TELECOM CHURN CASE STUDY

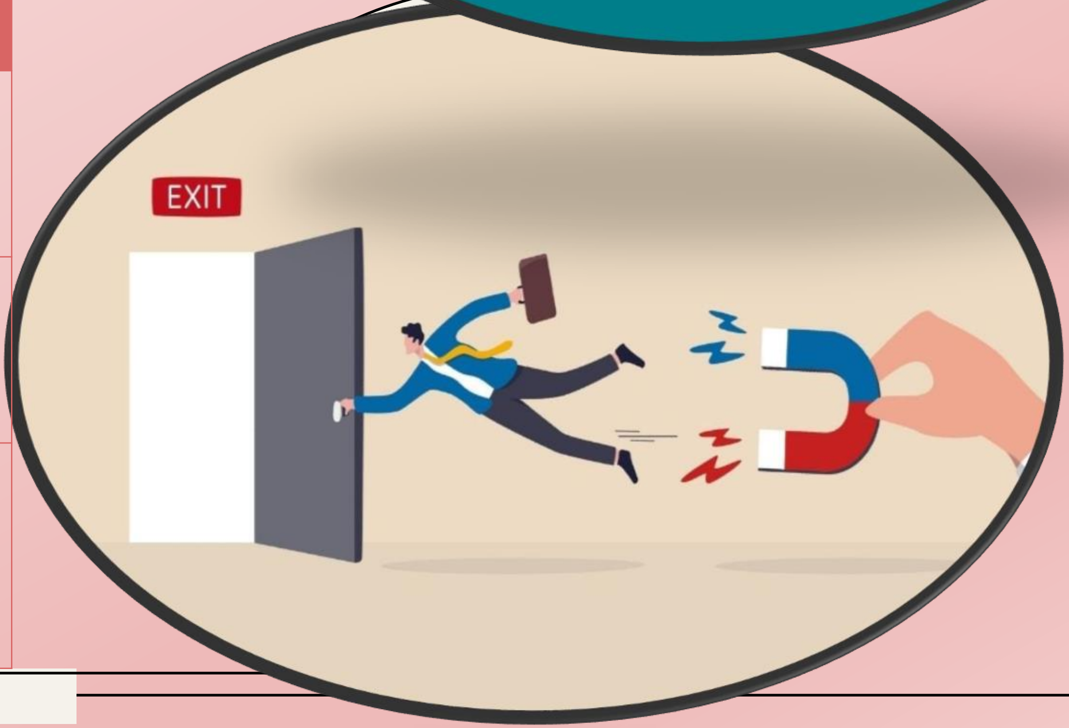
Y. SWETHA ASHERVADAM

CAPSTONE PROJECT

CUSTOMER BEHAVIOUR

Customers usually do not decide to switch to another competitor instantly but rather over a period of time (this is especially applicable to high-value customers). In churn prediction we assume that there are three phases of the customer lifecycle :

Good Phase	The customer is happy with the service
Action Phase	The customer experience starts to become sore
Churn Phase	The customer is said to have churned



BUSINESS OBJECTIVE

The **business objective** is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months. To do this task well understanding the typical customer behaviour during churn will be helpful.

In this case the dataset is over a four-month window

- The first two months are the good phase – **June July**
- The third month is the action phase - **August**
- The fourth month is the churn phase - **September**

STEPS INVOLVED

Filter High Value Customers

Identify high-value customers based on their recharge amounts in the first two months, ensuring a robust dataset for churn prediction.

Tag Churners

Label customers as churners based on their activity (or lack thereof) during the churn phase, and remove attributes corresponding to this phase from the dataset.

Deal with Missing Values

Drop columns which are more than 40% and delete the rows which are having null values less than 6%.

Build Model

Predictive Model using Random Forest Classifier

Logistic Regression Model to identify predictor variables

Test Model

Deploy the model on the test set

FILTER HIGH-VALUE CUSTOMERS

As mentioned earlier the need to predict churn is only for high-value customers.

Define high-value customers as follows:

Those who have recharged with an amount more than or equal to X where X is the 70th percentile of the average recharge amount in the first two months (the good phase).

After filtering the high-value customers there are about 30011 rows.

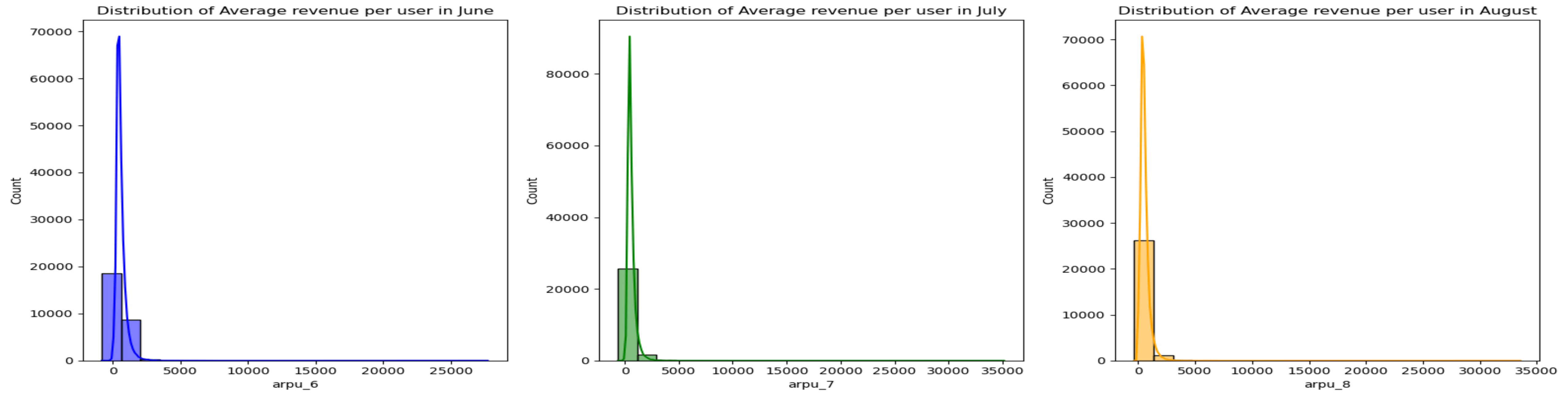
TAG CHURNERS

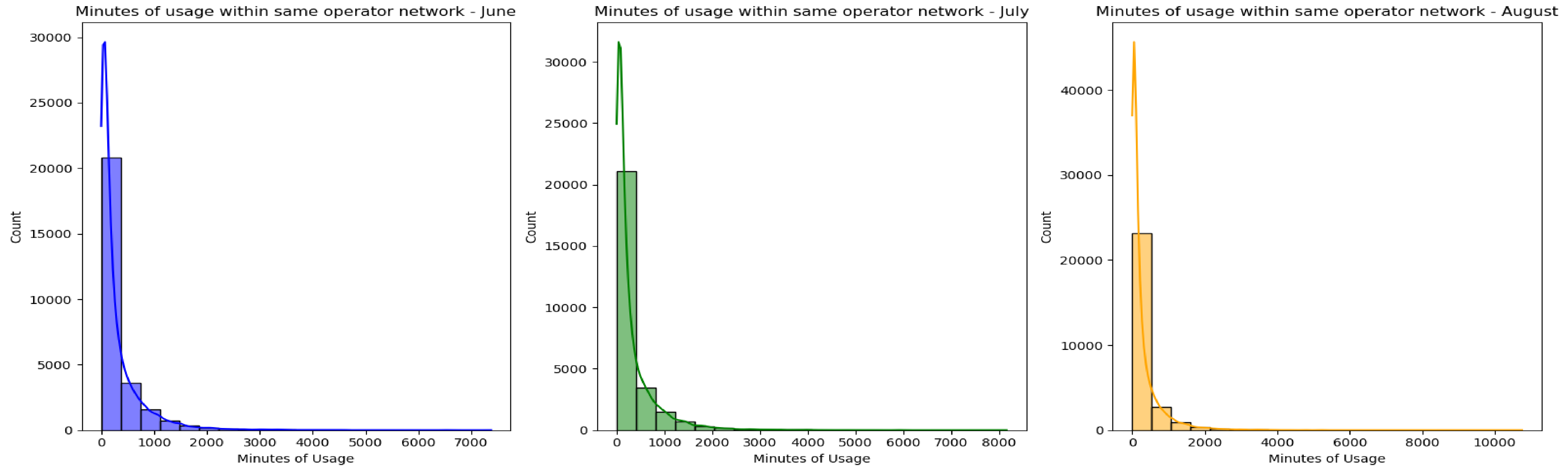
Now tag the churned customers (churn=1 else 0) based on the fourth month as follows: Those who have not made any calls (either incoming or outgoing) AND have not used mobile internet even once in the churn phase.

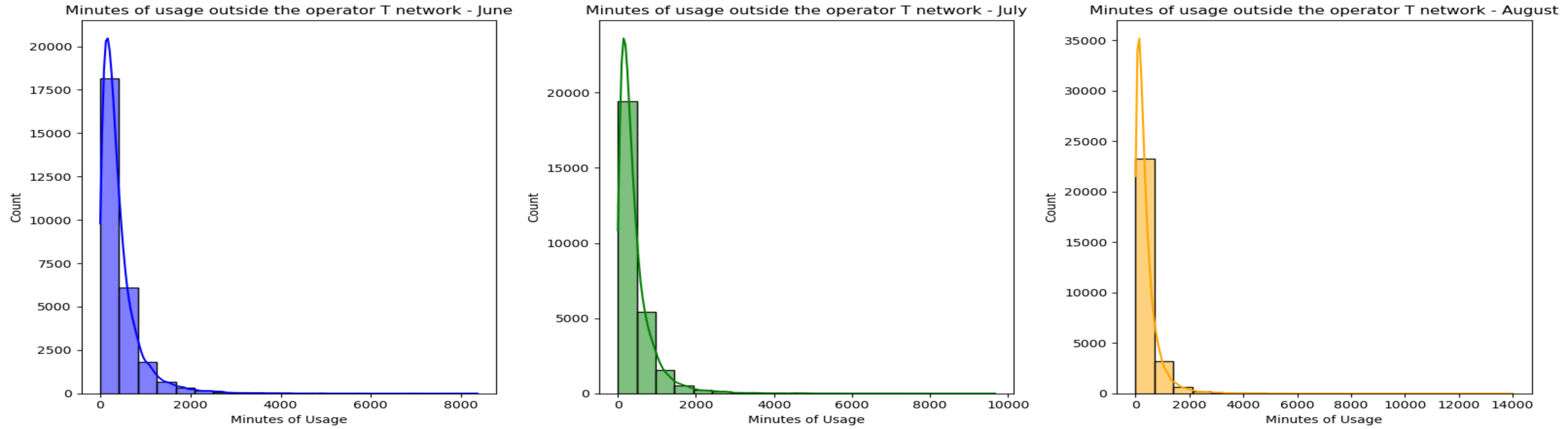
The attributes you need to use to tag churners are:
total_ic_mou_9
total_og_mou_9
vol_2g_mb_9
vol_3g_mb_9

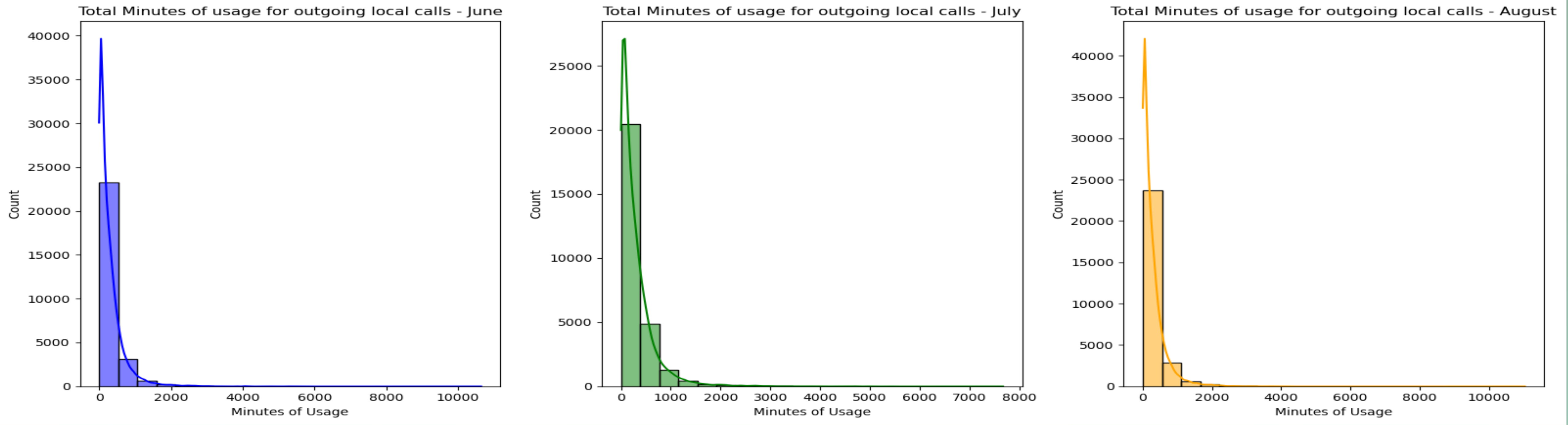
After tagging churners remove all the attributes corresponding to the churn phase (September)

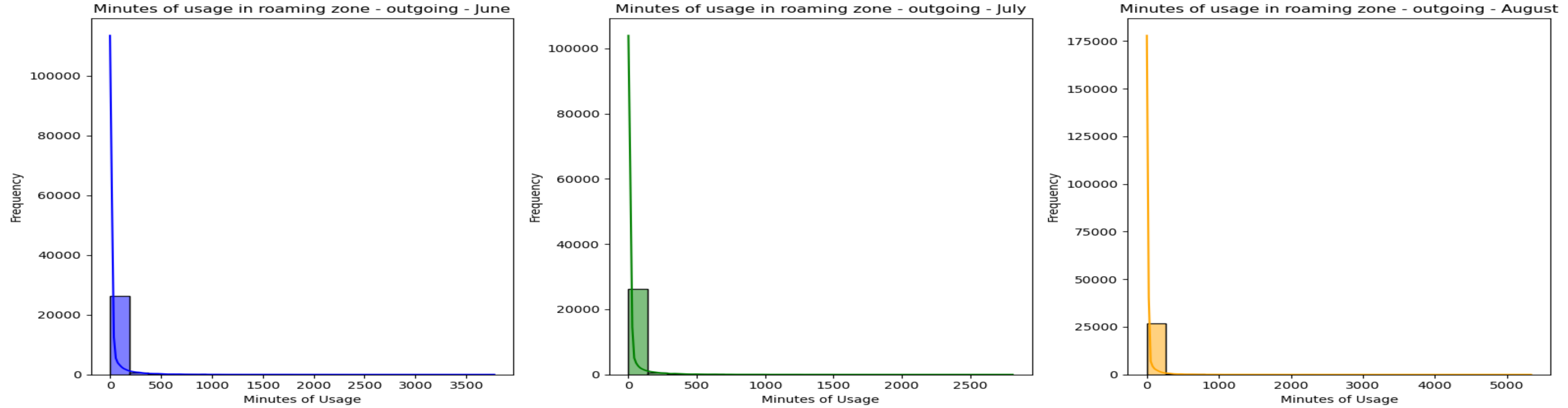
UNIVARIATE ANALYSIS

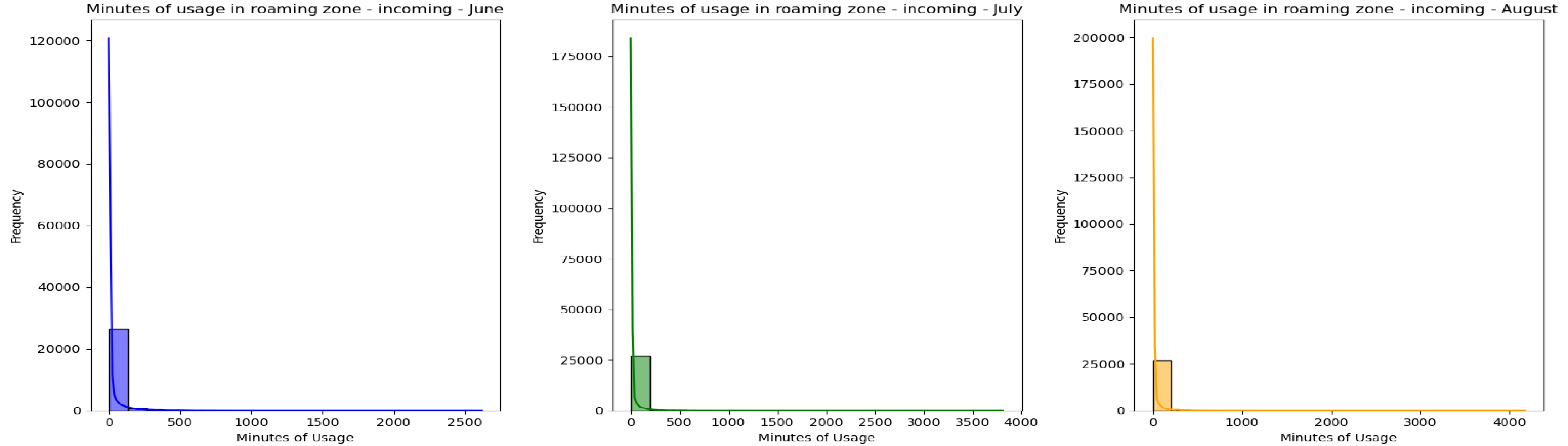

Distribution of Average revenue per user in June
Count
arpu_6
Distribution of Average revenue per user in July
Count
arpu_7
Distribution of Average revenue per user in August
Count
arpu_8

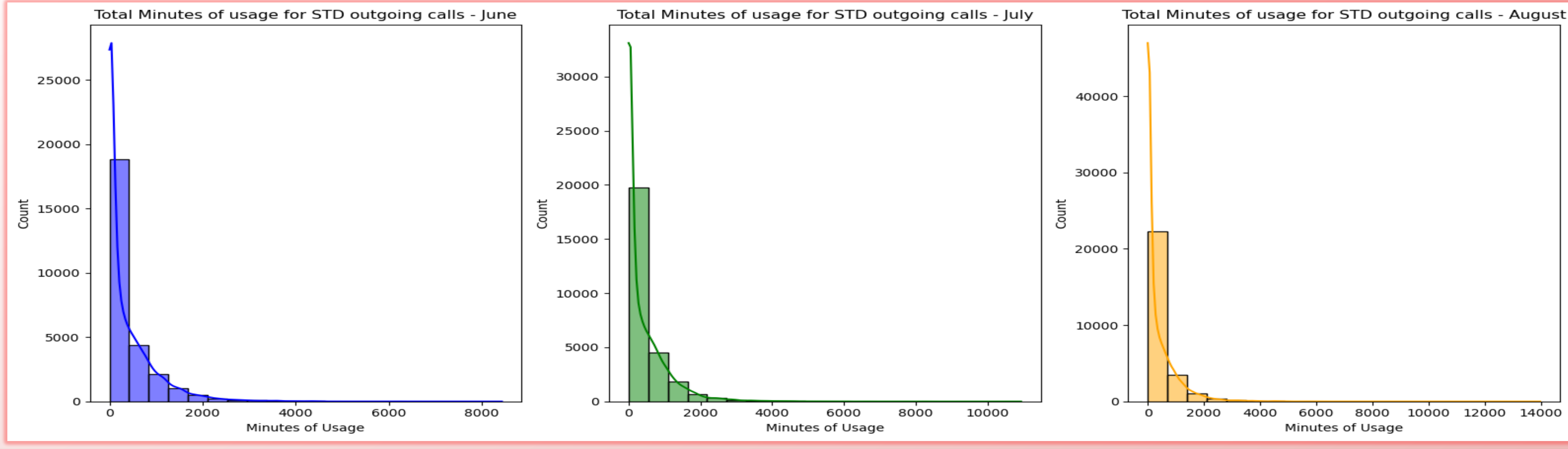

Minutes of usage within same operator network - June
Count
Minutes of Usage
Minutes of usage within same operator network - July
Count
Minutes of Usage
Minutes of usage within same operator network - August
Count
Minutes of Usage

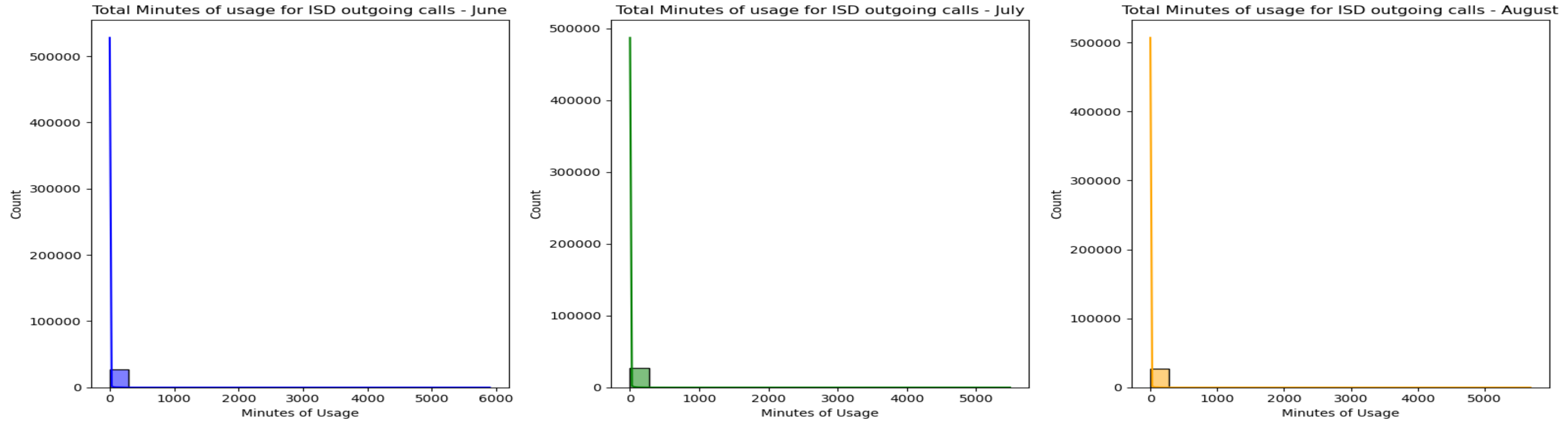

Minutes of usage outside the operator T network - June
Minutes of usage outside the operator T network - July
Minutes of usage outside the operator T network - August
Count
Minutes of Usage

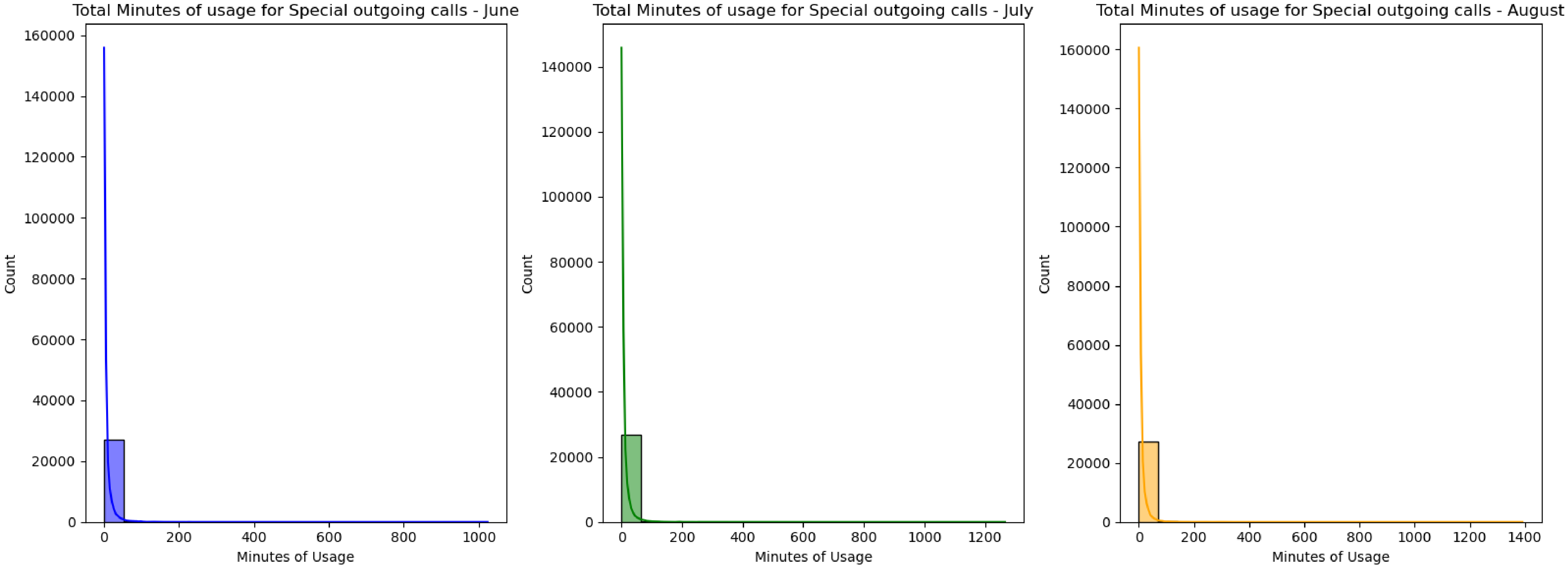

Total Minutes of usage for outgoing local calls - June
Total Minutes of usage for outgoing local calls - July
Total Minutes of usage for outgoing local calls - August
Count
Minutes of Usage

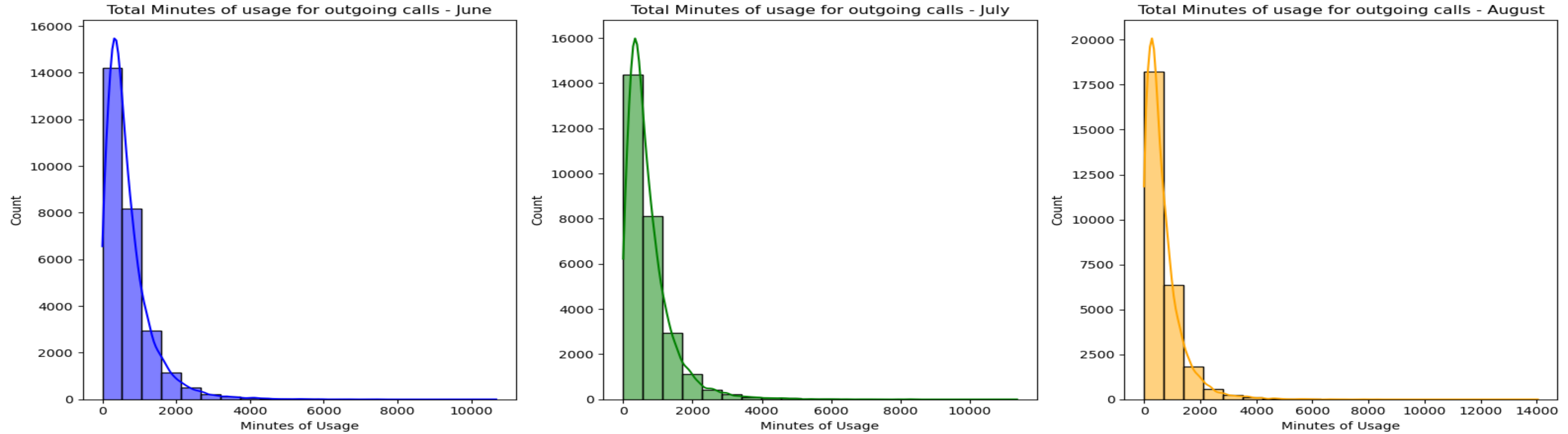

Minutes of usage in roaming zone - outgoing - June
Minutes of usage in roaming zone - outgoing - July
Minutes of usage in roaming zone - outgoing - August
Frequency
Minutes of Usage
0
20000
40000
60000
80000
100000
500
1000
1500
2000
2500
3000
3500
25000
50000
75000
125000
150000
175000
4000
5000

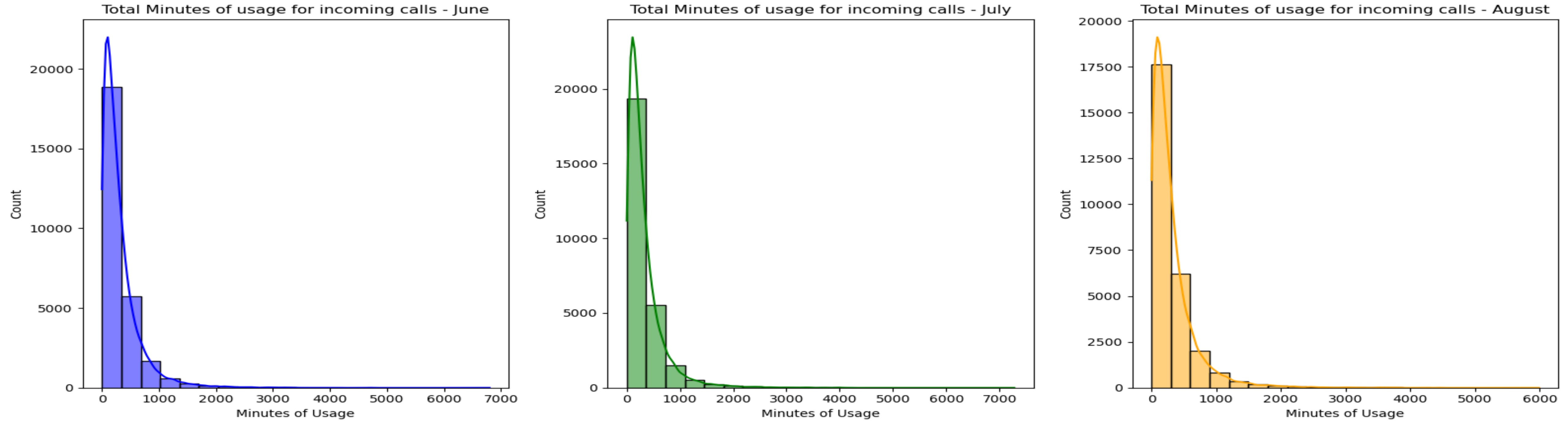

Minutes of usage in roaming zone - incoming - June
Minutes of usage in roaming zone - incoming - July
Minutes of usage in roaming zone - incoming - August
Frequency
Minutes of Usage
0
20000
40000
60000
80000
100000
120000
25000
50000
75000
125000
150000
175000
200000
500
1000
1500
2000
2500
3000
3500
4000

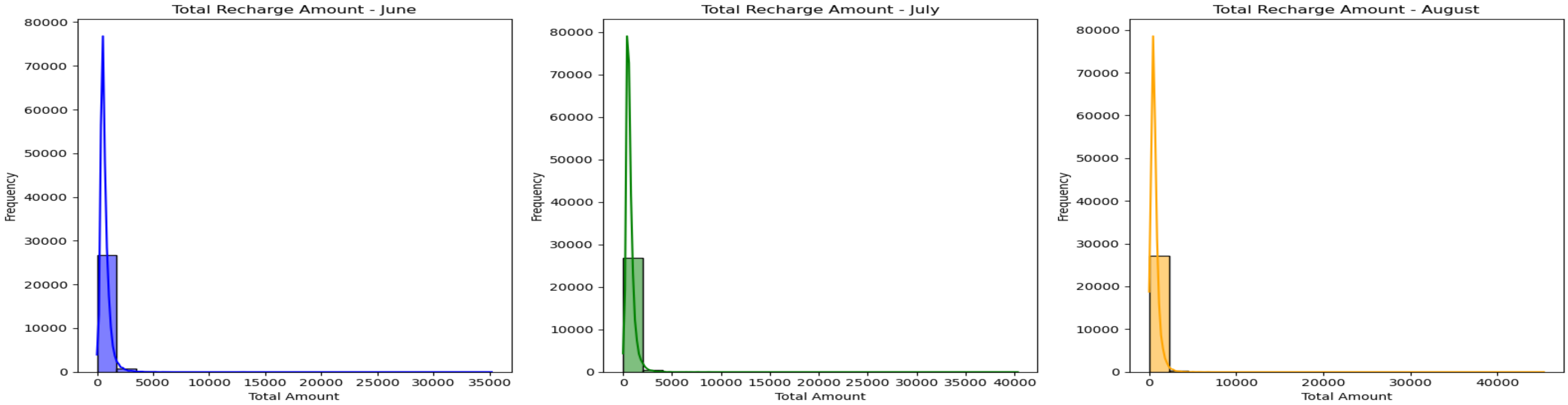

Total Minutes of usage for STD outgoing calls - June
Count
Minutes of Usage
0
5000
10000
15000
20000
25000
0
2000
4000
6000
8000
Total Minutes of usage for STD outgoing calls - July
Count
Minutes of Usage
0
5000
10000
15000
20000
25000
30000
0
2000
4000
6000
8000
10000
Total Minutes of usage for STD outgoing calls - August
Count
Minutes of Usage
0
10000
20000
30000
40000
0
2000
4000
6000
8000
10000
12000
14000

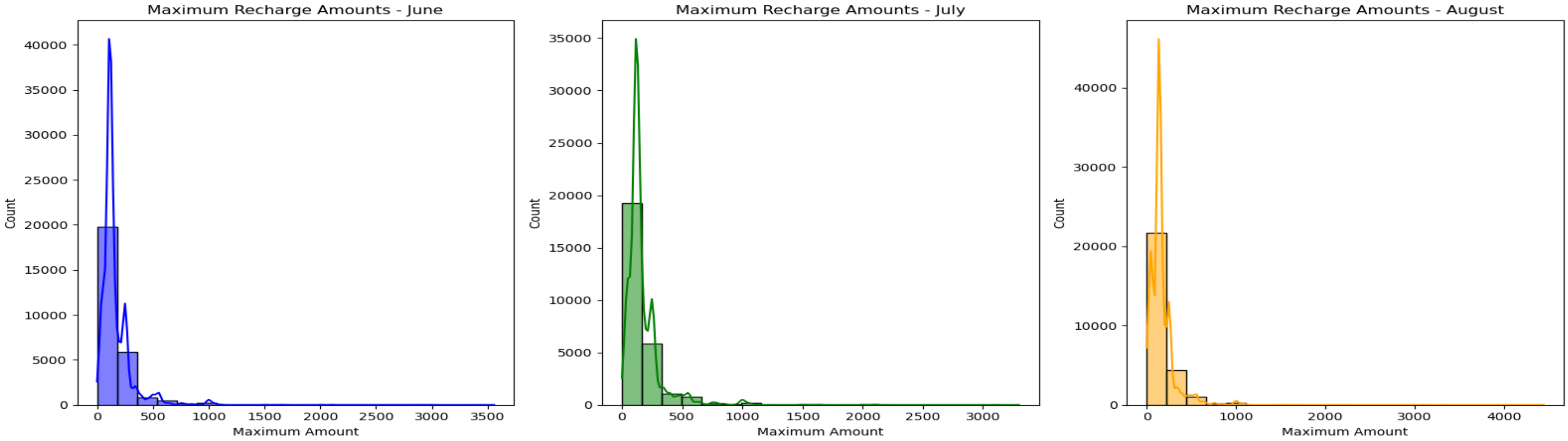

Total Minutes of usage for ISD outgoing calls - June
Count
Minutes of Usage
0
100000
200000
300000
400000
500000
0
1000
2000
3000
4000
5000
6000
Total Minutes of usage for ISD outgoing calls - July
Count
Minutes of Usage
0
100000
200000
300000
400000
500000
0
1000
2000
3000
4000
5000
Total Minutes of usage for ISD outgoing calls - August
Count
Minutes of Usage
0
100000
200000
300000
400000
500000
0
1000
2000
3000
4000
5000

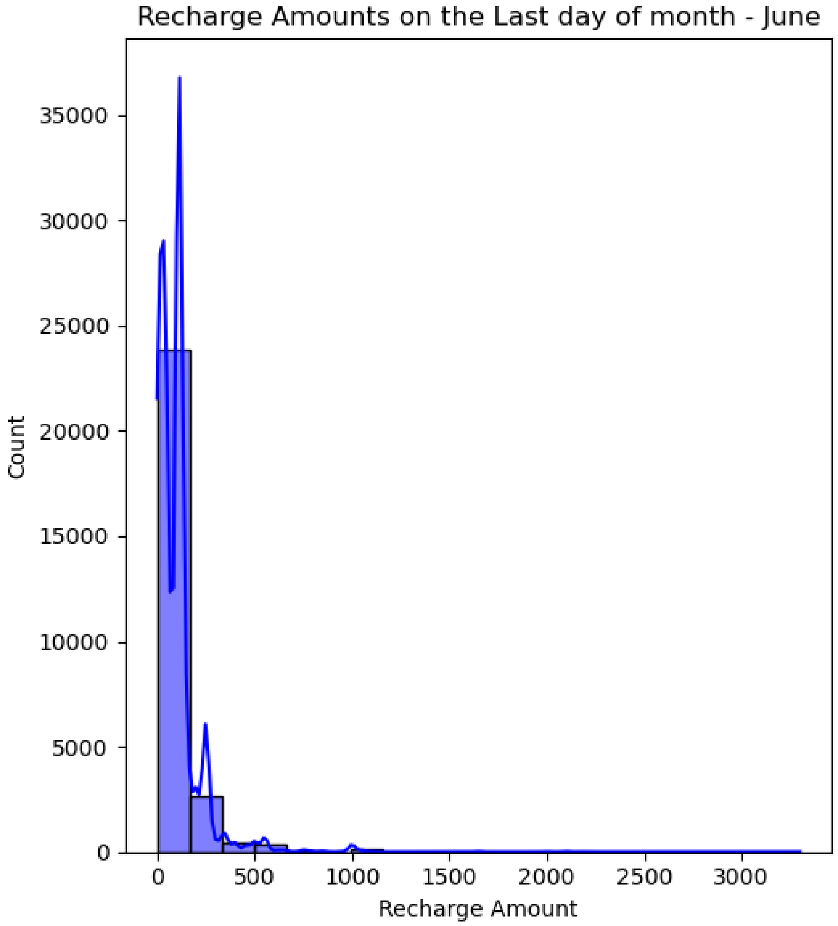

Total Minutes of usage for Special outgoing calls - June
Count
Minutes of Usage
Total Minutes of usage for Special outgoing calls - July
Count
Minutes of Usage
Total Minutes of usage for Special outgoing calls - August
Count
Minutes of Usage

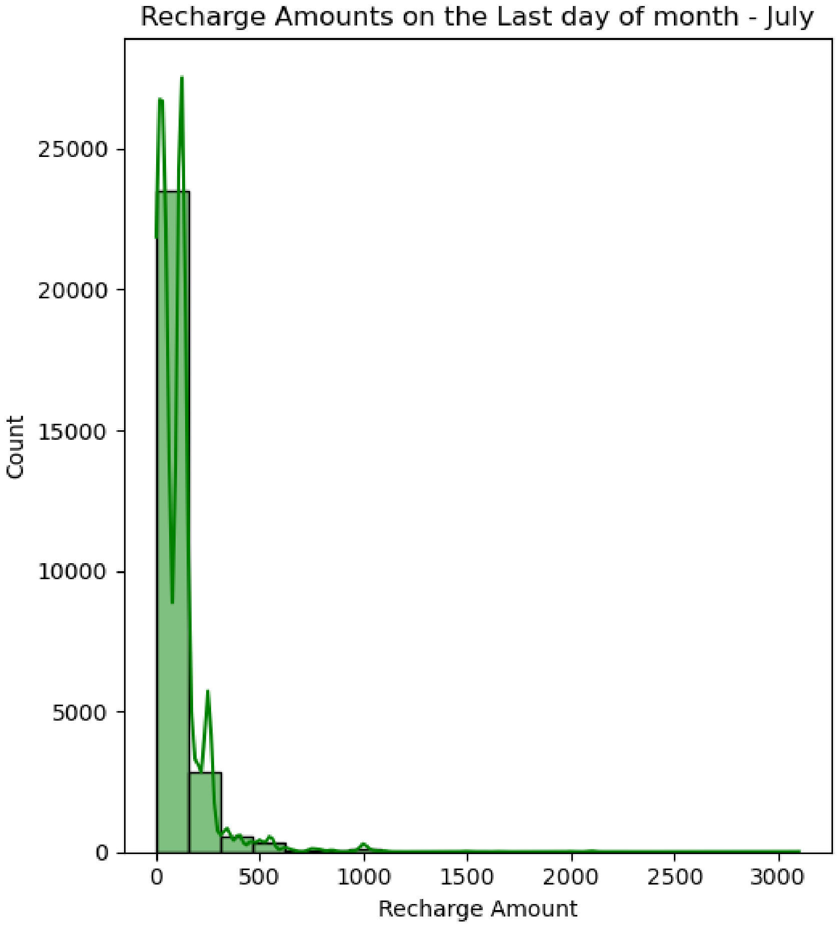

Total Minutes of usage for outgoing calls - June
Total Minutes of usage for outgoing calls - July
Total Minutes of usage for outgoing calls - August
Count
Minutes of Usage

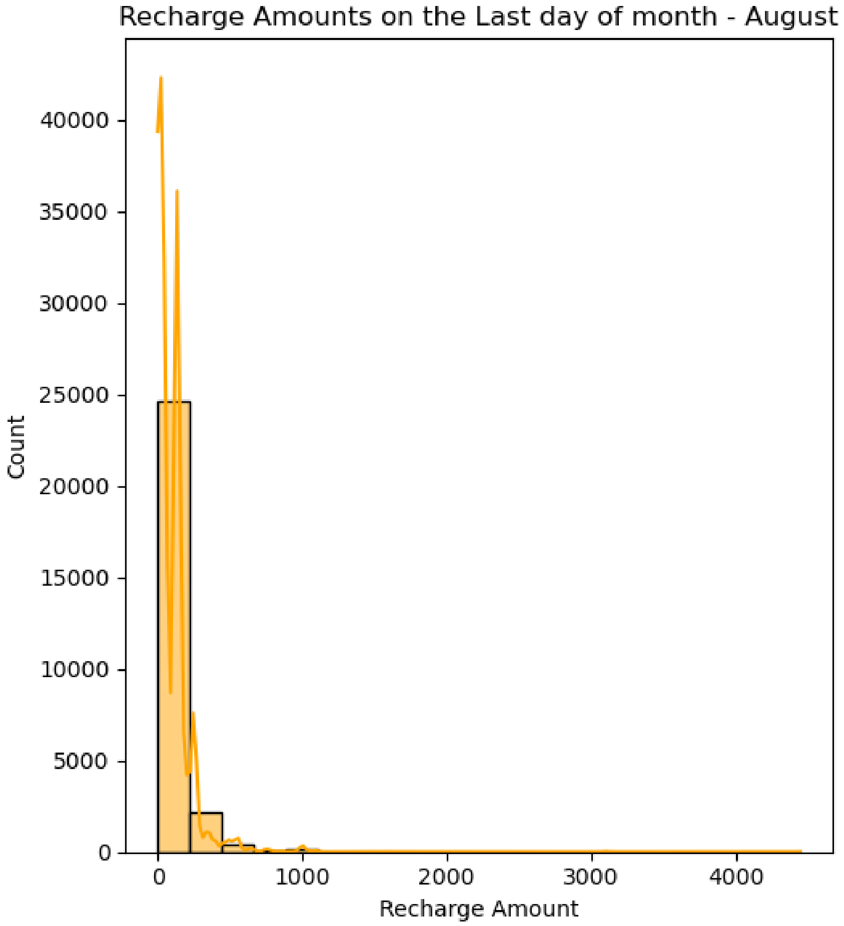

Total Minutes of usage for incoming calls - June
Total Minutes of usage for incoming calls - July
Total Minutes of usage for incoming calls - August
Count
Minutes of Usage

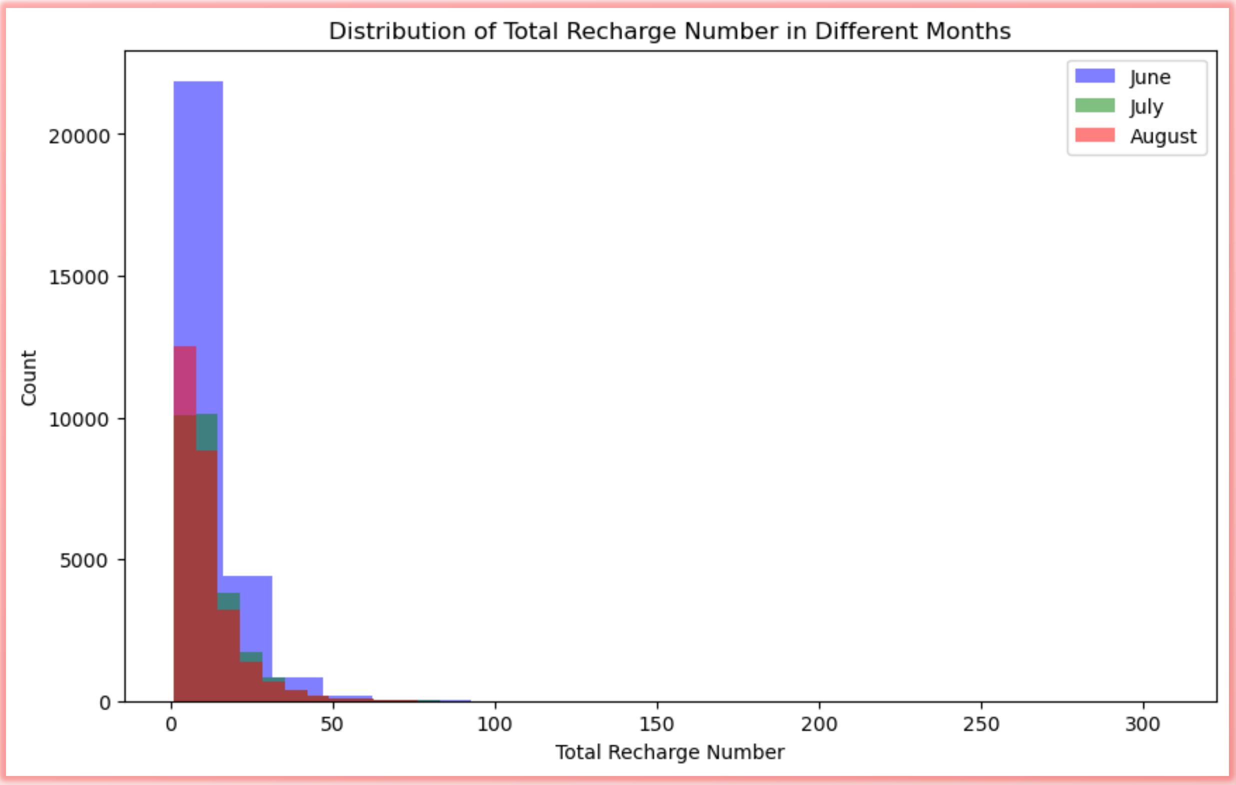

Total Recharge Amount - June
Total Recharge Amount - July
Total Recharge Amount - August
Frequency
Total Amount
0 10000 20000 30000 40000 50000 60000 70000 80000
0 5000 10000 15000 20000 25000 30000 35000
0 5000 10000 15000 20000 25000 30000 35000 40000
0 10000 20000 30000 40000

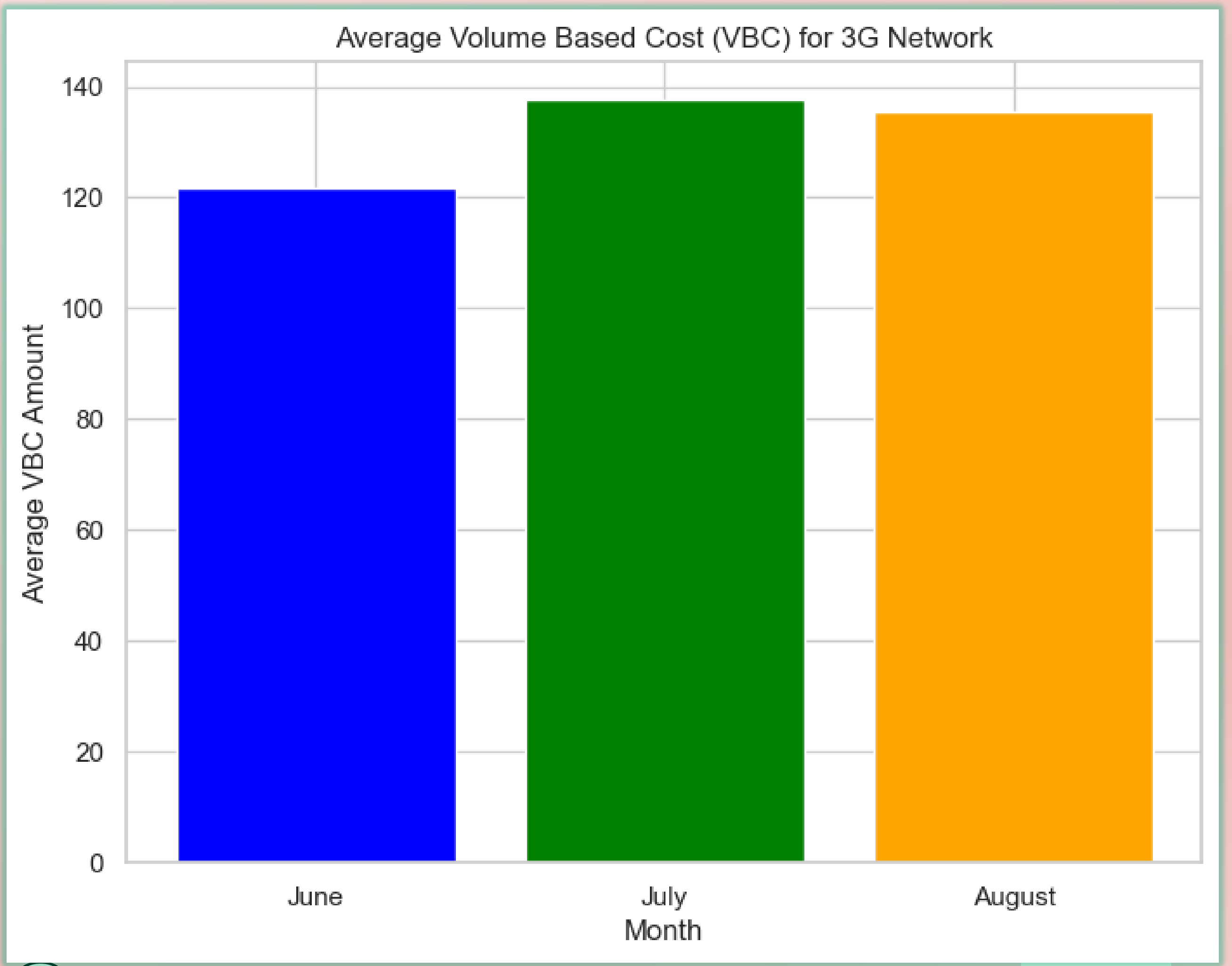

Maximum Recharge Amounts - June
Maximum Recharge Amounts - July
Maximum Recharge Amounts - August
Count
Maximum Amount
0 5000 10000 15000 20000 25000 30000 35000 40000
0 500 1000 1500 2000 2500 3000 3500
0 500 1000 1500 2000 2500 3000
0 10000 20000 30000 40000
0 1000 2000 3000 4000

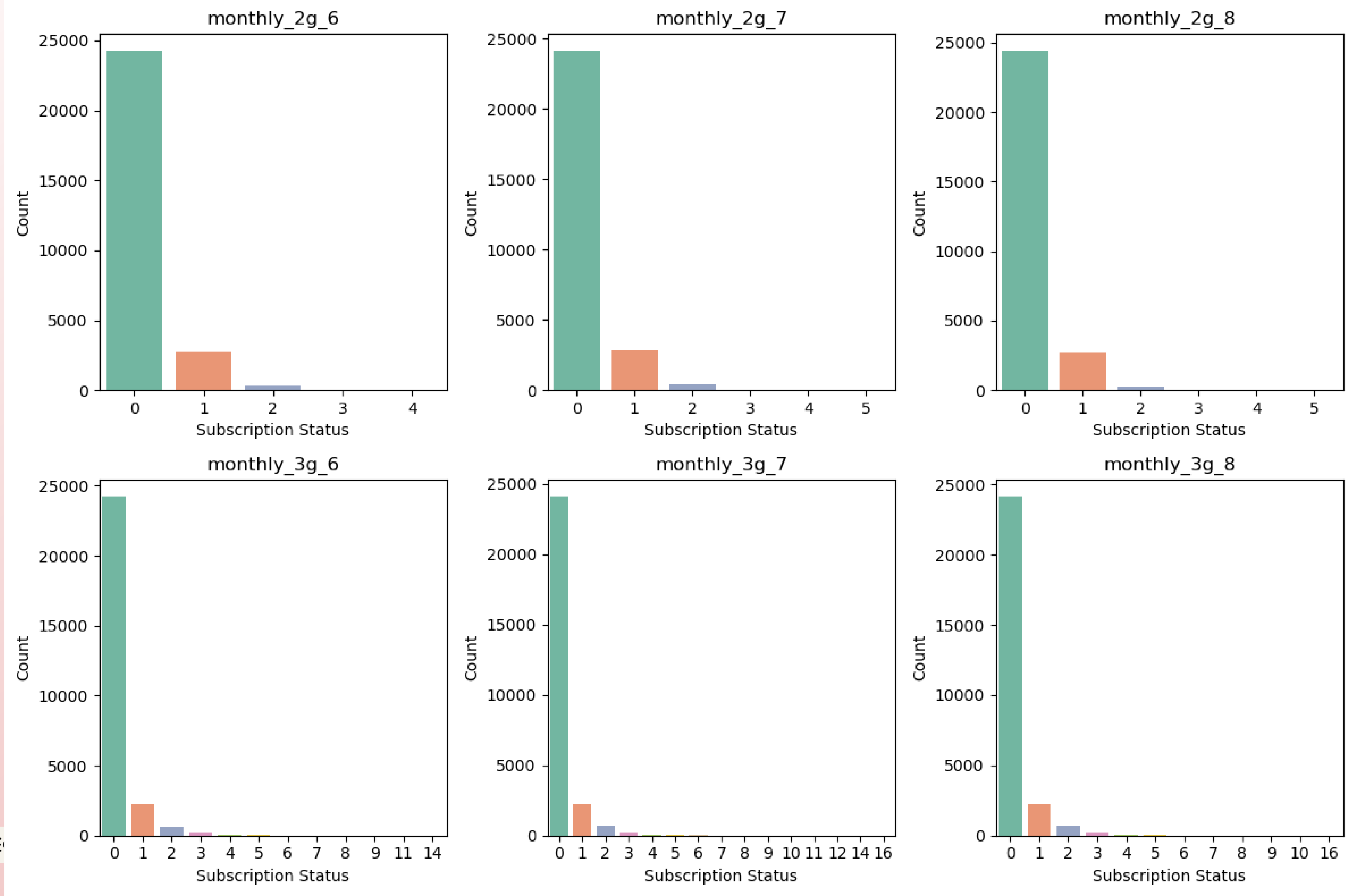

Recharge Amounts on the Last day of month - June
Count
Recharge Amount
0
5000
10000
15000
20000
25000
30000
35000
0
500
1000
1500
2000
2500
3000

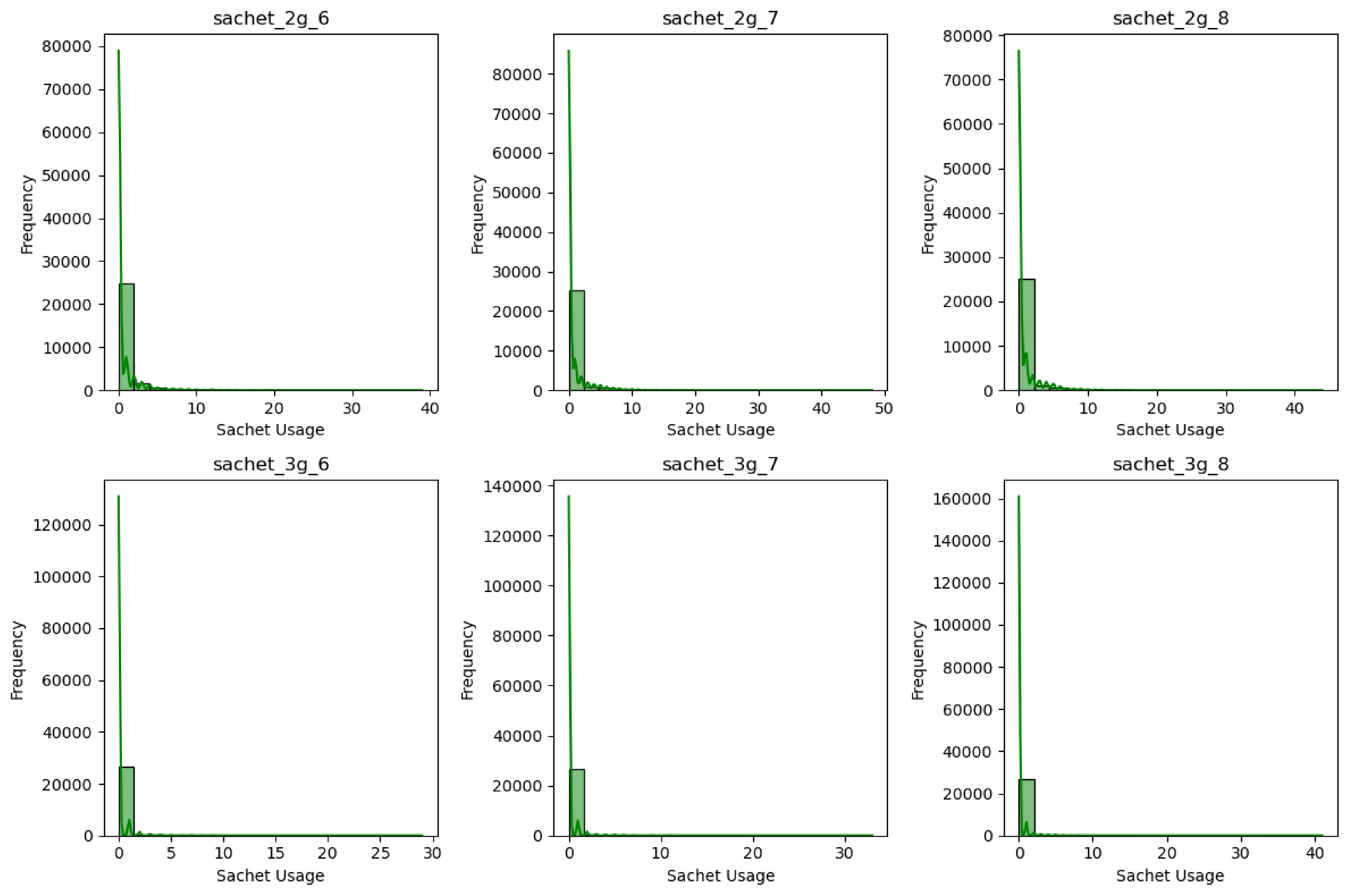

Recharge Amounts on the Last day of month - July
Count
Recharge Amount
0
5000
10000
15000
20000
25000
0
500
1000
1500
2000
2500
3000

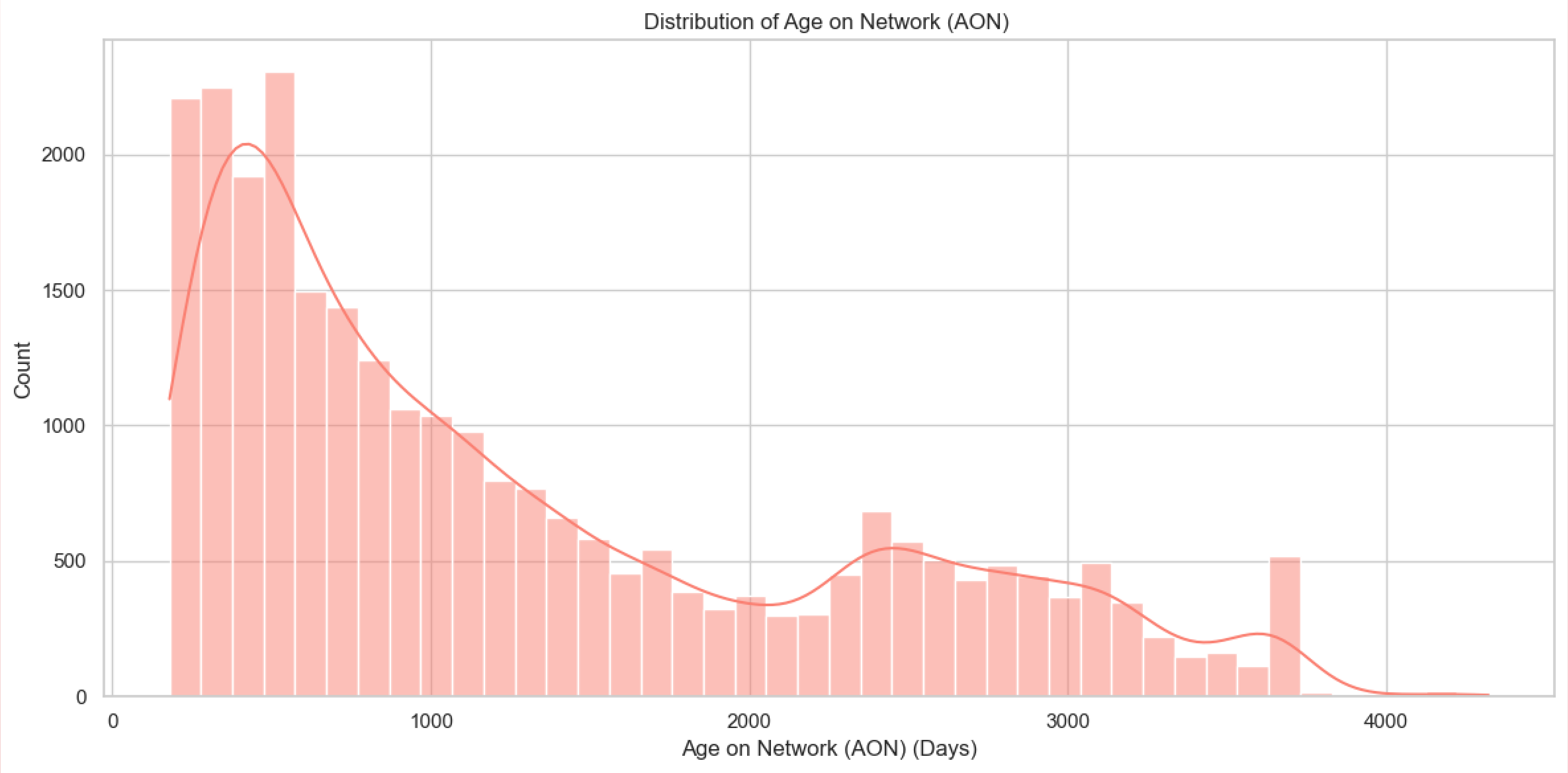

Recharge Amounts on the Last day of month - August
Count
Recharge Amount
0
5000
10000
15000
20000
25000
30000
35000
40000
0
1000
2000
3000
4000


Distribution of Total Recharge Number in Different Months
June
July
August
Count
20000
15000
10000
5000
0
0
50
100
150
200
250
300
Total Recharge Number

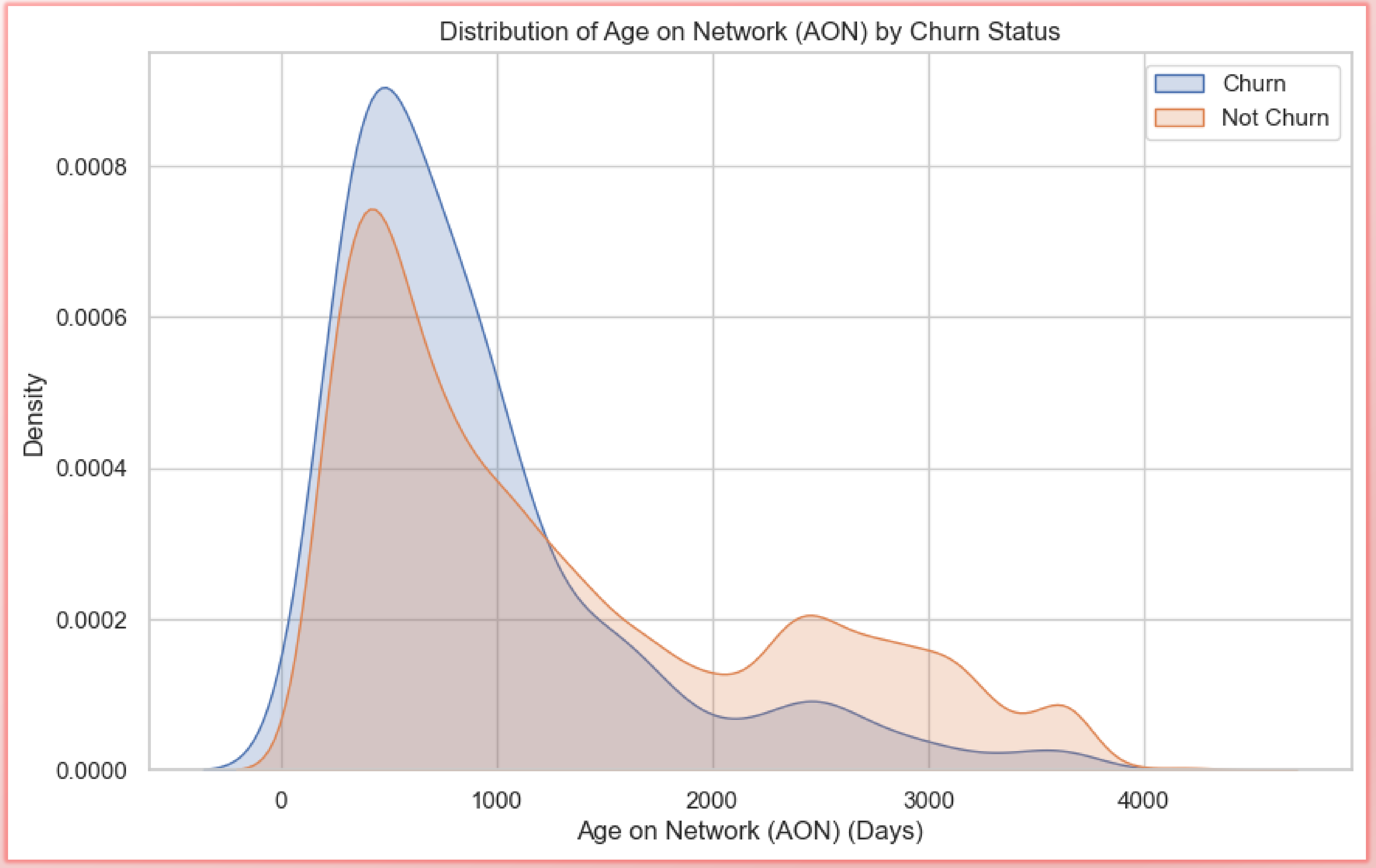

Average Volume Based Cost (VBC) for 3G Network
Average VBC Amount
140
120
100
80
60
40
20
0
June
July
August
Month

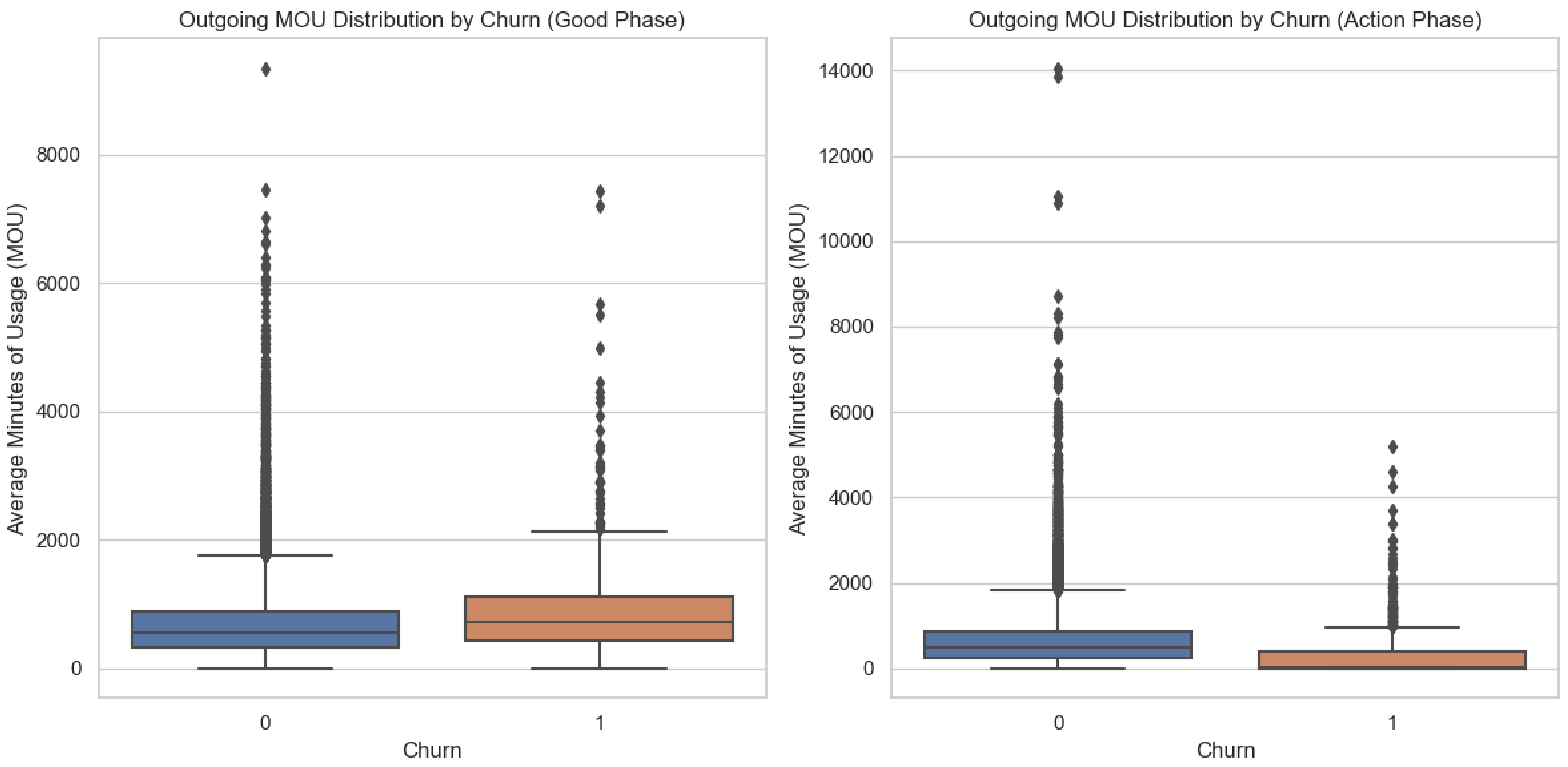

monthly_2g_6
monthly_2g_7
monthly_2g_8
monthly_3g_6
monthly_3g_7
monthly_3g_8
Count
Subscription Status
0
5000
10000
15000
20000
25000
0 1 2 3 4
0 1 2 3 4 5
0 1 2 3 4 5
0 1 2 3 4 5 6 7 8 9 11 14
0 1 2 3 4 5 6 7 8 9 10 11 12 14 16
0 1 2 3 4 5 6 7 8 9 10 16

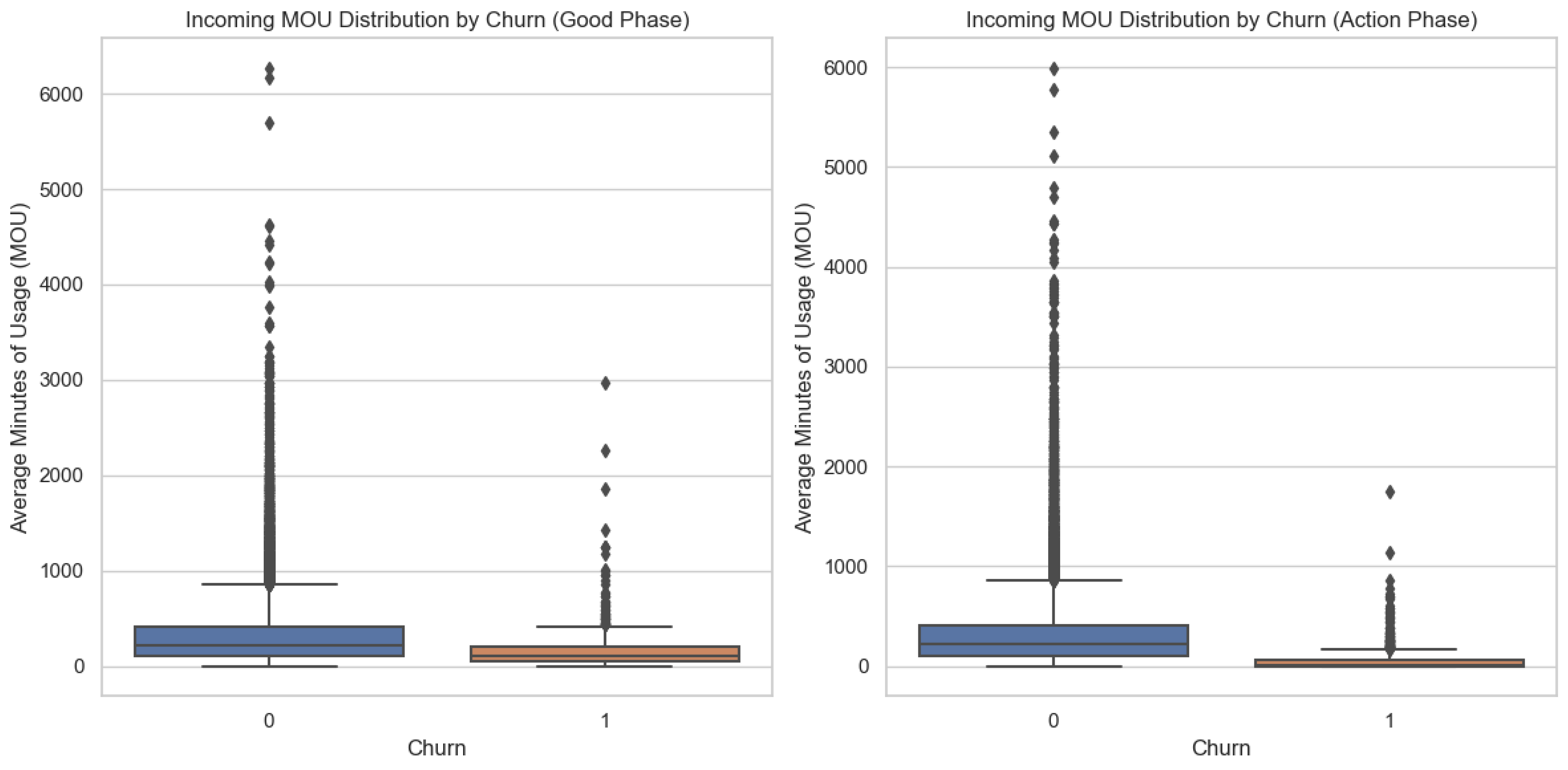

sachet_2g_6
sachet_2g_7
sachet_2g_8
sachet_3g_6
sachet_3g_7
sachet_3g_8
Frequency
Sachet Usage

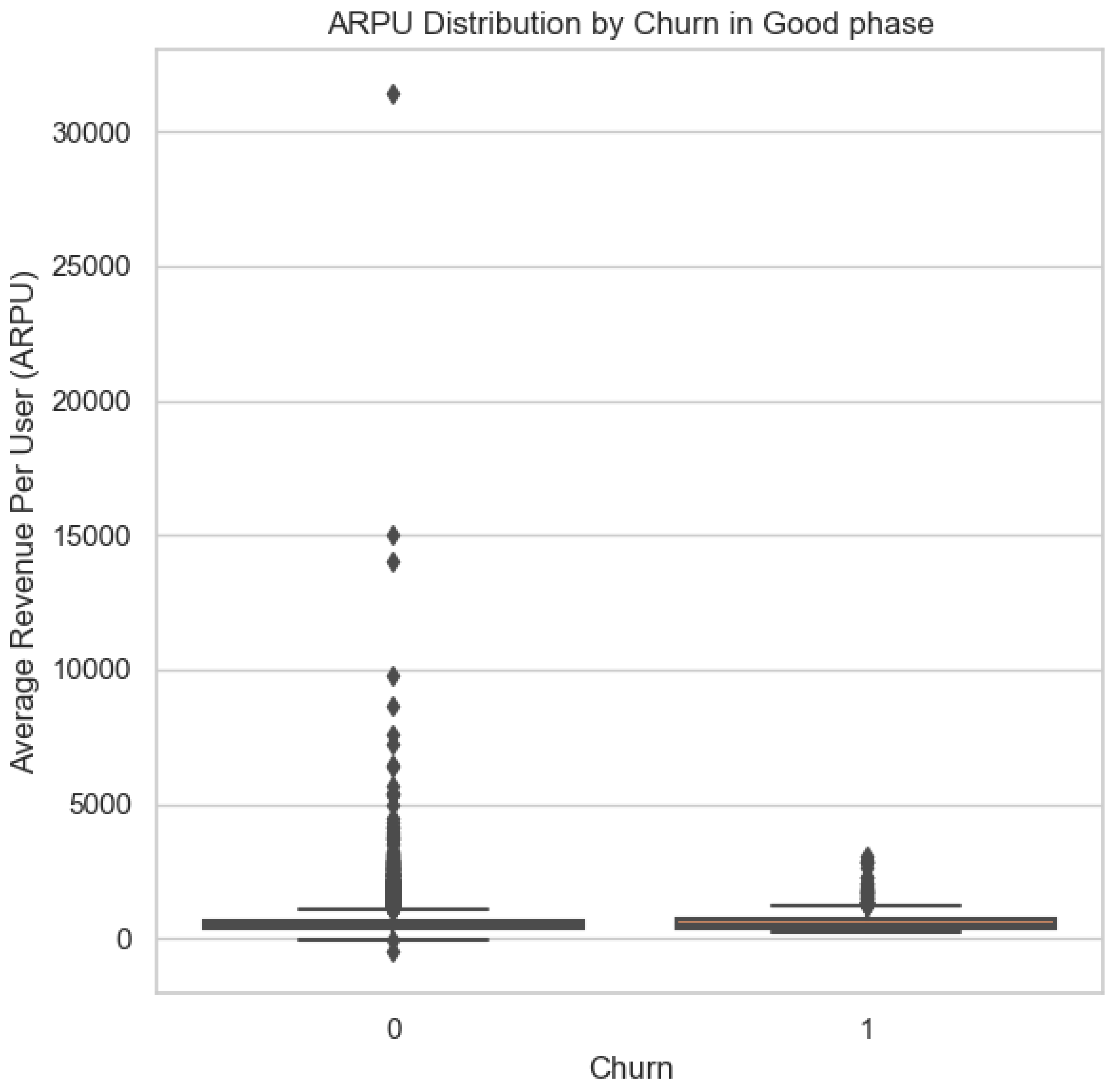

Distribution of Age on Network (AON)
Count
0
500
1000
1500
2000
0
1000
2000
3000
4000
Age on Network (AON) (Days)

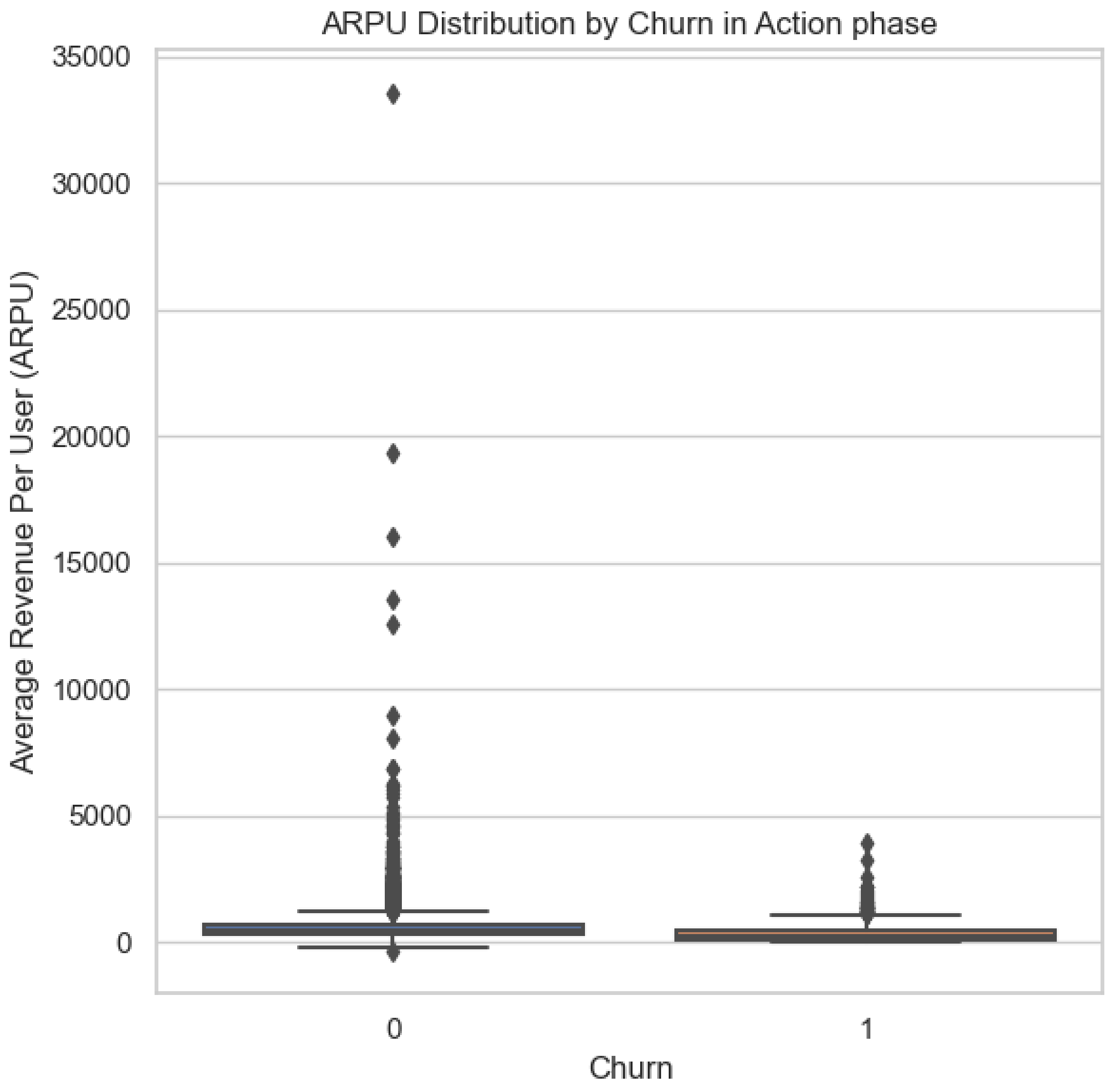
BIVARIATE ANALYSIS


Distribution of Age on Network (AON) by Churn Status
Churn
Not Churn
Density
0.0000
0.0002
0.0004
0.0006
0.0008
0
1000
2000
3000
4000
Age on Network (AON) (Days)

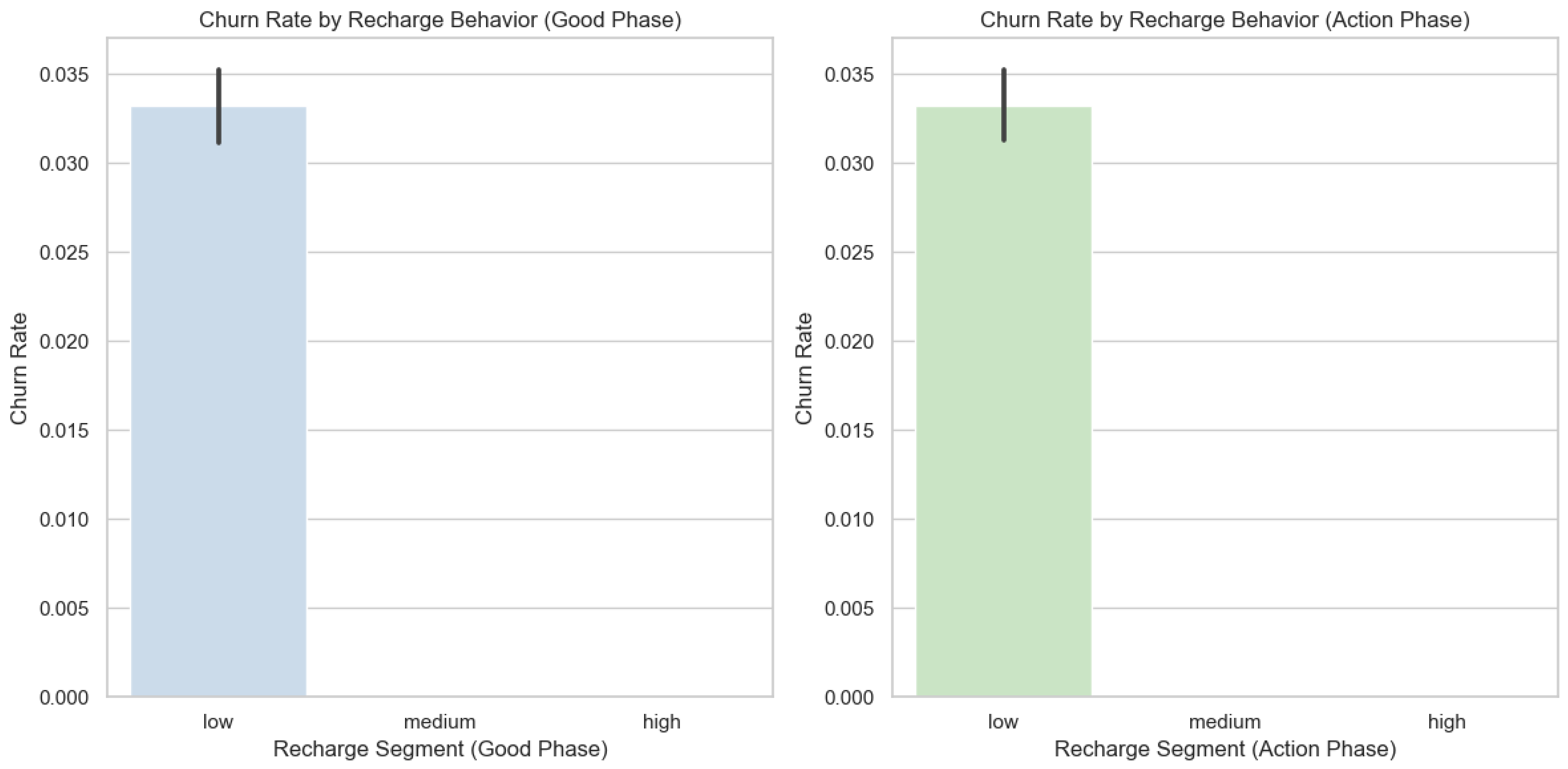

Outgoing MOU Distribution by Churn (Good Phase)
Average Minutes of Usage (MOU)
8000
6000
4000
2000
0
0
1
Churn
Outgoing MOU Distribution by Churn (Action Phase)
Average Minutes of Usage (MOU)
14000
12000
10000
8000
6000
4000
2000
0
0
1
Churn

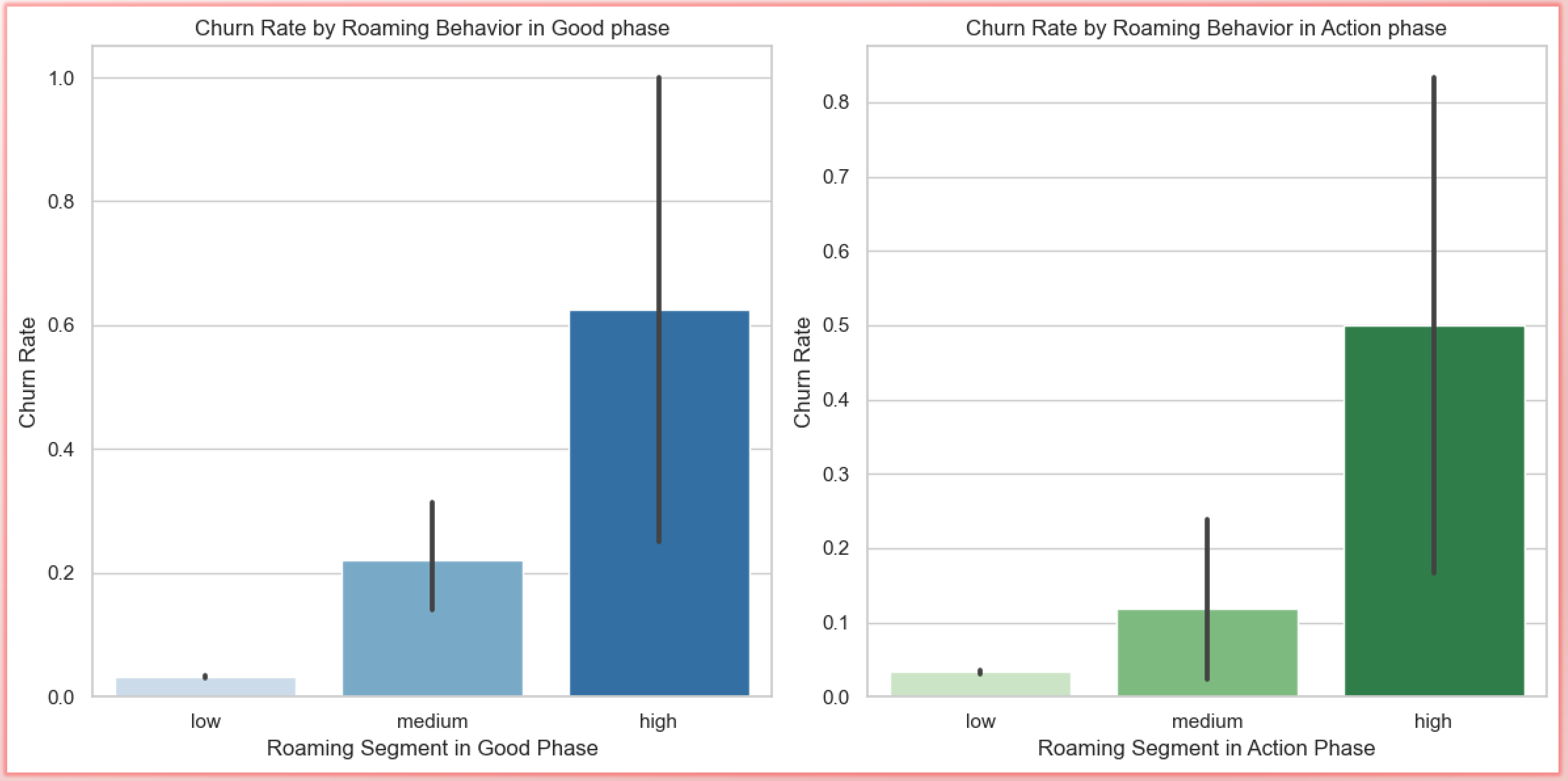

Incoming MOU Distribution by Churn (Good Phase)
Average Minutes of Usage (MOU)
0
1000
2000
3000
4000
5000
6000
0
1
Churn
Incoming MOU Distribution by Churn (Action Phase)
Average Minutes of Usage (MOU)
0
1000
2000
3000
4000
5000
6000
0
1
Churn

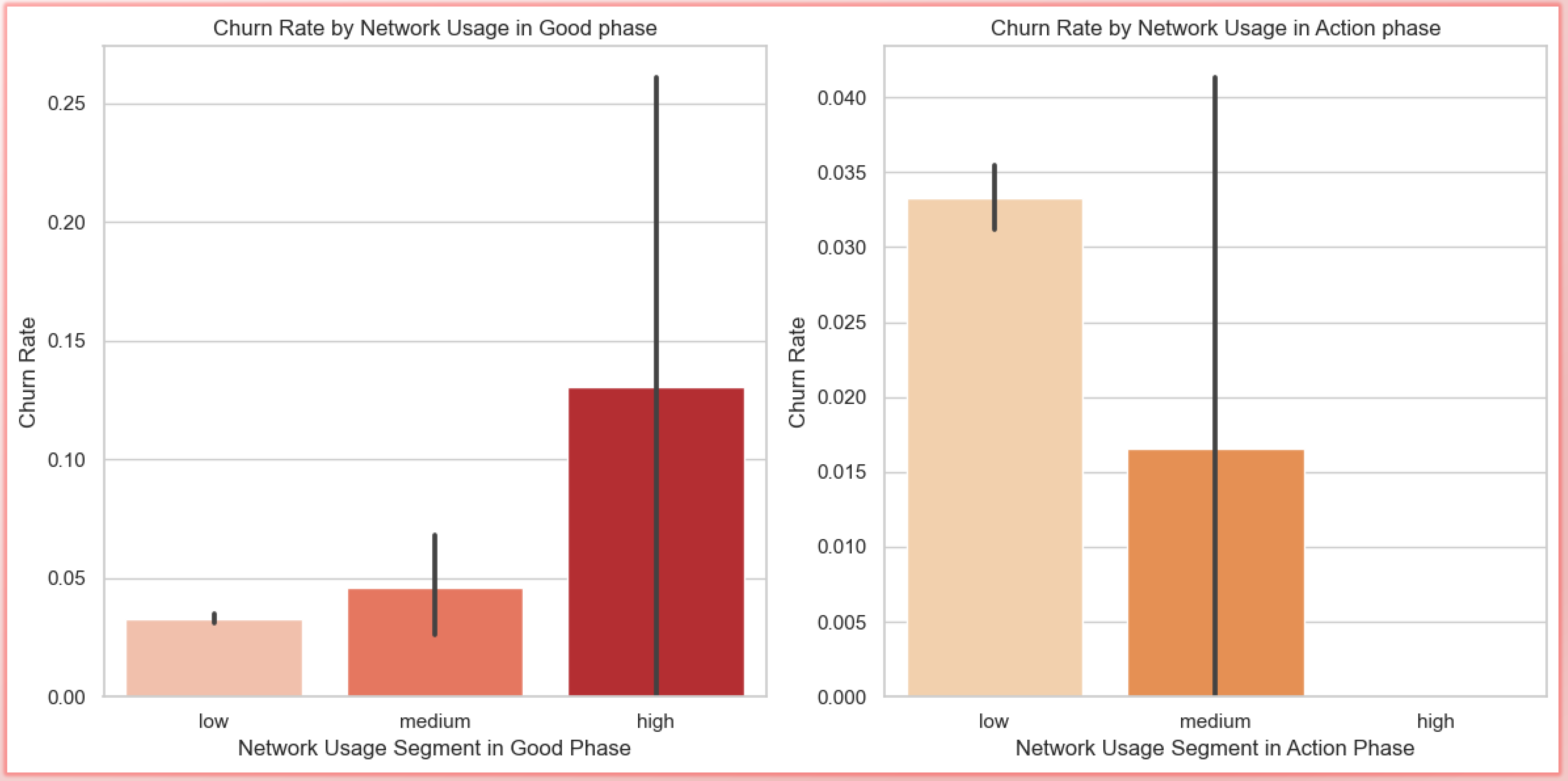

ARPU Distribution by Churn in Good phase
Average Revenue Per User (ARPU)
30000
25000
20000
15000
10000
5000
0
0
1
Churn


ARPU Distribution by Churn in Action phase
Average Revenue Per User (ARPU)
35000
30000
25000
20000
15000
10000
5000
0
0
1
Churn

SEGMENTATION ANALYSIS


Churn Rate by Recharge Behavior (Good Phase)
Churn Rate
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
low
medium
high
Recharge Segment (Good Phase)
Churn Rate by Recharge Behavior (Action Phase)
Churn Rate
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
low
medium
high
Recharge Segment (Action Phase)


Churn Rate by Roaming Behavior in Good phase
Churn Rate
1.0
0.8
0.6
0.4
0.2
0.0
low
medium
high
Roaming Segment in Good Phase
Churn Rate by Roaming Behavior in Action phase
Churn Rate
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
low
medium
high
Roaming Segment in Action Phase


Churn Rate by Network Usage in Good phase
Churn Rate
0.25
0.20
0.15
0.10
0.05
0.00
low
medium
high
Network Usage Segment in Good Phase
Churn Rate by Network Usage in Action phase
Churn Rate
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
low
medium
high
Network Usage Segment in Action Phase

CHECKED HIGHLY CORRELATED VARIABLES

- isd_og_mou_8
- isd_og_mou_7
- total_rech_amt_8
- isd_og_mou_8
- total_rech_amt_6
- total_rech_amt_7
- total_ic_mou_6
- total_ic_mou_8
- total_ic_mou_7
- std_og_t2t_mou_8
- std_og_t2m_mou_7
- std_og_t2m_mou_8
- std_og_t2t_mou_6

These variables were deleted as they were highly correlated.

MODEL BUILDING

Cleaned dataset consists of 26495 rows and 112 columns.

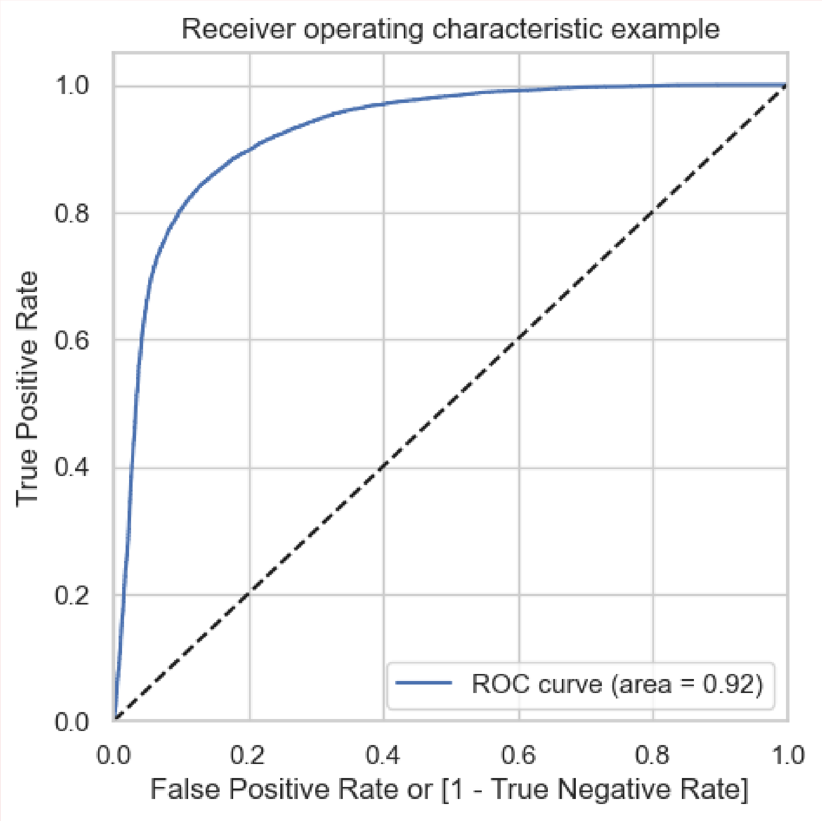
Performed feature scaling on the dataset. Handled class imbalance using SMOTE

Conducted train-test split with a ratio of 80% for training data and 20% for testing data.

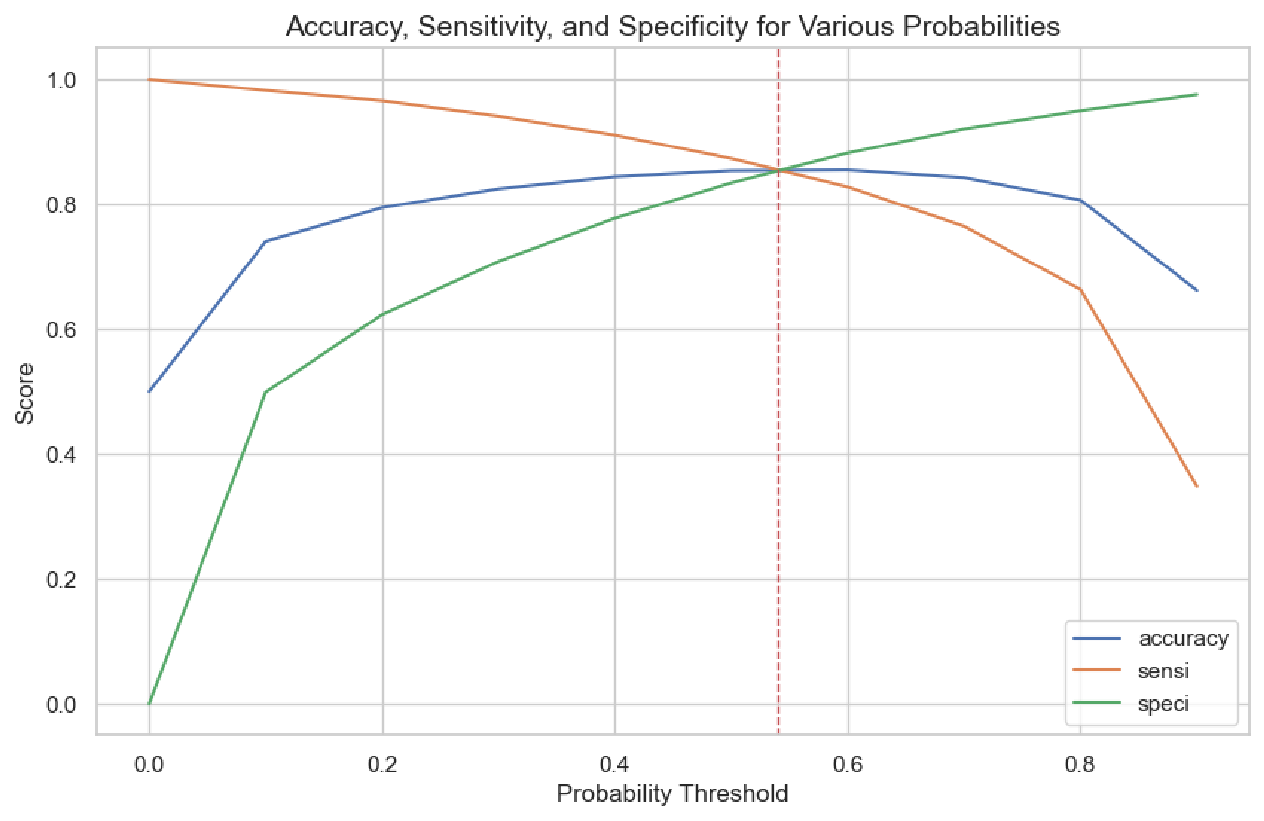
Utilized Logistic Regression model for prediction. Employed Recursive Feature Elimination (RFE) technique with 20 variables as output for feature selection.

Finalized model iteratively, removing features with p-values greater than 0.05 and Variance Inflation Factor (VIF) greater than 5 for Testing data.

ROC CURVE



Accuracy, Sensitivity, And Specificity Curves



- With the current cut off as seen in the graph as **0.54**, the **accuracy** is **85.56%**, **sensitivity** is **85.59%** and **specificity** is **85.53%**
- The **precision** is **85.54%** , **Recall** is **85.59%** and **F1** score is **85.56%**.

PREDICTIONS ON THE TEST DATA

- Performed scaling on the required variables as part of preprocessing.
- Trained the test set using the final model derived from the feature selection process.
- Utilized the trained model to make predictions on the test data.
- Employed a cutoff value of 0.54 for classification.
- Achieved an **Accuracy** of **86.36%** on the test data.
- **Sensitivity**, representing the true positive rate, stood at **85.64%**.
- **Specificity**, representing the true negative rate, was measured at **86.38%**.

RECOMMENDATIONS

- **Improve Network Coverage:** Invest in improving network coverage and quality, especially for local and roaming calls, as these factors significantly influence churn.
- **Roaming Services Enhancement:** Enhance roaming services to provide a seamless experience for customers traveling outside their home network area.
- **Recharge Offers:** Design attractive recharge offers and schemes, focusing on the last day recharge amounts and the total number of recharges, to retain customers and incentivize usage.
- **Service Scheme Customization:** Offer personalized service schemes with validity equivalent to a month for both 2G and 3G networks, targeting specific customer segments based on usage patterns.
- **Data Usage Optimization:** Provide value-added services and data usage optimization tips to customers, considering factors like mobile internet usage volume for 2G network in August.
- **Customer Engagement:** Implement proactive customer engagement strategies, such as loyalty programs and personalized communication, to address customer needs and concerns effectively.


Thank you