TELECOM CHURN CASE STUDY

Author – Piyush Gupta, Santanu Das, Nakul Jain

INTRODUCTION

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, customer retention has now become even more important than customer acquisition.

For many incumbent operators, retaining high profitable customers is the number one business goal.

To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.

In this project, we will analyze customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

Understanding and defining churn

- There are two main models of payment in the telecom industry - postpaid (customers pay a monthly/annual bill after using the services) and prepaid (customers pay/recharge with a certain amount in advance and then use the services).

- In the postpaid model, when customers want to switch to another operator, they usually inform the existing operator to terminate the services, and you directly know that this is an instance of churn.

- However, in the prepaid model, customers who want to switch to another network can simply stop using the services without any notice, and it is hard to know whether someone has actually churned or is simply not using the services temporarily (e.g. someone may be on a trip abroad for a month or two and then intend to resume using the services again).

- Thus, churn prediction is usually more critical (and non-trivial) for prepaid customers, and the term 'churn' should be defined carefully. Also, prepaid is the most common model in India and Southeast Asia, while postpaid is more common in Europe in North America.

- This project is based on the Indian and Southeast Asian market.

Definitions of churn

There are various ways to define churn, such as:

- **Revenue-based churn:** Customers who have not utilized any revenue-generating facilities such as mobile internet, outgoing calls, SMS etc. over a given period of time. One could also use aggregate metrics such as 'customers who have generated less than INR 4 per month in total/average/median revenue.

The main shortcoming of this definition is that there are customers who only receive calls/SMSes from their wage-earning counterparts, i.e. they don't generate revenue but use the services. For example, many users in rural areas only receive calls from their wage-earning siblings in urban areas.

- **Usage-based churn:** Customers who have not done any usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time.
- A potential shortcoming of this definition is that when the customer has stopped using the services for a while, it may be too late to take any corrective actions to retain them. For e.g., if you define churn based on a 'two-months zero usage' period, predicting churn could be useless since by that time the customer would have already switched to another operator.

- In this project, we will use the usage-based definition to define churn.

Understanding the business objective and the data

- The dataset contains customer-level information for a span of four consecutive months - June, July, August and September. The months are encoded as 6, 7, 8 and 9, respectively.

- 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 behavior during churn will be helpful.

Data Set

- The data dictionary contains meanings of abbreviations. Some frequent ones are loc (local), IC (incoming), OG (outgoing), T2T (telecom operator to telecom operator), T2O (telecom operator to another operator), RECH (recharge) etc.

- The attributes containing 6, 7, 8, 9 as suffixes imply that those correspond to the months 6, 7, 8, 9 respectively.

.

Understanding Data Points

Acronyms	Descriptions
MOBILE_NUMBER	Customer phone number
CIRCLE_ID	Telecom circle area to which the customer belongs to
LOC	Local calls - within same telecom circle
STD	STD calls - outside the calling circle
IC	Incoming calls
OG	Outgoing calls
T2T	Operator T to T, i.e. within same operator (mobile to mobile)
T2M	Operator T to other operator mobile
T2O	Operator T to other operator fixed line
T2F	Operator T to fixed lines of T
T2C	Operator T to it's own call center
ARPU	Average revenue per user
MOU	Minutes of usage - voice calls
AON	Age on network - number of days the customer is using the operator T network
ONNET	All kind of calls within the same operator network
OFFNET	All kind of calls outside the operator T network
ROAM	Indicates that customer is in roaming zone during the call
SPL	Special calls
ISD	ISD calls
RECH	Recharge

Acronyms	Descriptions
NUM	Number
AMT	Amount in local currency
MAX	Maximum
DATA	Mobile internet
3G	3G network
AV	Average
VOL	Mobile internet usage volume (in MB)
2G	2G network
PCK	Prepaid service schemes called - PACKS
NIGHT	Scheme to use during specific night hours only
MONTHLY	Service schemes with validity equivalent to a month
SACHET	Service schemes with validity smaller than a month
*.6	KPI for the month of June
*.7	KPI for the month of July
*.8	KPI for the month of August
*.9	KPI for the month of September
FB_USER	Service scheme to avail services of Facebook and similar social networking sites
VBC	Volume based cost - when no specific scheme is not purchased and paid as per usage

Data preparation

The following data preparation steps are crucial for this problem:

1. **Filter high-value customers**

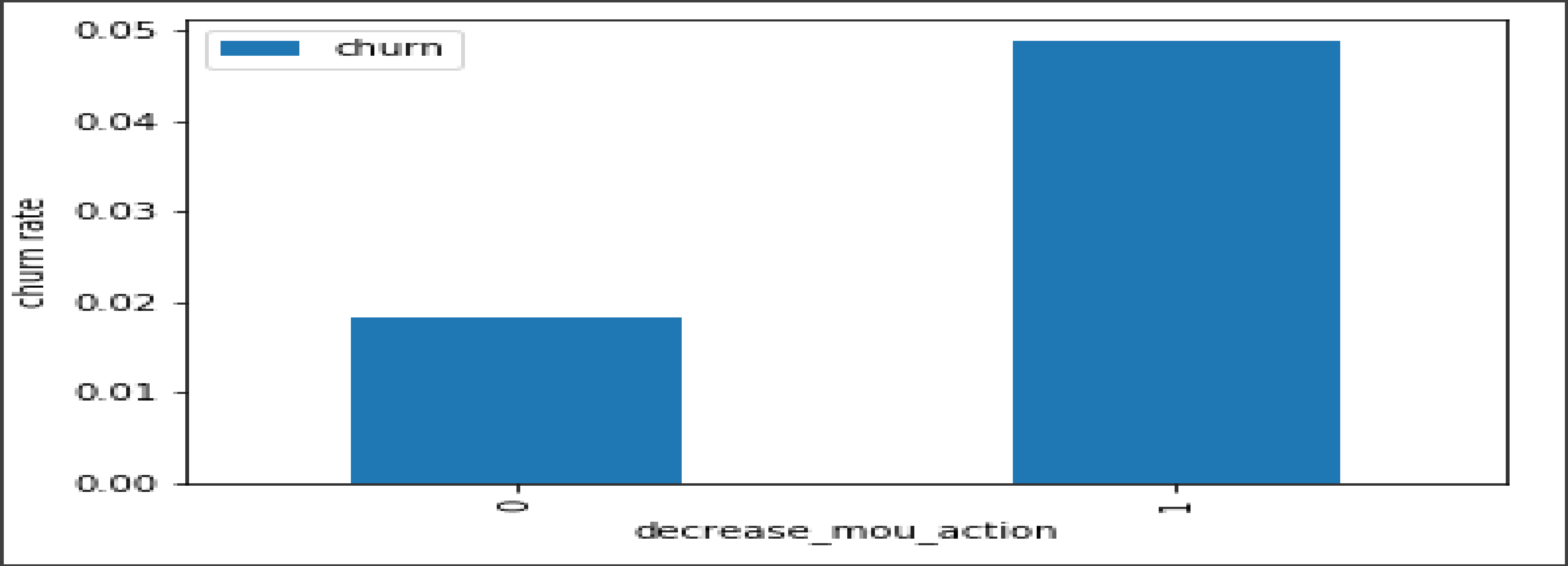
- As mentioned above, you need to predict churn only for high-value customers. Define high-value customers as follows: Those who have recharged with an amount more than or equal to X, where X is the 70th percentile of the average recharge amount in the first two months (the good phase).
- After filtering the high-value customers, you should get about 30k rows.

2. **Tag churners and remove attributes of the churn phase**

- 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
-

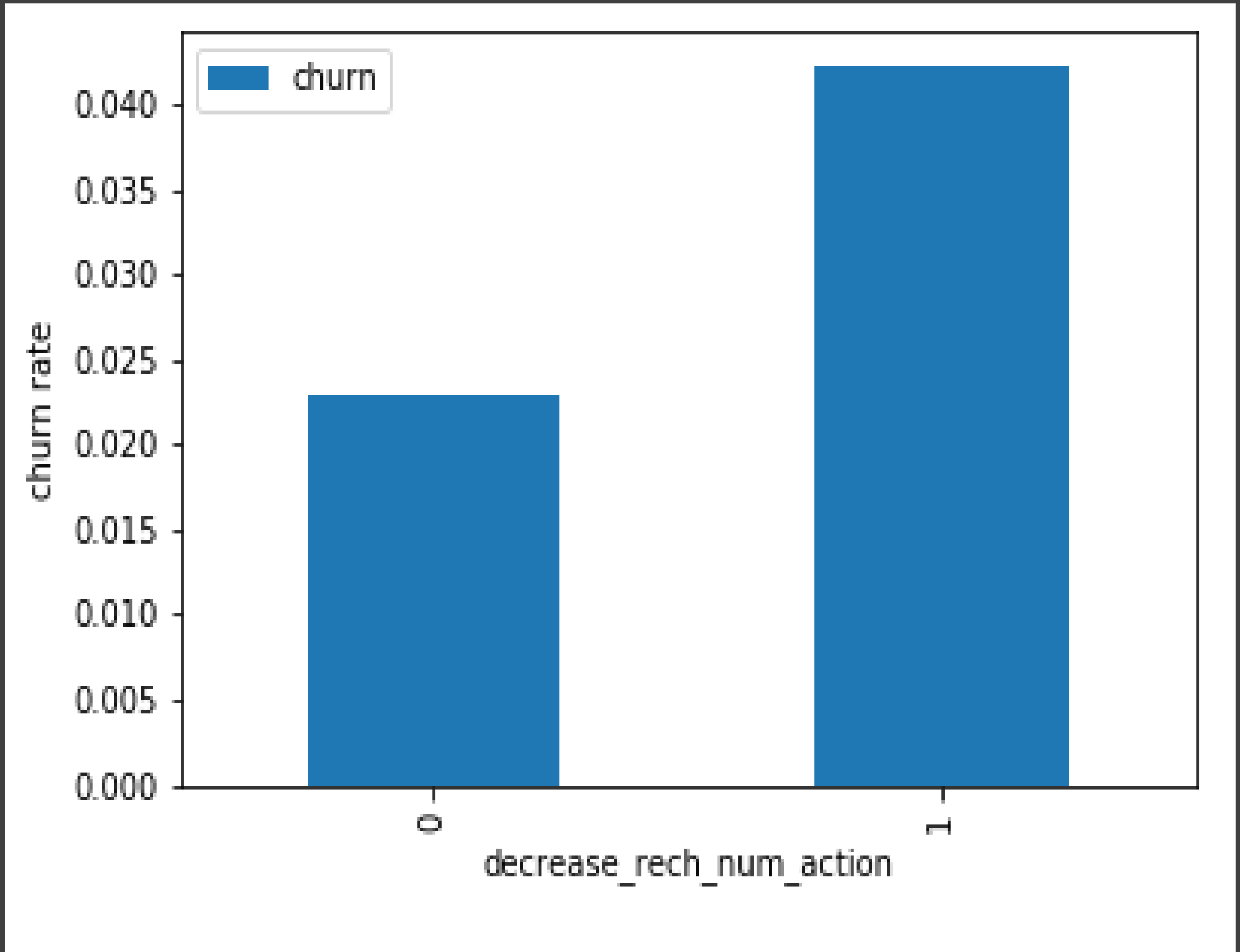
  After tagging churners, remove all the attributes corresponding to the churn phase (all attributes having ‘ _9’, etc. in their names).

Analysis



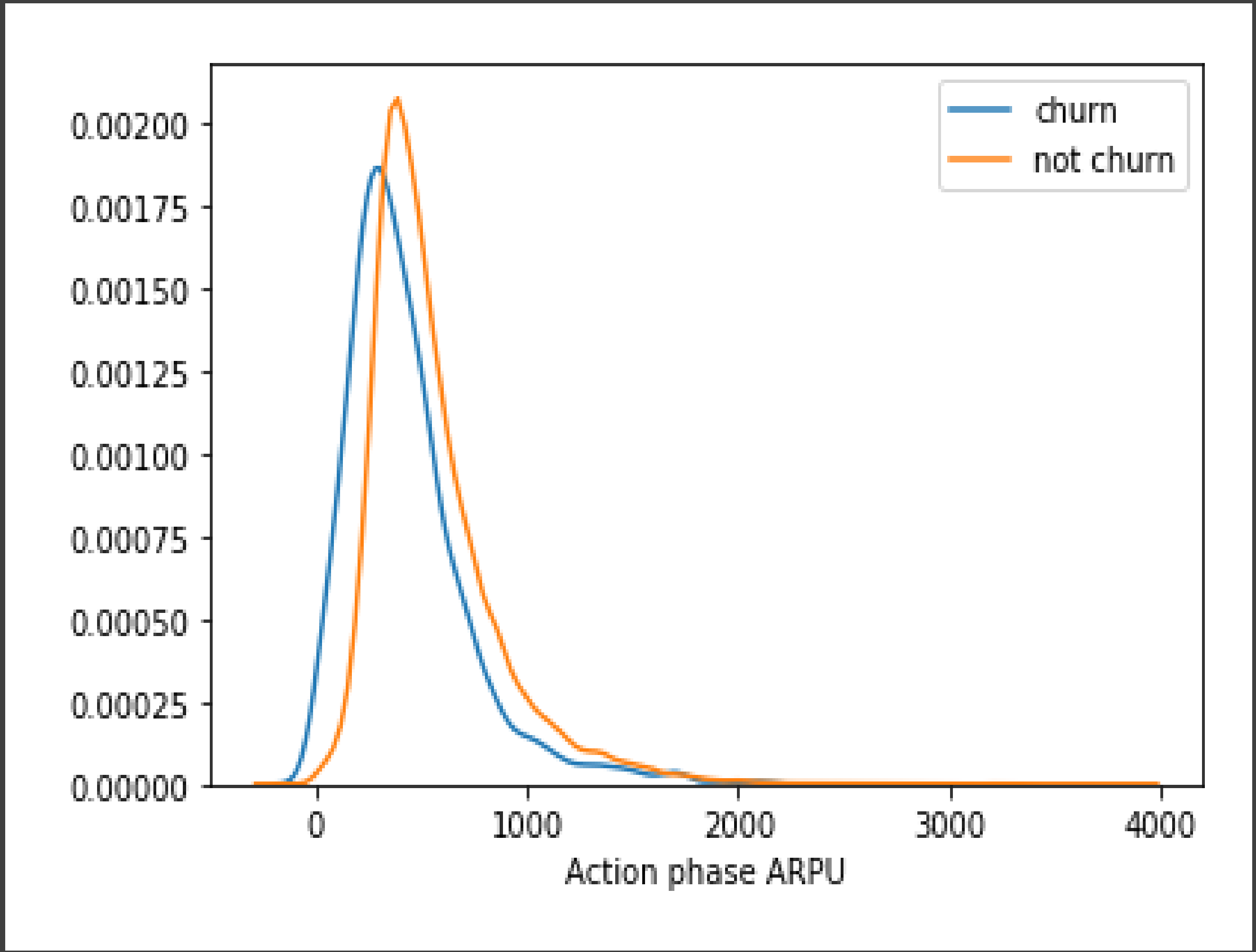
- We can see that the churn rate is more for the customers, whose minutes of usage(mou) decreased in the action phase than the good phase.

Analysis – AMT_ANNUITY Variable



- As expected, the churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase.

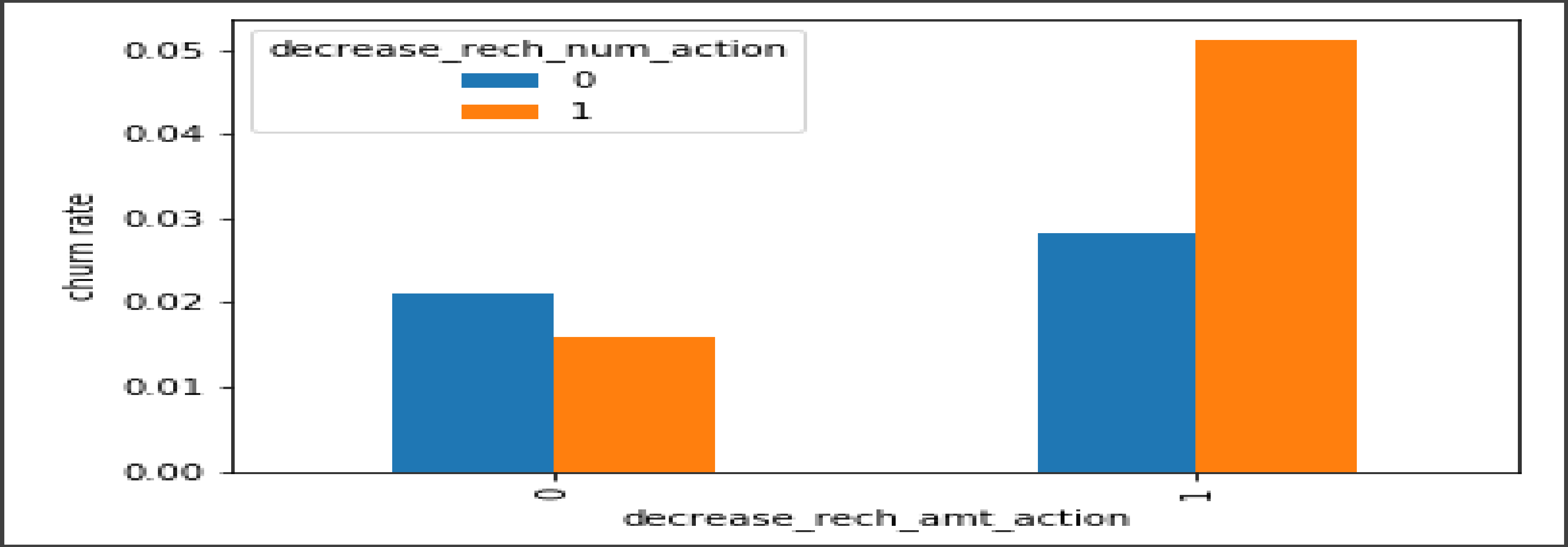
Analysis of the average revenue per customer (churn and not churn) in the action phase



- Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned.

- ARPU for the not churned customers is mostly densed on the 0 to 1000.

Analysis of churn rate by the decreasing recharge amount and number of recharge in the action phase



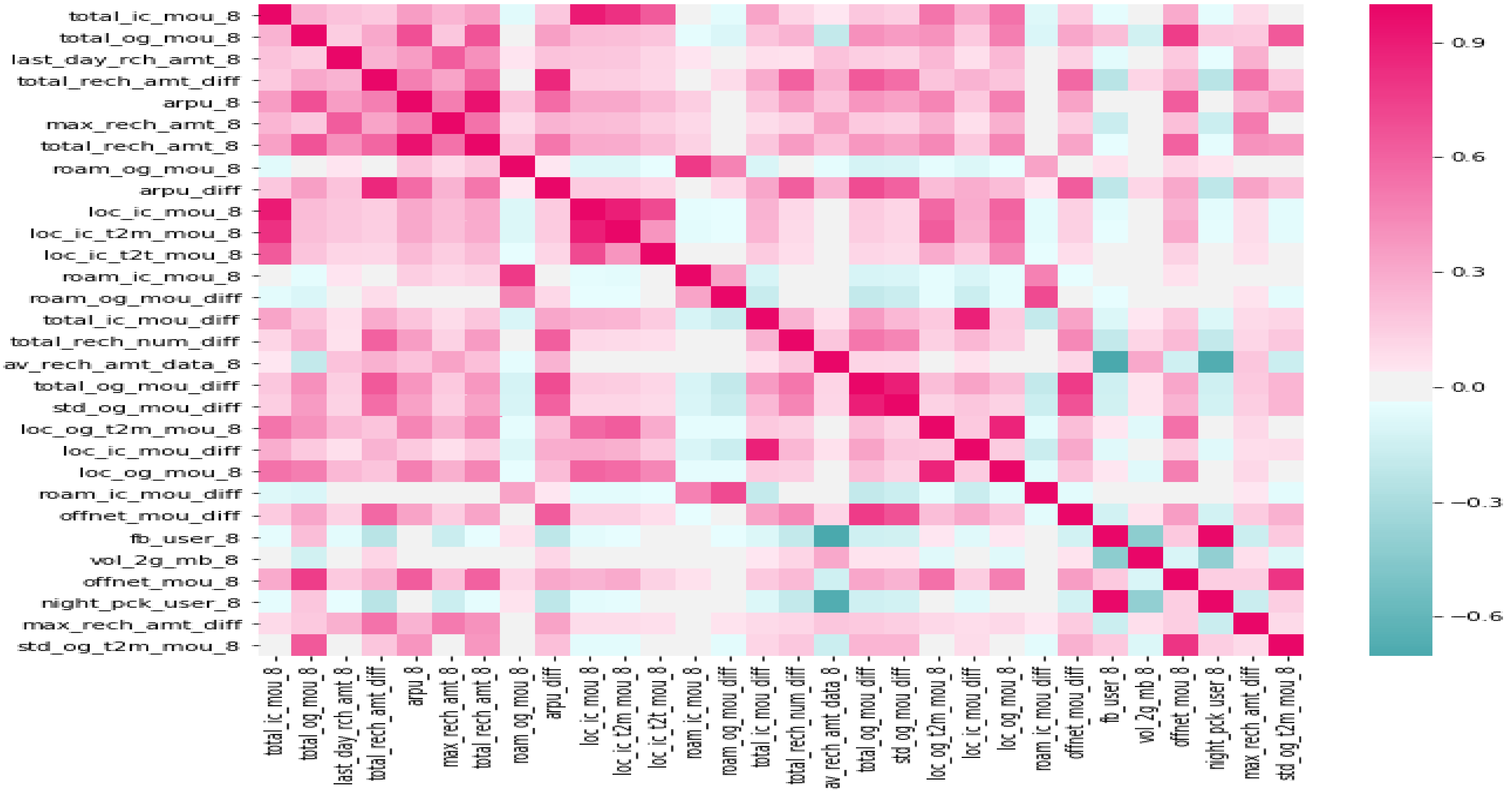
We can see from the above plot, that the churn rate is more for the customers, whose recharge amount as well as number of recharge have decreased in the action phase than the good phase.

Evaluate on test data

```
# predict churn on test data
y_pred = pipeline.predict(X_test)

# create onfusion matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)

# check sensitivity and specificity
sensitivity, specificity, _ = sensitivity_specificity_support(y_test, y_pred, average='binary')
print("Sensitivity: \t", round(sensitivity, 2), "\n", "Specificity: \t", round(specificity, 2), sep='')

# check area under curve
y_pred_prob = pipeline.predict_proba(X_test)[:, 1]
print("AUC:    \t", round(roc_auc_score(y_test, y_pred_prob),2))
```

```
[[5614 1277]
 [  95  515]]
Sensitivity:    0.84
Specificity:    0.81
AUC:             0.9
```

Feature Correlation



Extract the intercept and the coefficients from the logistic model

In [92]:
```
logistic_model = model.best_estimator_.named_steps['logistic']
```

In [93]:
```
# intercept
intercept_df = pd.DataFrame(logistic_model.intercept_.reshape((1,1)), columns = ['intercept'])
```

In [94]:
```
# coefficients
coefficients = logistic_model.coef_.reshape((9, 1)).tolist()
coefficients = [val for sublist in coefficients for val in sublist]
coefficients = [round(coefficient, 3) for coefficient in coefficients]

logistic_features = list(X_train.columns)
coefficients_df = pd.DataFrame(logistic_model.coef_, columns=logistic_features)
```

In [95]:
```
# concatenate dataframes
coefficients = pd.concat([intercept_df, coefficients_df], axis=1)
coefficients
```

Out[95]:

	intercept	total_ic_mou_8	total_rech_amt_diff	total_og_mou_8	arpu_8	roam_ic_mou_8	roam_og_mou_8	std_ic_mou_8	av_rech_amt_data_8	std_og_mou_8
0	-1.46829	-1.227629	-0.68268	-1.067155	0.169269	0.042355	0.194784	0.075685	-0.794502	0.56011

Conclusion and Recommendations:

- Total Incoming Minutes of Usage (total_ic_mou_8): To encourage higher volumes of incoming calls, businesses should consider implementing strategies such as attractive call packages or promotional offers.

- STD Outgoing Minutes of Usage in August: The observation that customers making more STD outgoing calls in August tend to demonstrate higher value or engagement with the business suggests an opportunity for leveraging this insight. Businesses can offer targeted STD calling plans or special promotions during this period to foster customer engagement.

- STD Incoming Minutes of Usage in July (std_ic_mou_7): The finding that customers receiving more STD incoming calls during July exhibit higher value or engagement highlights an area where the business can focus on enhancing customer experience. This can be achieved through improvements in call quality, reliable connections, and attractive STD incoming call packages.

- Roaming Outgoing Minutes of Usage in July (roam_og_mou_7): The observation that customers making more outgoing calls while in roaming during July tend to be more valuable presents an opportunity for businesses to provide special roaming offers, affordable international calling rates, or tailored packages. These measures can cater to the needs of customers who frequently travel or utilize roaming services.

- On-net Minutes of Usage in June (onnet_mou_6): Businesses should explore ways to encourage customers to diversify their calling patterns or explore other services within the business.

- STD Outgoing Minutes of Usage to Fixed Lines in August (std_og_t2f_mou_7): Analyzing strategies to improve the value proposition for customers making STD calls to fixed lines is recommended. This can involve offering bundled packages or discounted rates for these specific types of calls.
- ISD Incoming Minutes of Usage in July : The finding that customers receiving more ISD incoming calls during July tend to have higher value or engagement suggests an opportunity to offer attractive ISD incoming call plans, competitive international calling rates, or promotions catering to customers with international connections.
- Local Outgoing Minutes of Usage to Other Operator Mobiles in July (loc_og_t2m_mou_7): Businesses should consider strategies to enhance the attractiveness of their local outgoing call offerings to customers using other operator mobiles. This can be achieved through competitive pricing, value-added services, or targeted marketing campaigns.
- Special Outgoing Minutes of Usage in July (spl_og_mou_7): Evaluating the nature and purpose of these special outgoing calls and identifying ways to enhance their value or relevance for customers is recommended. Introducing customized special call packages, exclusive features, or discounts can encourage customers to make more special outgoing calls.
- 2G Sachet Recharge in July (sachet_2g_7): The finding that customers opting for 2G sachet recharges during July tend to exhibit lower value or engagement suggests a need for analysis of the underlying reasons. Businesses may consider promoting higher-speed data services or encouraging customers to upgrade to more advanced data plans.

In order to address the aforementioned issues, it is advisable for the telecom company to pay careful attention to roaming rates and provide enticing offers to customers using services from roaming zones. Additionally, the company should focus on reviewing and potentially adjusting STD and ISD rates, possibly by introducing STD and ISD packages. To gain deeper insights into these matters, collecting customer query and complaint data would be beneficial, allowing the company to tailor its services according to the needs of customers.

Thank You