Telecom Customer Churn Analysis

By Sanjay Pal

Date: 02-08-2024

Introduction to Customer Churn Analysis in Telecom

Business Problem Overview

- **Competitive Telecom Market:**
  - Customers frequently switch service providers, leading to high churn rates.
  - Average annual churn rate: **15-25%**.
- **Importance of Customer Retention:**
  - Acquiring new customers costs **5-10 times** more than retaining existing ones.
  - Retaining profitable customers is critical for business success.
- **Objective:**
  - Predict customers at high risk of churn and identify key churn indicators.
  - Focus on retaining highly profitable customers to minimize revenue leakage.

Understanding and Defining Churn

- **Churn Models in Telecom:**
  - **Postpaid:** Customers notify operators when switching.
  - **Prepaid:** Customers can leave without notice, making churn prediction challenging.
- **Focus on Prepaid Customers:**
  - More common in India and Southeast Asia.
  - Critical to define churn accurately for prepaid models.
- **Churn Definition:**
  - **Usage-based Churn:** Customers with no usage (calls, internet) over a period are considered churned.

High-Value Customer Churn

- **Revenue Distribution:**
  - **80%** of revenue from the top **20%** of customers.
  - Focus on reducing churn among high-value customers to protect significant revenue streams.
- **Project Objective:**
  - Define high-value customers using specific metrics.
  - Build predictive models targeting churn among high-value customers.

Data preparation

Filtering High-Value Customers

- **Objective:** Focus churn prediction on high-value customers for more impactful insights.
- **Definition of High-Value Customers:**
  - Customers who have recharged with an amount ≥ **X**.
  - **X** is the 70th percentile of the average recharge amount during the Good Phase (Months 6 & 7).
- **Outcome:** After filtering, approximately **30,000** high-value customer records are retained for analysis.

Tagging Churners:

- **Criteria for Churn (Churn = 1):**
  - No calls (incoming or outgoing) AND
  - No mobile internet usage in the Churn Phase (Month 9).
- **Attributes Used for Churn Tagging:**
  - total_ic_mou_9 (Total Incoming Calls - Month 9)
  - total_og_mou_9 (Total Outgoing Calls - Month 9)
  - vol_2g_mb_9 (2G Data Volume - Month 9)
  - vol_3g_mb_9 (3G Data Volume - Month 9)

Removing Churn Phase Attributes

•All attributes related to Month 9 are removed to prevent data leakage and focus on predictors from previous months.
•Attributes removed include any with the suffix _9, etc.

```
columns_9 = [col for col in df.columns.to_list() if '_9' in col]
print(columns_9)
```

```
df = df.drop(columns_9, axis=1)
```

Feature Engineering

Adding new features

- diff_arpu ---> Difference of Average Revenue Per User during the good phase (6 and 7 month) a

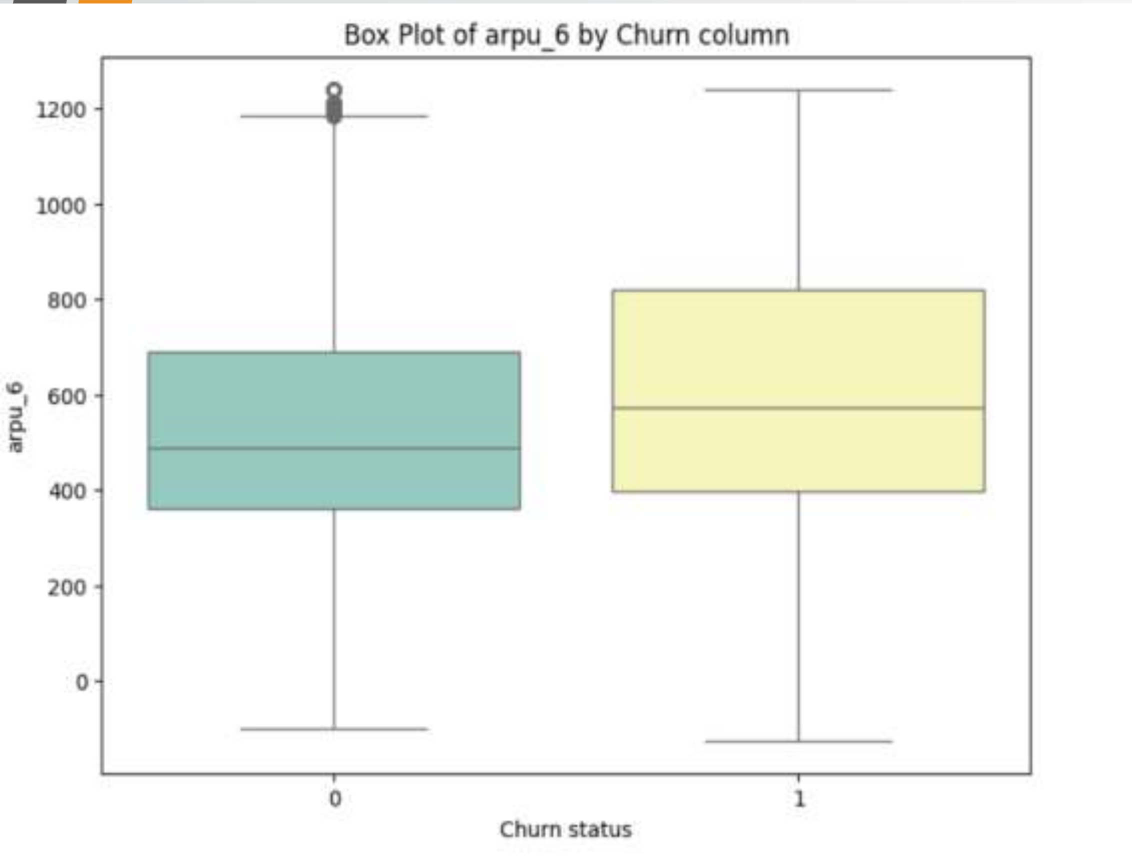
```python
# Taking the average for good phase ARPU
df1['avg_arpu_good'] = (df1['arpu_7'] + df1['arpu_6'])/2
# Difference of good and action phase ARPU
df1['diff_arpu'] = df1['arpu_8'] - df1['avg_arpu_good']
```

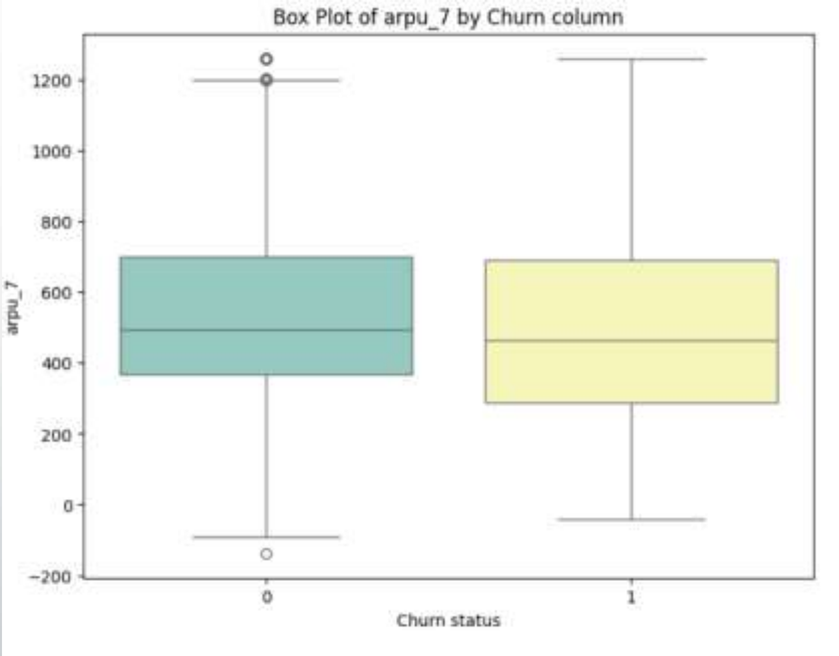
- dif action ph

```python
# Taking the average for good phase Recharge Amount
df1['avg_rech_amt_action'] = (df1['total_rech_amt_7'] + df1['total_rech_amt_
# Difference of action phase recharge amount and good phase recharge amount
df1['diff_rech_amt'] = df1['total_rech_amt_8'] - df1['avg_rech_amt_action']
```

EDA (Exploratory Data Analysis)

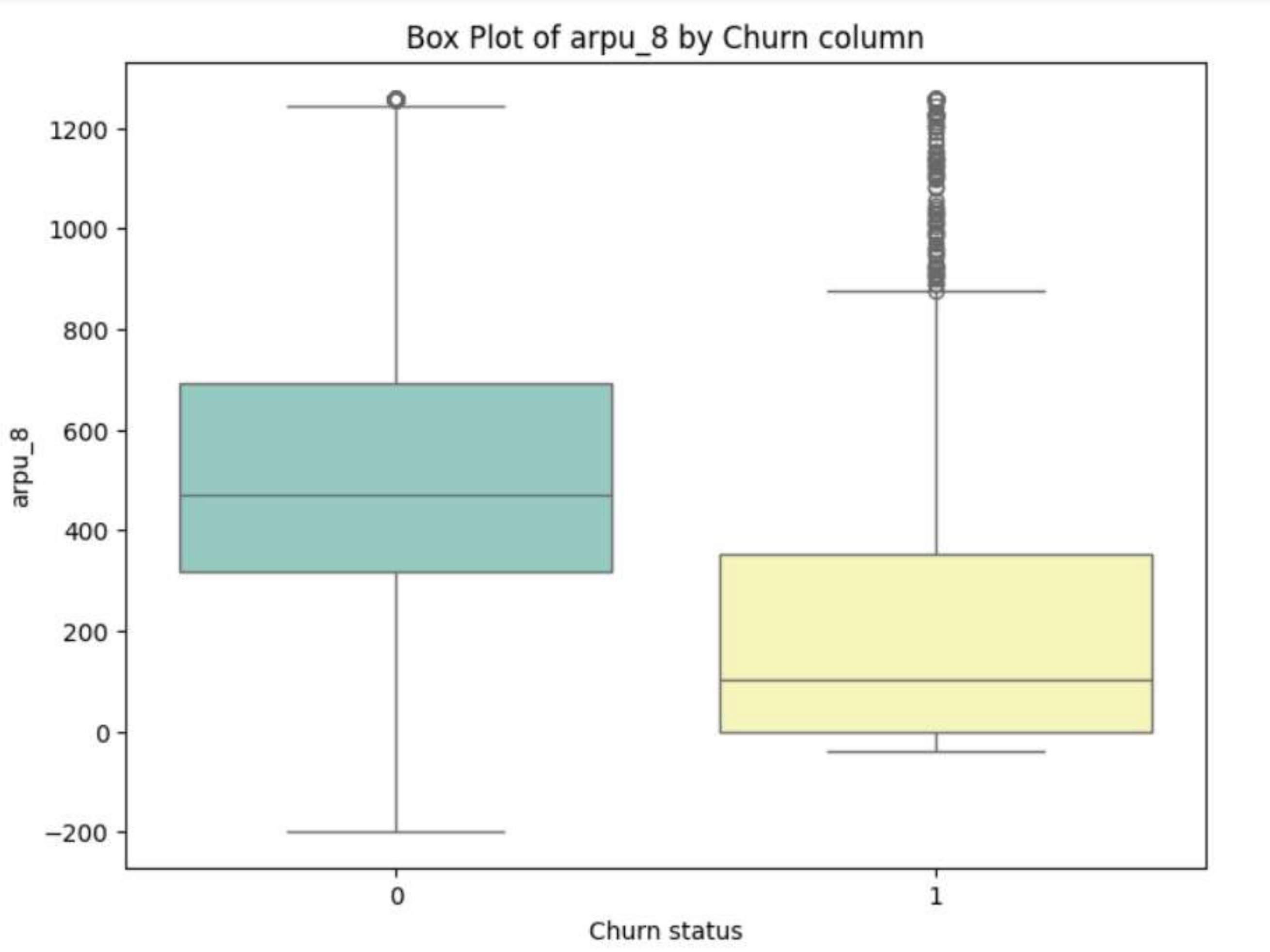
Box plot of ARPU(average revenue per user) with respect to Churn status



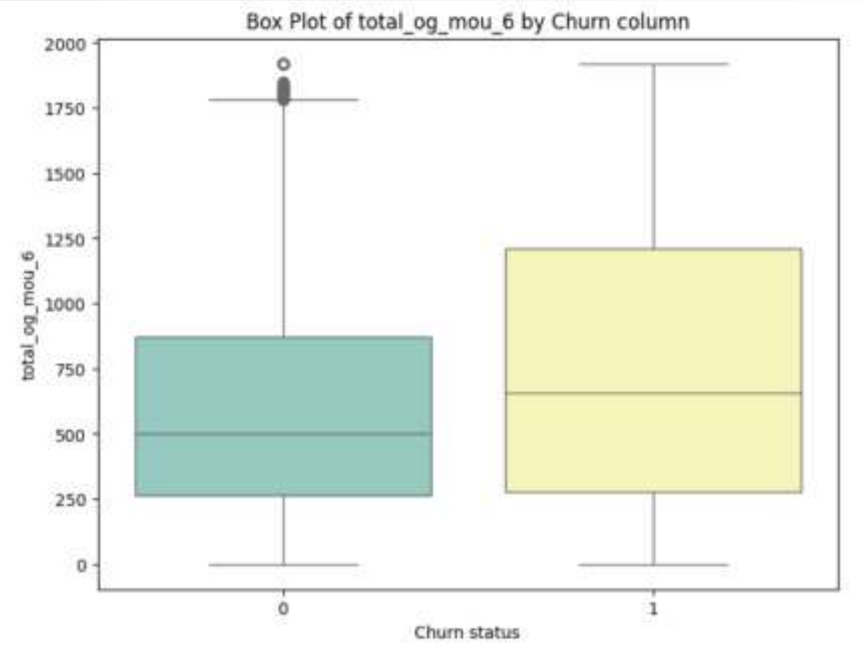


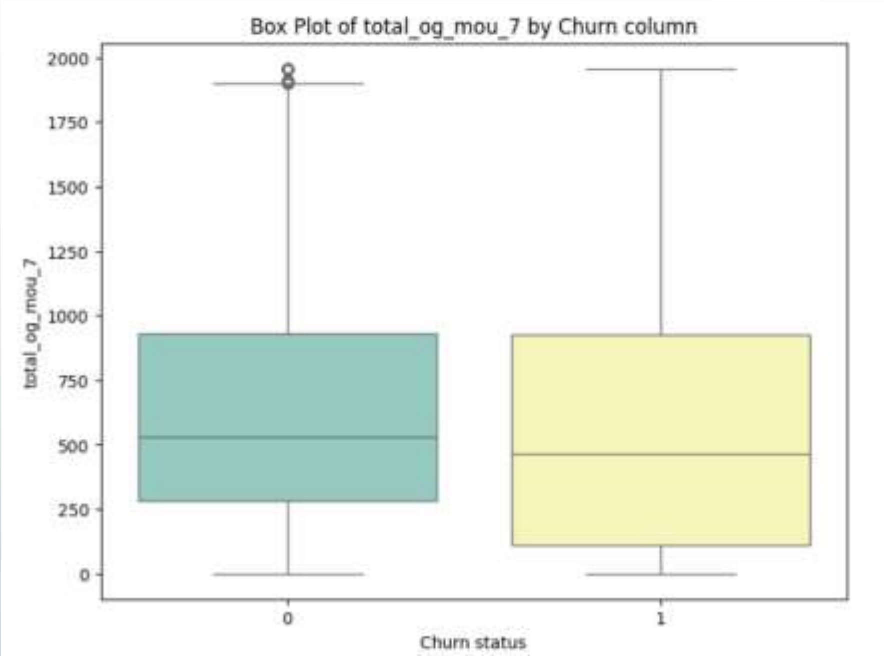
Observation

- As we can see, churn customers exhibit similar ARPU in good phase(months 6 and 7), but show significantly lower ARPU in action phase(month 8) compared to non-churn customers, indicating a decline in engagement.



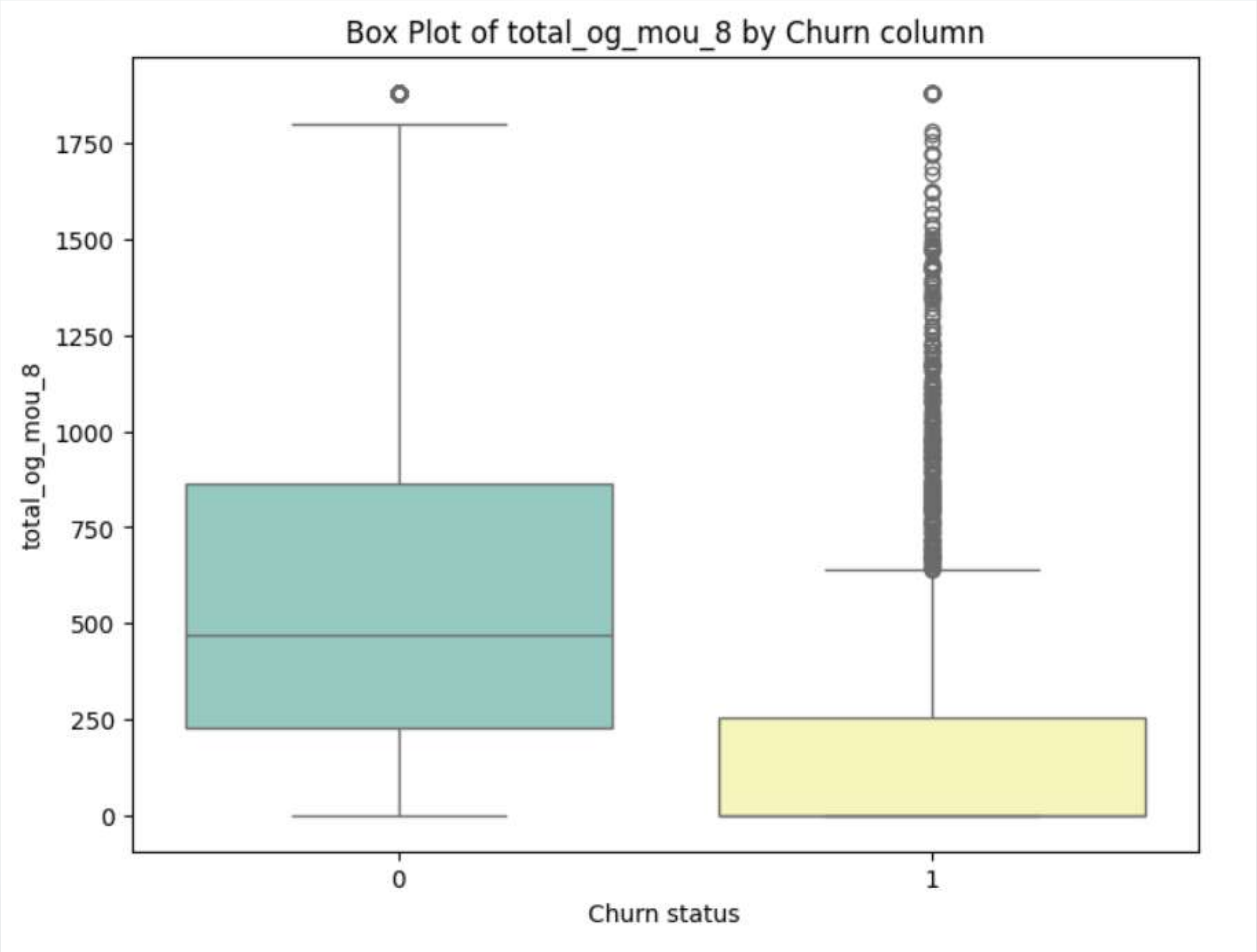
Box plot of total outgoing MOU(minutes of usages) with respect to churn status



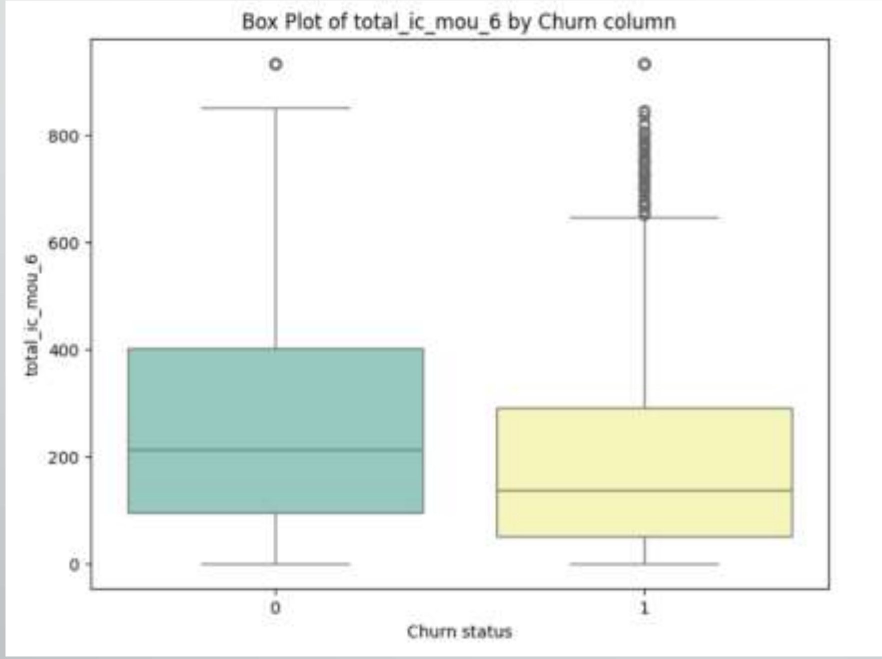


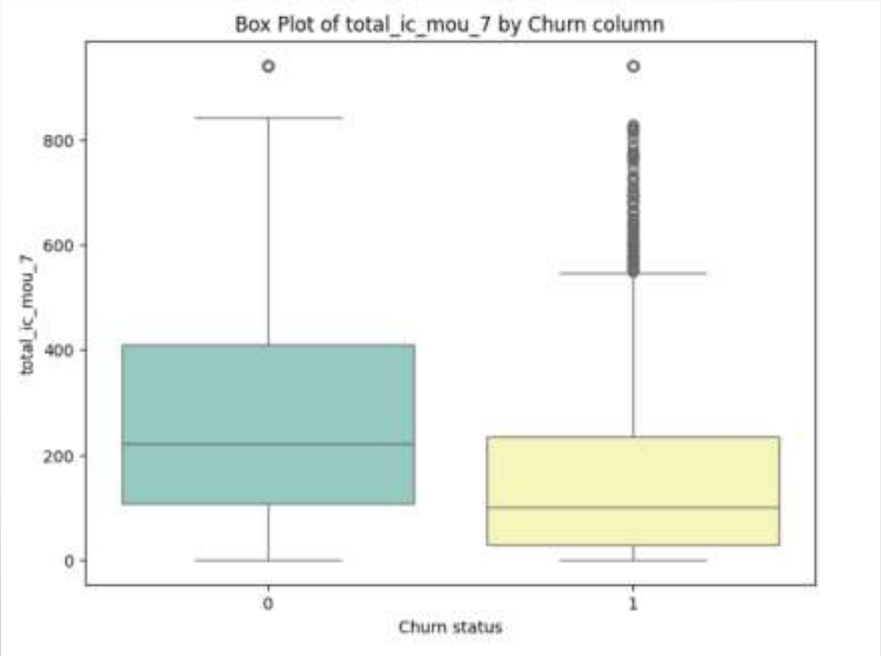
Observation

- As we can see the same pattern, churn customers exhibit similar total outgoing MOU in good phase(months 6 and 7), but show significantly lower total outgoing MOU in action phase(month 8) compared to non-churn customers, indicating a decline in engagement.



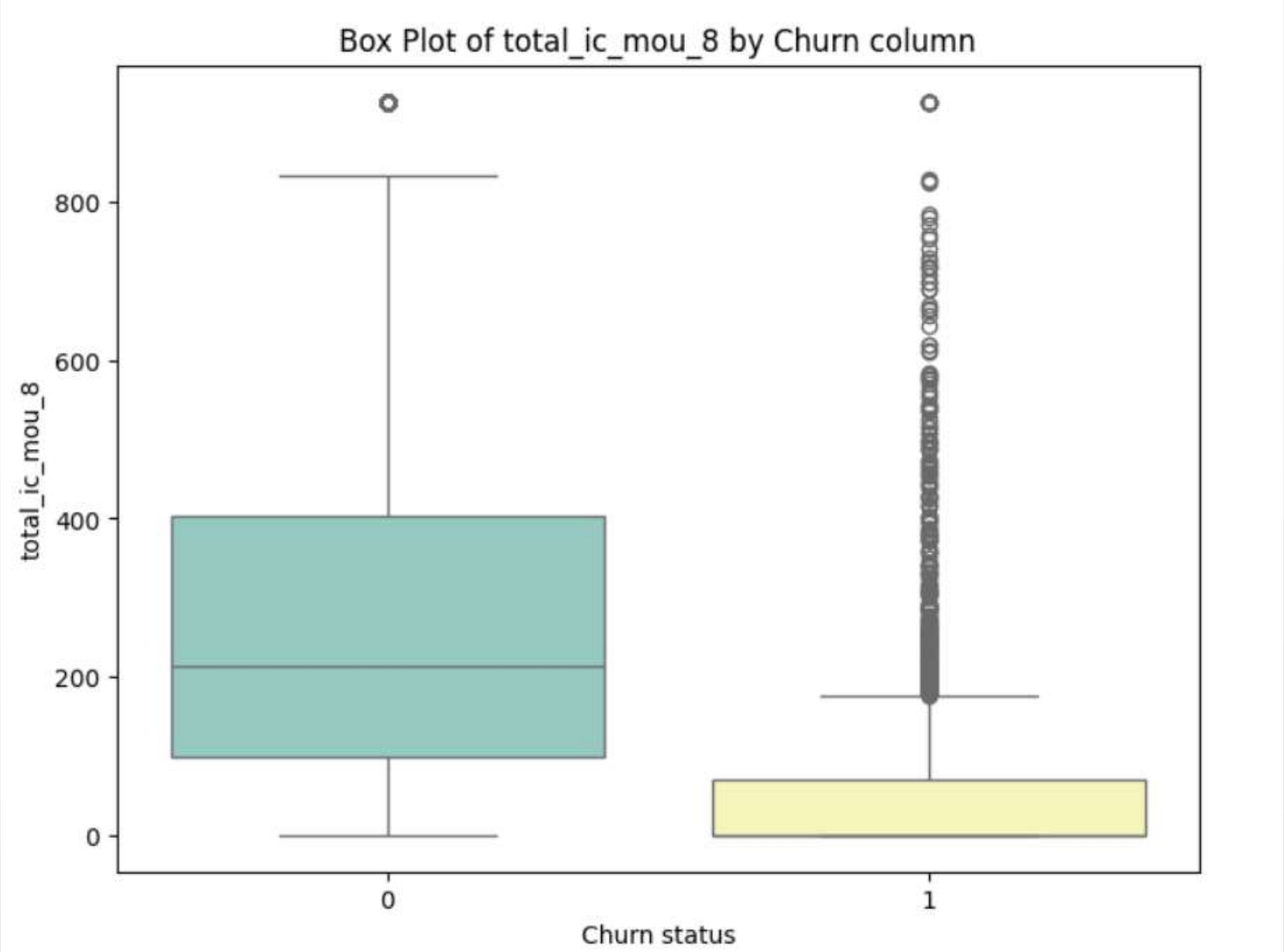
Box plot of total incoming MOU(minutes of usages) with respect to churn status





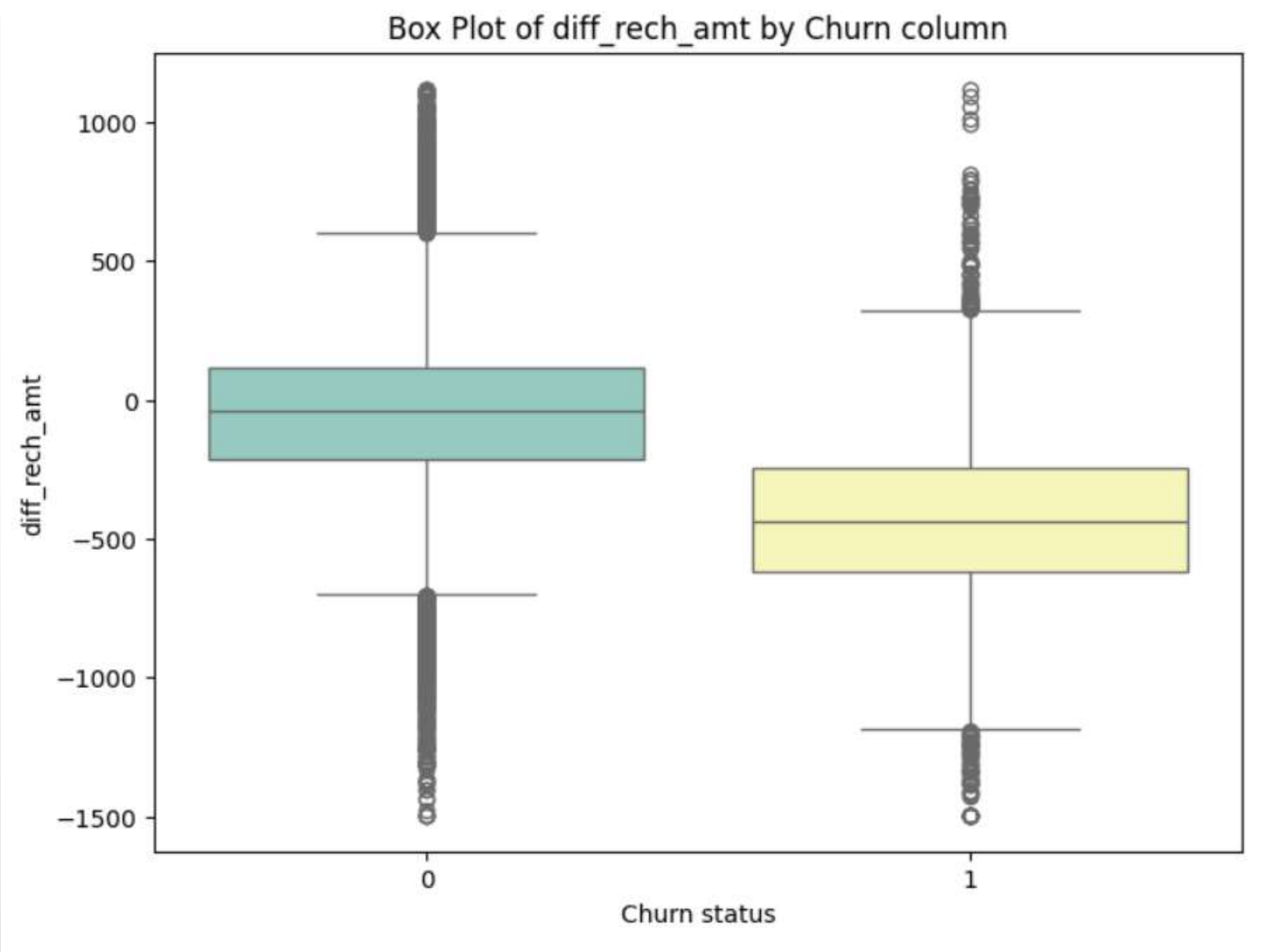
Observation

- As we can see the same pattern, churn customers exhibit similar total incoming MOU in good phase(months 6 and 7), but show significantly lower total incoming MOU in action phase(month 8) compared to non-churn customers, indicating a decline in engagement.



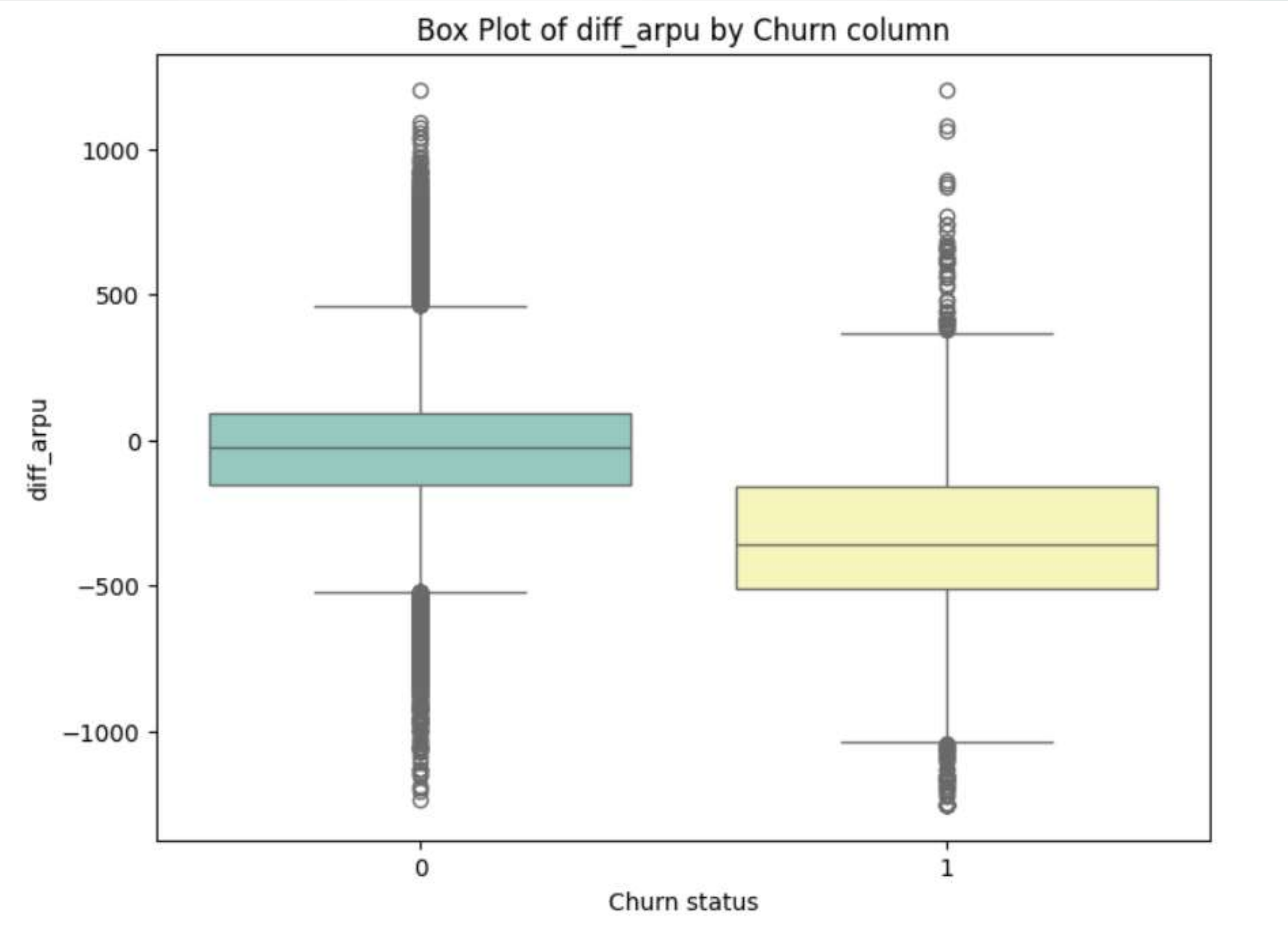
Difference of Recharge Amount during good phase and action phase with respect to churn status

- As we can see the difference of recharge amount during good phase and action phase is tend to more negative for churn customer as compare to non churn customer

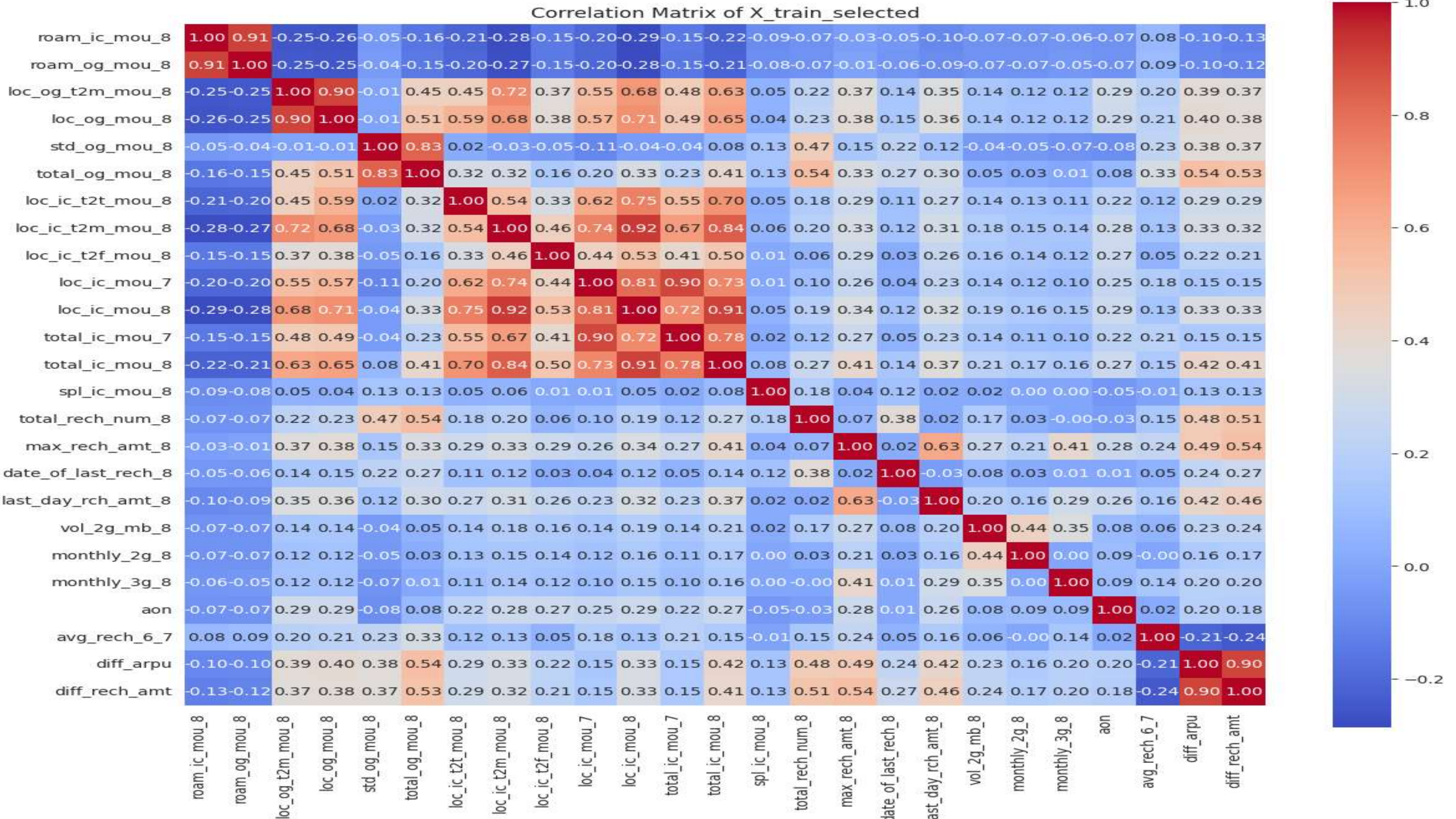


Difference of Average Revenue Per User during the good phase (6 and 7 month) and action phase (8th month) with respect to churn status

- As we can see difference of ARPU during good phase and action phase is tend to more negative for churn customer as compare to non churn customer



- **Calculating the correlation matrix and creating the heatmap to see the relationship between variables**


Correlation Matrix of X_train_selected
roam_ic_mou_8
roam_og_mou_8
loc_og_t2m_mou_8
loc_og_mou_8
std_og_mou_8
total_og_mou_8
loc_ic_t2t_mou_8
loc_ic_t2m_mou_8
loc_ic_t2f_mou_8
loc_ic_mou_7
loc_ic_mou_8
total_ic_mou_7
total_ic_mou_8
spl_ic_mou_8
total_rech_num_8
max_rech_amt_8
date_of_last_rech_8
last_day_rch_amt_8
vol_2g_mb_8
monthly_2g_8
monthly_3g_8
aon
avg_rech_6_7
diff_arpu
diff_rech_amt
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Some of the top correlation

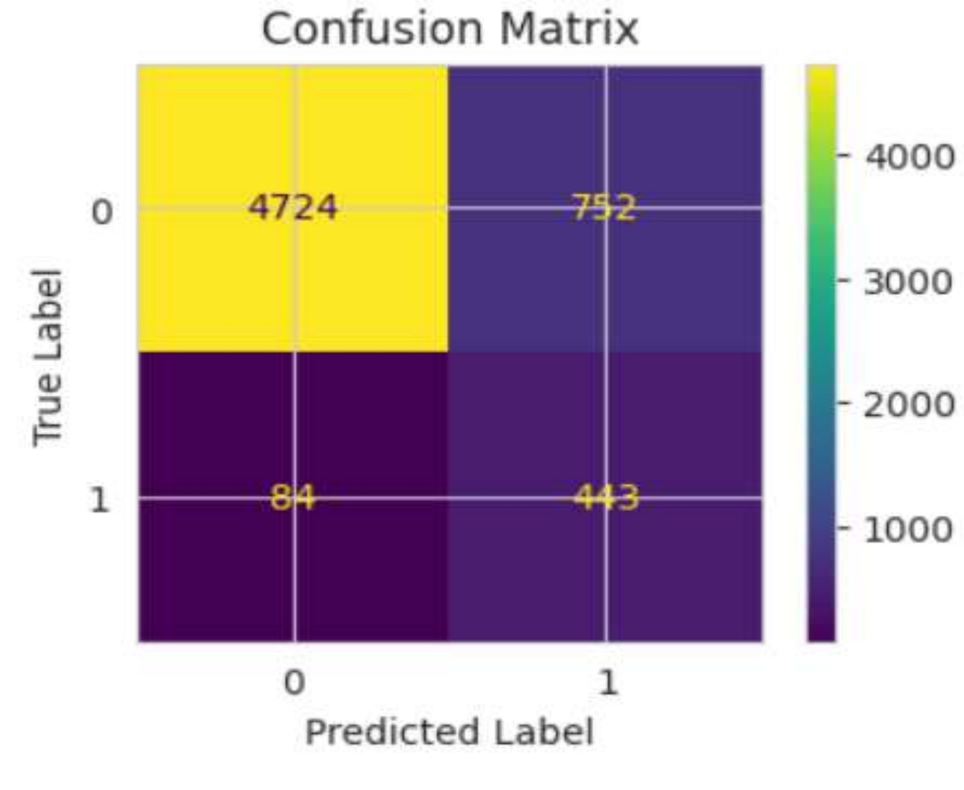
- diff_arpu and diff_rech_amt 0.90
- roam_ic_mou_8 and roam_og_mou_8 0.91
- loc_og_mou_8 and loc_og_t2m_mou_8 0.90
- total_ic_mou_7 and loc_ic_mou_7 0.90
- loc_ic_mou_8 and loc_ic_t2m_mou_8 0.92
- total_ic_mou_8 and loc_ic_mou_8 0.91

Model Evaluation (Logistic Regression)

Initial Model Evaluation

- **Accuracy**: 86.07%

- **Classification Metrics**:
- **Precision**:
    - Class 0: 0.98
    - Class 1: 0.37
- **Recall**:
    - Class 0: 0.86
    - Class 1: 0.84
- **F1-score**:
    - Class 0: 0.92
    - Class 1: 0.51



Hyperparameter Tuning

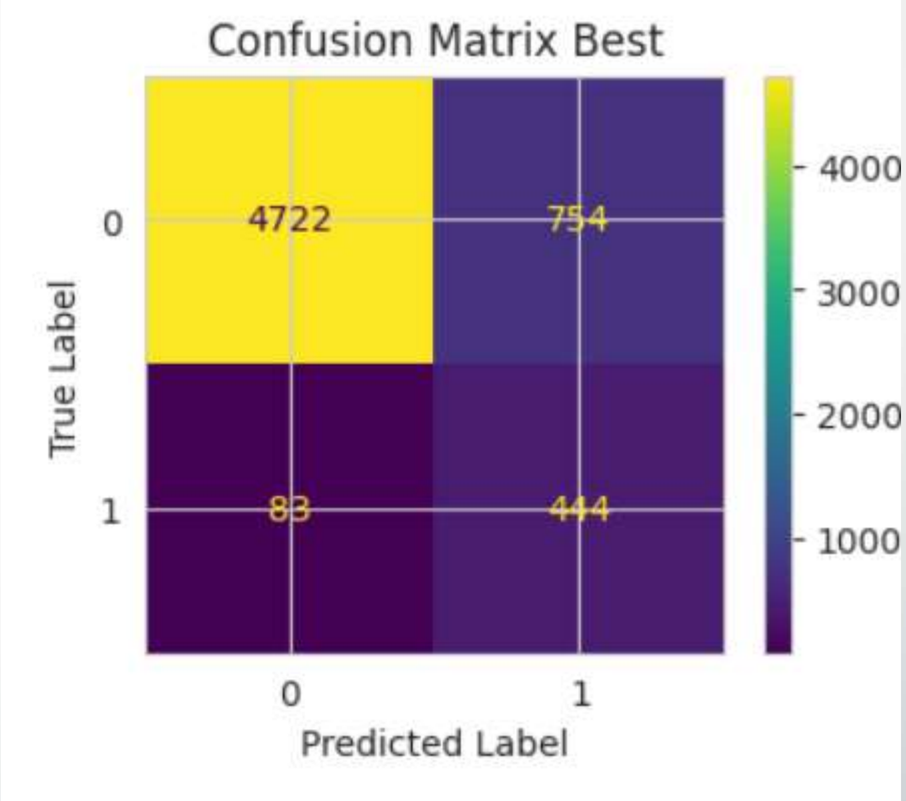
- **Method**: GridSearchCV
- **Parameter Grid**:
  - **Penalty**: ['l1', 'l2', 'elasticnet', 'none']
  - **C (Regularization)**: [0.01, 0.1, 1, 10, 100]
  - **Solver**: [‘lbfgs', 'liblinear', 'saga']
- **Goal**: Optimize model performance by finding the best combination of hyperparameters.

Tuned Model Results

- **Accuracy**: 86.05%

- **Classification Metrics**:
- **Precision**:
  - Class 0: 0.98
  - Class 1: 0.37
- **Recall**:
  - Class 0: 0.86
  - Class 1: 0.84
- **F1-score**:
  - Class 0: 0.92
  - Class 1: 0.51

- No changes in the tuned model indicate optimal parameter settings were initially chosen, ensuring no further improvement was necessary.



Conclusion and Recommendations

Summary of Key Insights

- **High Churn Risk Indicators:**
  - Significant declines in Average Revenue Per User (ARPU), total outgoing minutes of usage (MOU), and recharge amounts during the action phase (Month 8) were strong indicators of potential churn.
  - High-value customers showed distinct patterns that allowed for targeted analysis.
- **Model Performance:**
  - The initial logistic regression model achieved an accuracy of 86.07%.
  - Hyperparameter tuning confirmed the robustness of the initial settings, with no significant improvements observed.

Recommendations

- **Targeted Retention Strategies:**
  - **Personalized Offers:** Develop personalized retention strategies for identified high-risk customer segments, such as offering tailored promotions or enhanced customer service, especially focusing on customers with declining ARPU and recharge amounts.
- **High-Value Customer Focus:** Since 80% of revenue comes from the top 20% of customers, prioritize retention strategies for high-value customers to maximize revenue retention. Tailor customer support and enhance service quality for this segment to maintain satisfaction and loyalty.

Thank You