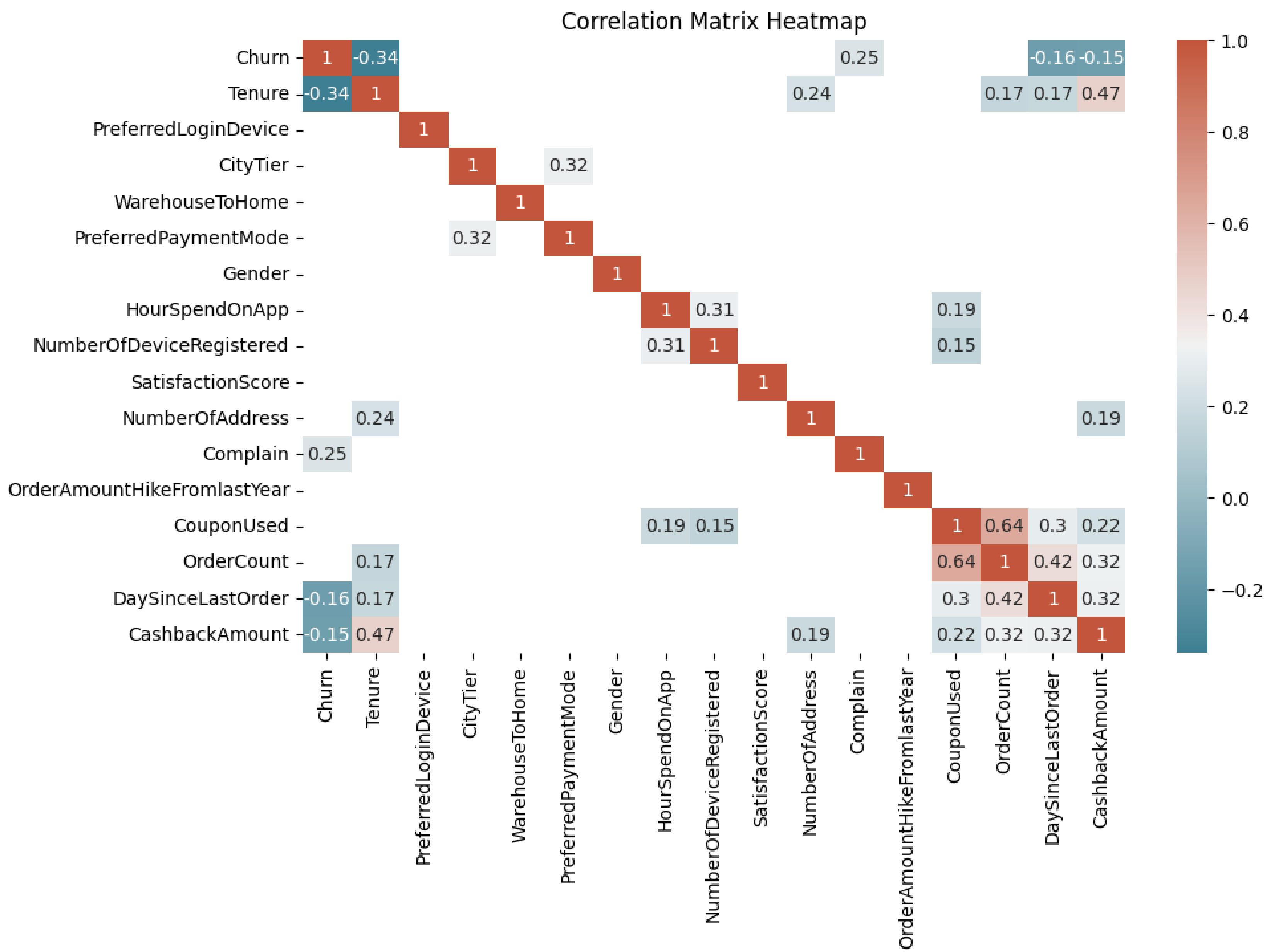

Hours Spent on App by Churn Customers
4
20.9%
2
24.5%
54.6%
3
Hours Spent on App by Non-Churn Customers
4
20.9%
0.1%
26.5%
2
51.8%
3

Churned customer with tenure

For most churned customers, their tenure period is usually only between 1 month and 2 months, it seems quite reasonable, we might be able to assume that they apply for a membership for the promotion or discount of the first order.

However, we can find out that almost 100 customers churned when their tenure reached nine months. This is more unusual, we can figure it out to find if there is special event.





Because there is no significant relationship between churn and others, so we set the threshold as absolute 0.15.

- **Churn**:
  - There is a moderate negative correlation between churn and tenure (-0.34), suggesting that customers with longer tenure are less likely to churn.
  - There is also a moderate negative correlation between churn and cashback amount (-0.15), indicating that customers who receive more cashback are slightly less likely to churn.

Longer-tenured customers are less likely to churn and tend to be more engaged with the app.

Model training

Model chosen

Because Random Forest can handle complex datasets, we decided to apply it in our project. It is the perfect fit for our diversified dataset because of its versatility in processing numerical and categorical (binary categorical) data. Our goal in using Random Forest for predictive modelling jobs is to obtain a high degree of accuracy and generalisation.

- Accuracy: 96%

It indeed trained an excellent model with significantly high accuracy.

+ Code + Markdown | Run All Restart Clear All Outputs | Variables Outline ···

```python
# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Initialize and train the RandomForest model
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Make predictions
y_pred_rf = rf_model.predict(X_test)

# Evaluate the RandomForest model
print("RandomForest Classifier Model Evaluation")
print("=======================================")
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred_rf))
print("\nClassification Report:")
print(classification_report(y_test, y_pred_rf))
print("\nAccuracy Score:")
print(accuracy_score(y_test, y_pred_rf))
```

[34]

```
RandomForest Classifier Model Evaluation
=======================================
Confusion Matrix:
[[1395   19]
 [  44  231]]

Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.99      0.98      1414
           1       0.92      0.84      0.88       275

    accuracy                           0.96      1689
   macro avg       0.95      0.91      0.93      1689
weighted avg       0.96      0.96      0.96      1689
```

Model chosen

We also implemented another model - Logistic Regression as a comparison with Random Forest.

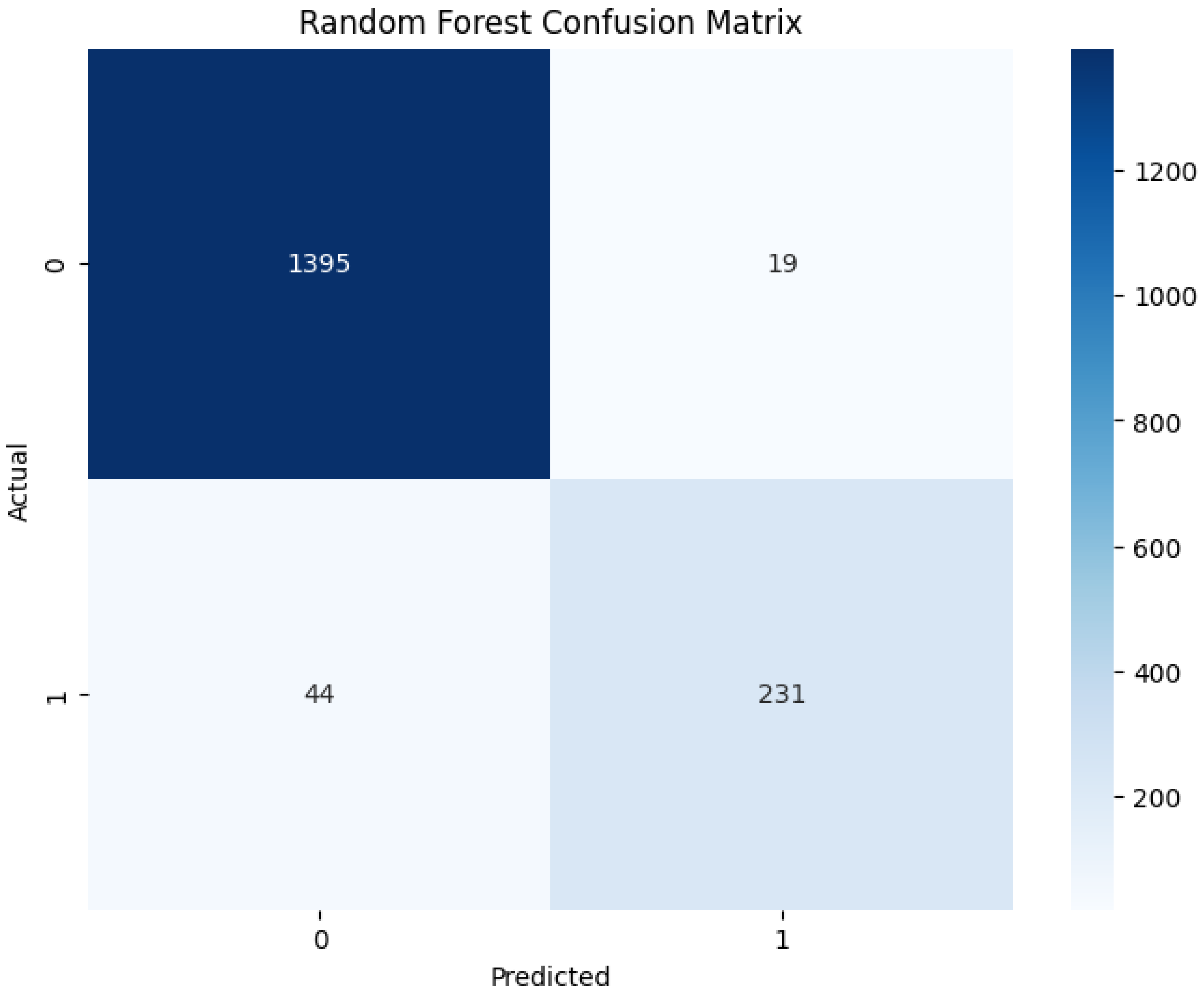
- Accuracy: 89%

The trained model of Logistic Regression also trained a good performance with 89% accuracy. Although, the precision, recall, f1-score are lower than Random Forest's, it can provide a good prediction on new data.

```python
# Initialize and train the Logistic Regression model
lr_model = LogisticRegression(max_iter=1000, random_state=42)
lr_model.fit(X_train, y_train)

# Make predictions
y_pred_lr = lr_model.predict(X_test)

# Evaluate the Logistic Regression model
print("\nLogistic Regression Model Evaluation")
print("====================================")
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred_lr))
print("\nClassification Report:")
print(classification_report(y_test, y_pred_lr))
print("\nAccuracy Score:")
print(accuracy_score(y_test, y_pred_lr))
```

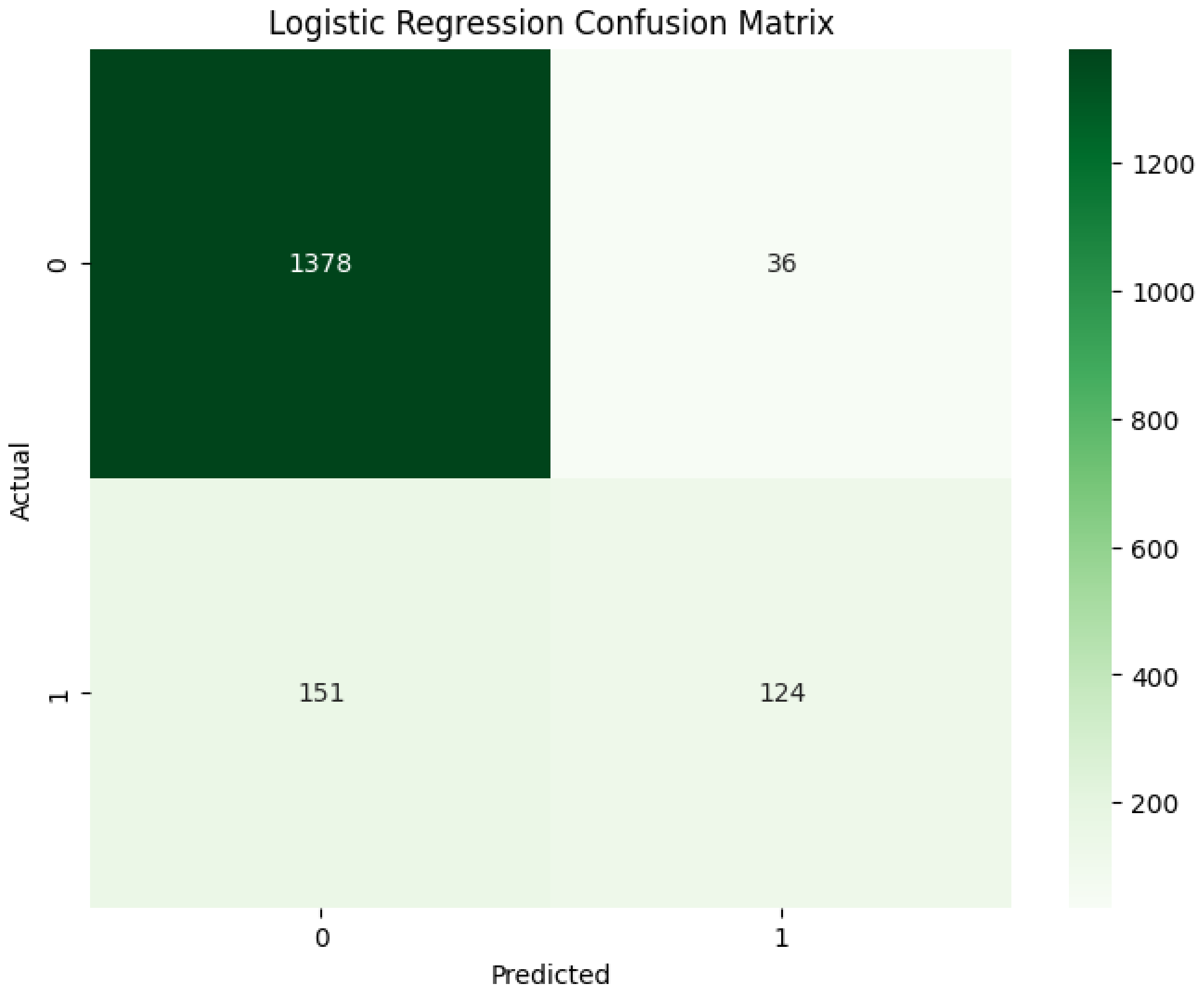
[37]

```
Logistic Regression Model Evaluation
====================================
Confusion Matrix:
[[1378   36]
 [ 151  124]]

Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.97      0.94      1414
           1       0.78      0.45      0.57       275

    accuracy                           0.89      1689
   macro avg       0.84      0.71      0.75      1689
weighted avg       0.88      0.89      0.88      1689


Accuracy Score:
0.8892835997631735
```



Confusion matrix - Random Forest

- Precision is high, indicating that when the model predicts a customer will churn, it is likely correct.
- Recall is slightly lower, indicating that some churners are being missed by the model.
- There is a small number of false positives and false negatives, suggesting the model is reliable but could be improved to catch more actual churners without increasing false positives significantly.



Confusion matrix - Logistic Regression

- Precision is reasonably good at 77.5%, suggesting that when the model predicts churn, it is correct a significant majority of the time.
- Recall is quite low at 45.1%, indicating the model misses more than half of the actual churners. This suggests the model is better at correctly identifying non-churners than churners.
- The F1 Score of 0.57 indicates a moderate balance between precision and recall, but there is room for improvement, particularly in identifying actual churners.

Business Decision and Marketing Strategies

- Coupons and Promotional Strategies: Focus on promoting coupon usage by offering attractive discounts and promotions related to specific order categories.
- Re-engagement campaigns: Leverage targeted re-engagement campaigns or timely follow-ups after purchase to maintain customer interest and prevent churn.
- Mobile Preference: More customers prefer using mobile phones than computers. Invest in enhancing the mobile app experience to encourage customers to spend more time in the app.
- Targeted retention plans: Develop retention plans for customers with longer tenures. Offer loyalty rewards, discounts or exclusive offers to retain customers.

Improvements after Implementing Business & Marketing Strategies

- By releasing more customized and attractive coupons, we can maximize customer stickiness with Loblaw.
- By following up on new customers' responses and satisfaction with their first order, we can encourage them to return or modify our new customer policies to maintain their interest.
- By investing in enhancing the mobile app experience, we can satisfy more customers and potentially attract new ones with inrresistible promote and app feature.
- By developing retention plans for long-term customers, we can significantly increase the short-term number of orders and retain customers, preventing them from churning.

Thank you