Telecom churn

Casestudy

Arun Mishra

Telecom Churn Case Study

Case Study : Telecom Churn Case Study

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

Business Goal

In this project, you will analyse 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.

Data Prepration

Data Preparation

In churn prediction, we assume that there are three phases of customer lifecycle :

- The ‘good’ phase [Month 6 & 7]
- The ‘action’ phase [Month 8]
- The ‘churn’ phase [Month 9]

In this case, since you are working over a four-month window, the first two months are the ‘good’ phase, the third month is the ‘action’ phase, while the fourth month is the ‘churn’ phase.

1. Derive new features

We can see more then **74%** values for **recharge** related data are missing.
We can create new feature as **total_rech_amt_data** using **total_rech_data** and **av_rech_amt_data** to capture amount utilized by customer for data.
Also as the minimum value is 1 we can impute the NA values by 0, Considering there were no recharges done by the custom

2. Filter 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 greater than 70th percentile of the average recharge amount in the first two months (the good phase)

3. Tag churners and remove attributes of the churn phase¶

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).

EDA

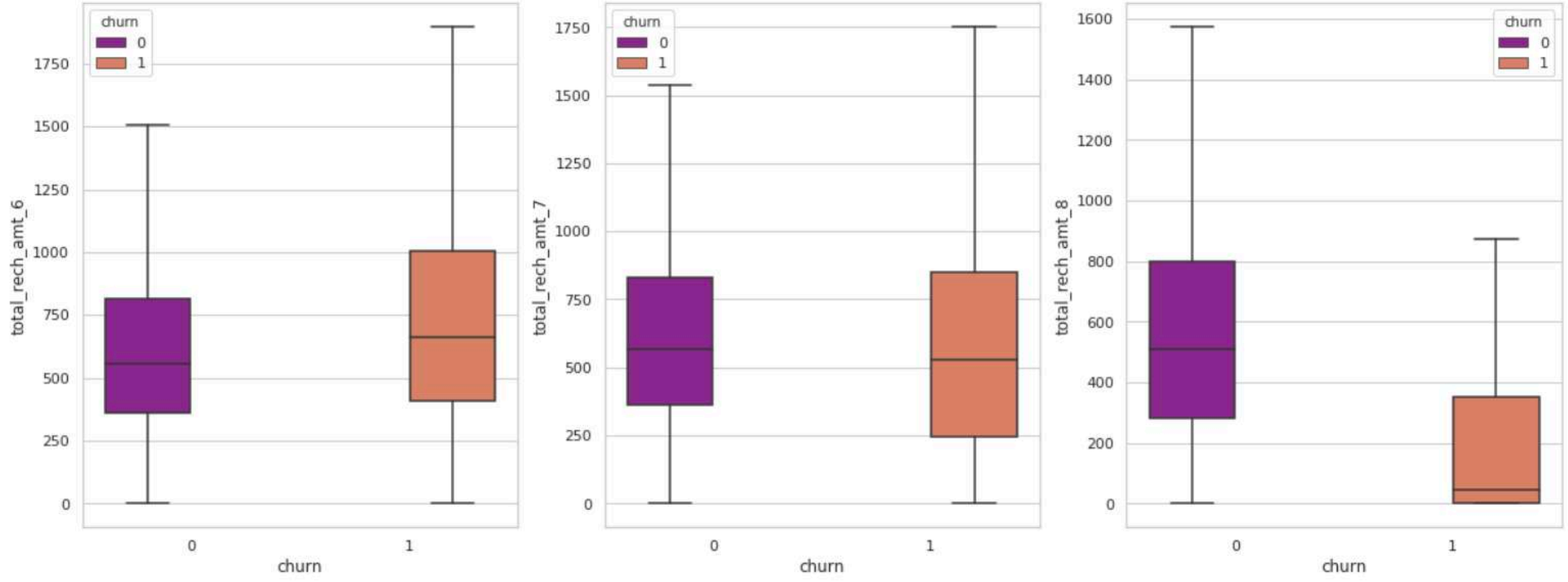
1. Preprocess data (convert columns to appropriate formats, handle missing values, etc.)

Exploring Date field

2. Conduct appropriate exploratory analysis to extract useful insights (whether directly useful for business or for eventual modelling/feature engineering).

a. Recharge amount related varial

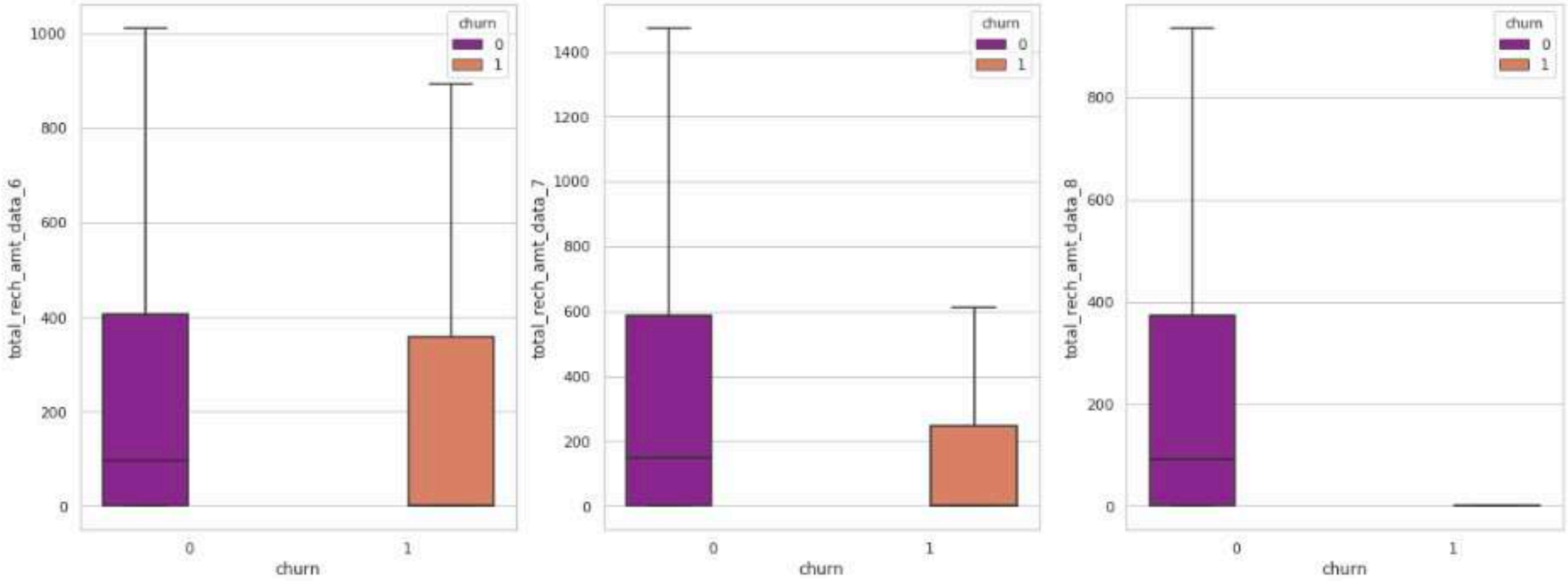
```python
# Ploting for total recharge amount:
plot_box_chart('total_rech_amt')
```



Slide Type

> **Analysis: We can see a drop in the total recharge amount for churned customers in the 8th Month (Action Phase).

We are getting a huge drop in 8th month recharge amount for churned customers.

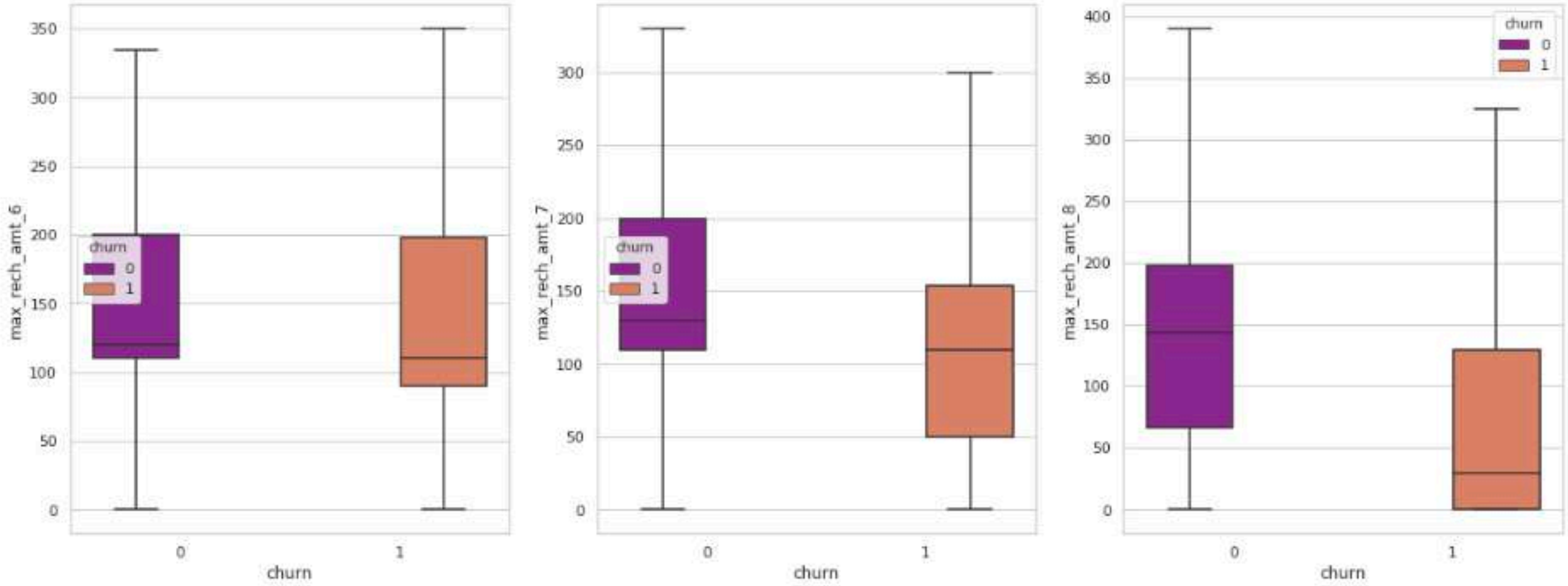


Slide Type

> **Analysis: We can see that there is a huge drop in total recharge amount for data in the 8th month (action phase) for churned customers.

5]: Slide Type

```
# Ploting for maximum recharge amount for data:
plot_box_chart('max_rech_amt')
```



Slide Type

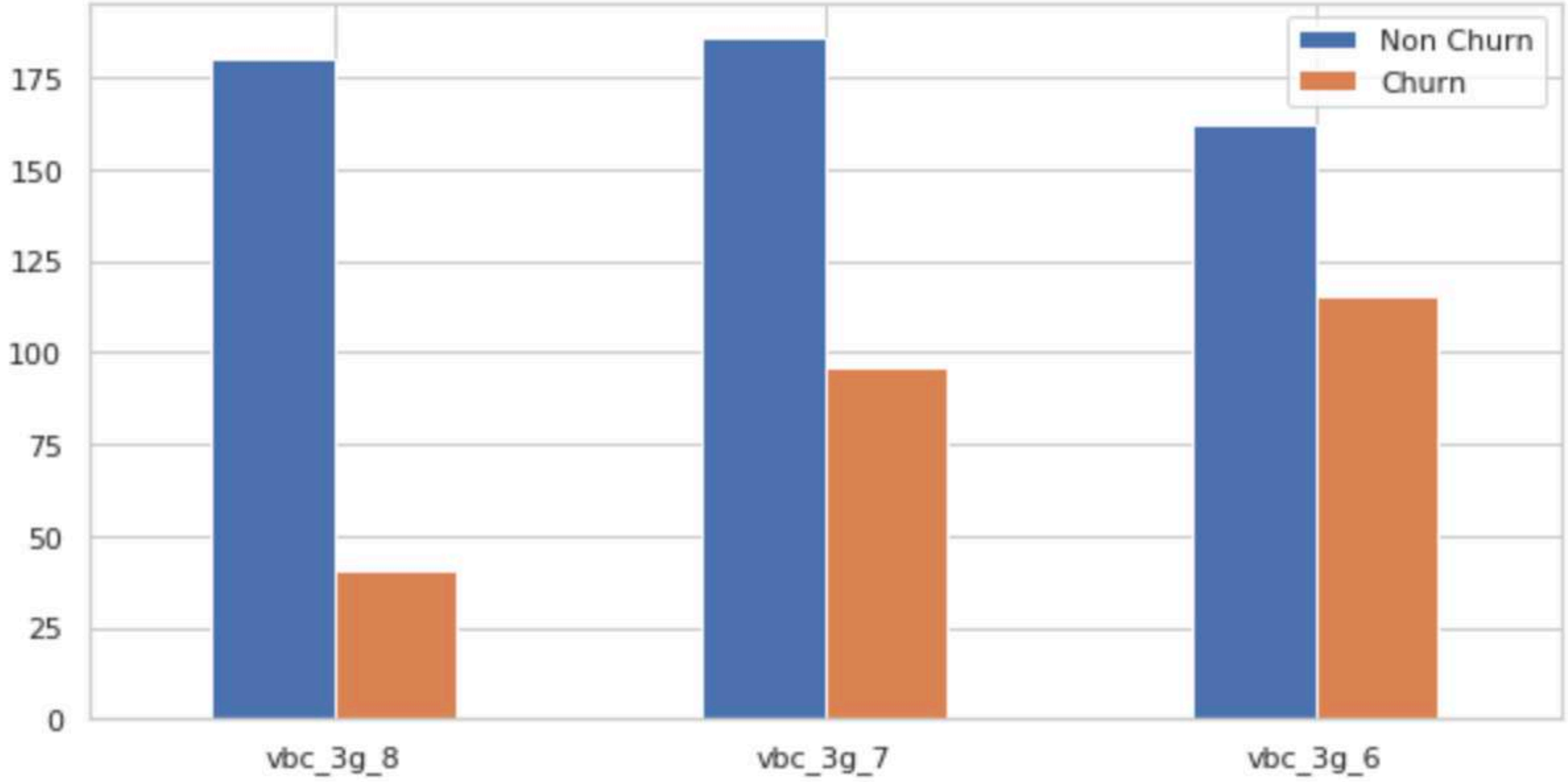
> **Analysis:** We can see that there is a huge drop in maximum recharge amount for data in the 8th month (action phase) for churned customers.

2G and 3G usage related attributes

We are getting more then 40% values are not available for count of recharge and average revenue for 2G/3G per user. Although we have 2G/3G volume usage based data available, we can drop these columns.

1) 2G and 3G usage for churned customers drops in 8th month
2) We also see that 2G/3G usage is higher for non-churned customers indicating that churned customers might be from areas where 2G/3G service is not properly available.

Significantly it showing that volume based cost for 3G is much lower for Churned customers as compared to Non-Churned Customers and also there is a drop in vbc in 8th month



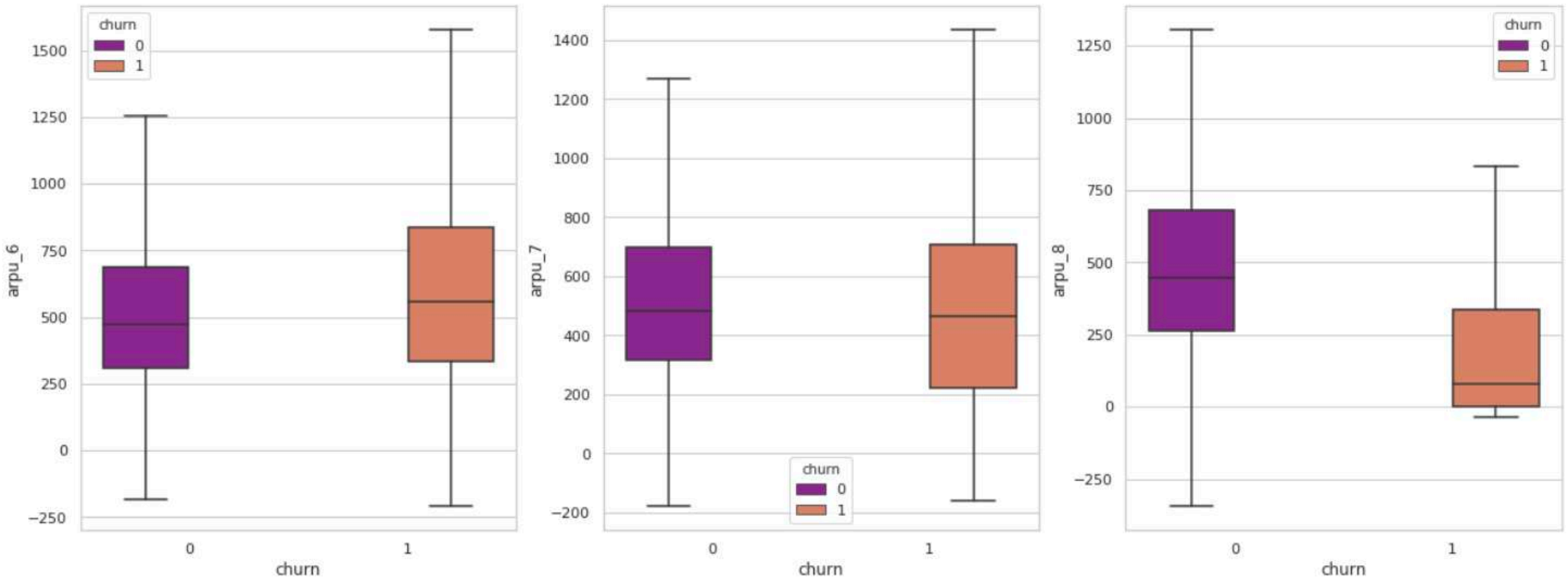
	vbc_3g_8	vbc_3g_7	vbc_3g_6
Non Churn	180.62	186.37	162.56
Churn	40.94	96.34	115.46

Slide Type

Analysis: Significantly it showing that volume based cost for 3G is much lower for Churned customers compared to Non-Churned Customers and also there is a drop in vbc in 8th month

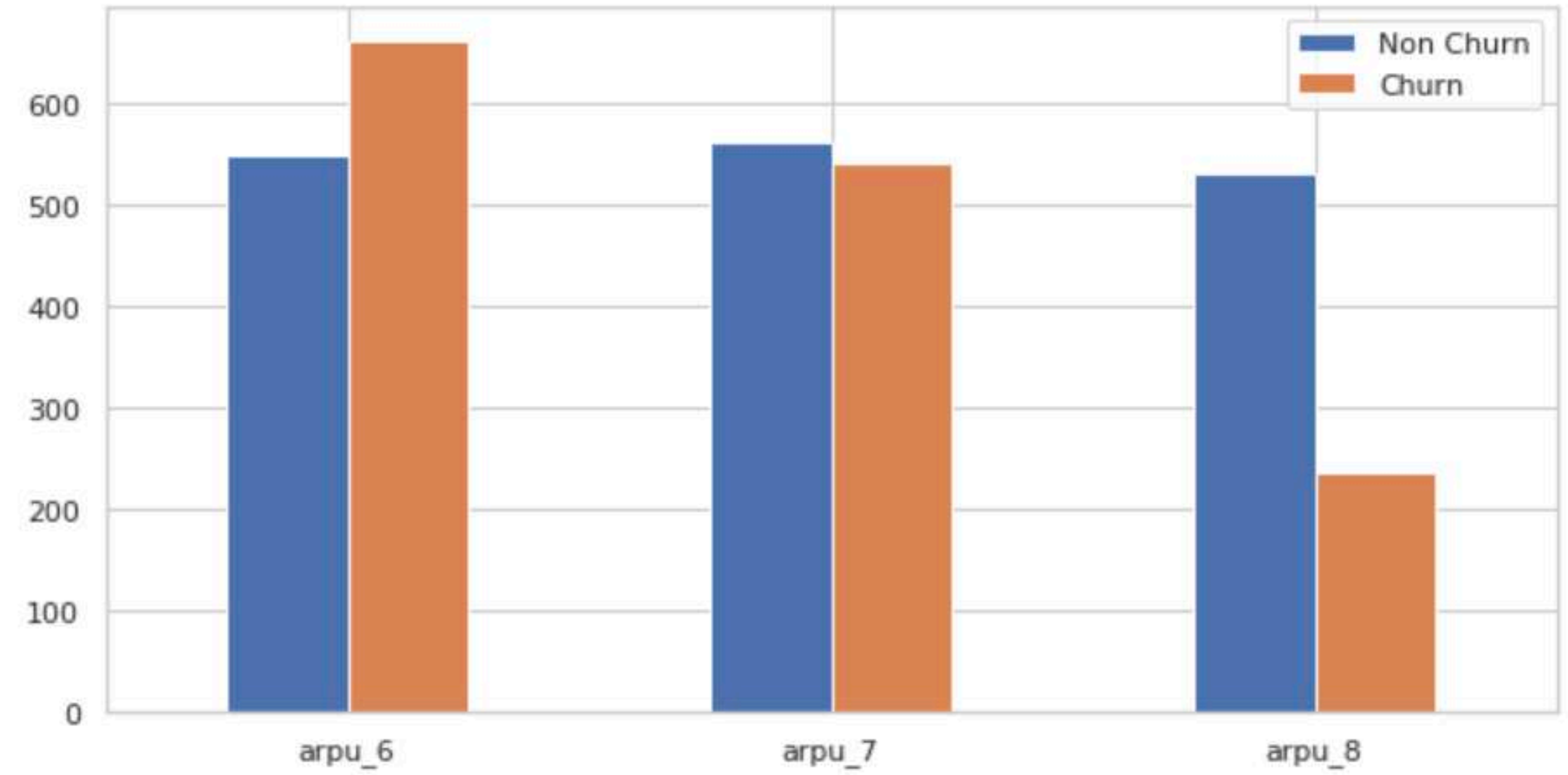
Average Revenue Per User

We can see that huge drops for Arpu in 8th month for churned customers



Slide Type

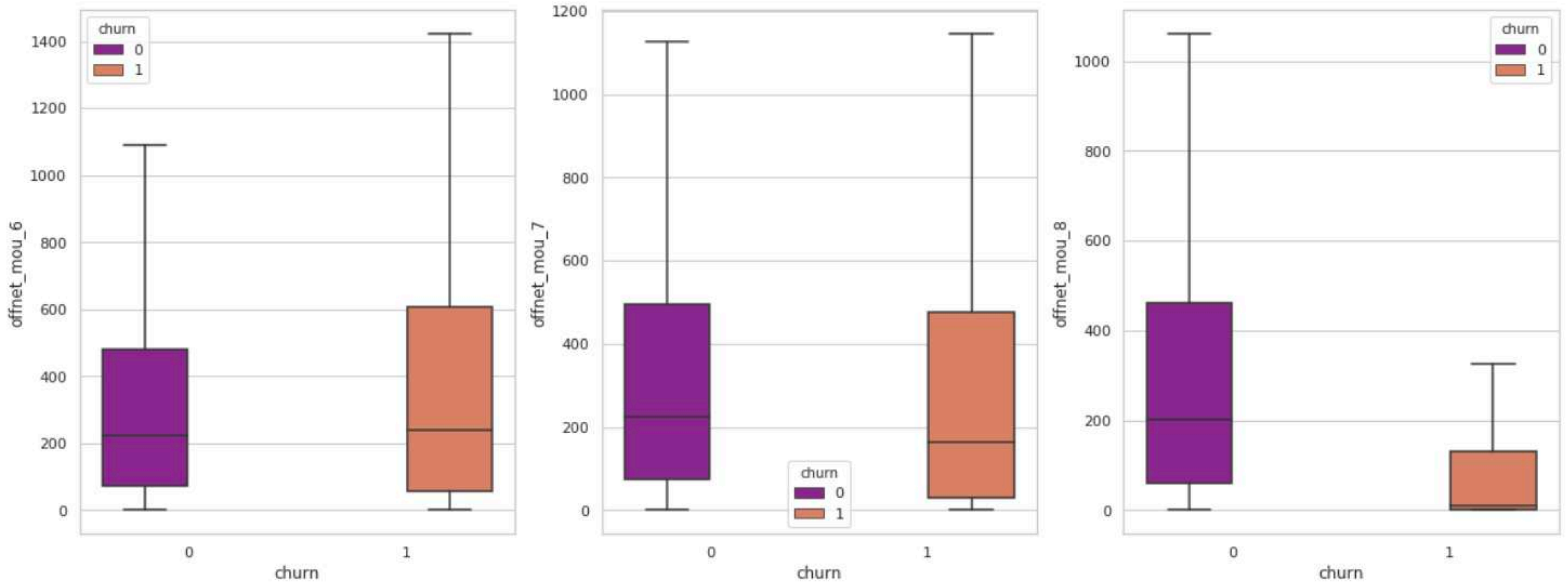
```
plot_mean_bar_chart(telecom_df_high_val_cust, arpu_cols)
```



	arpu_6	arpu_7	arpu_8
Non Churn	549.55	562.93	532.87
Churn	663.71	541.15	237.66

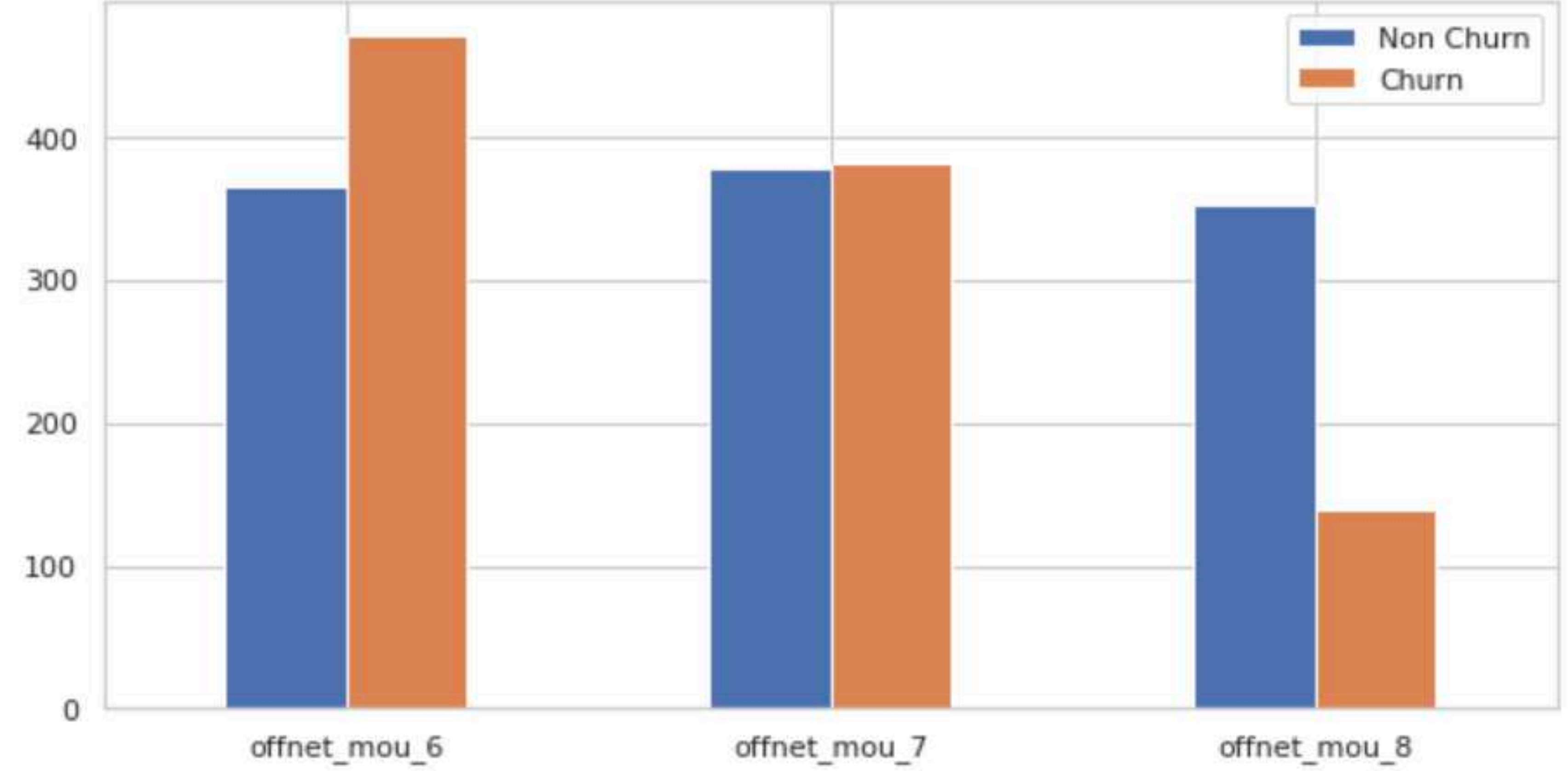
Offnet usage

We can see the drop for offnet mou services in the 8th month



Slide Type

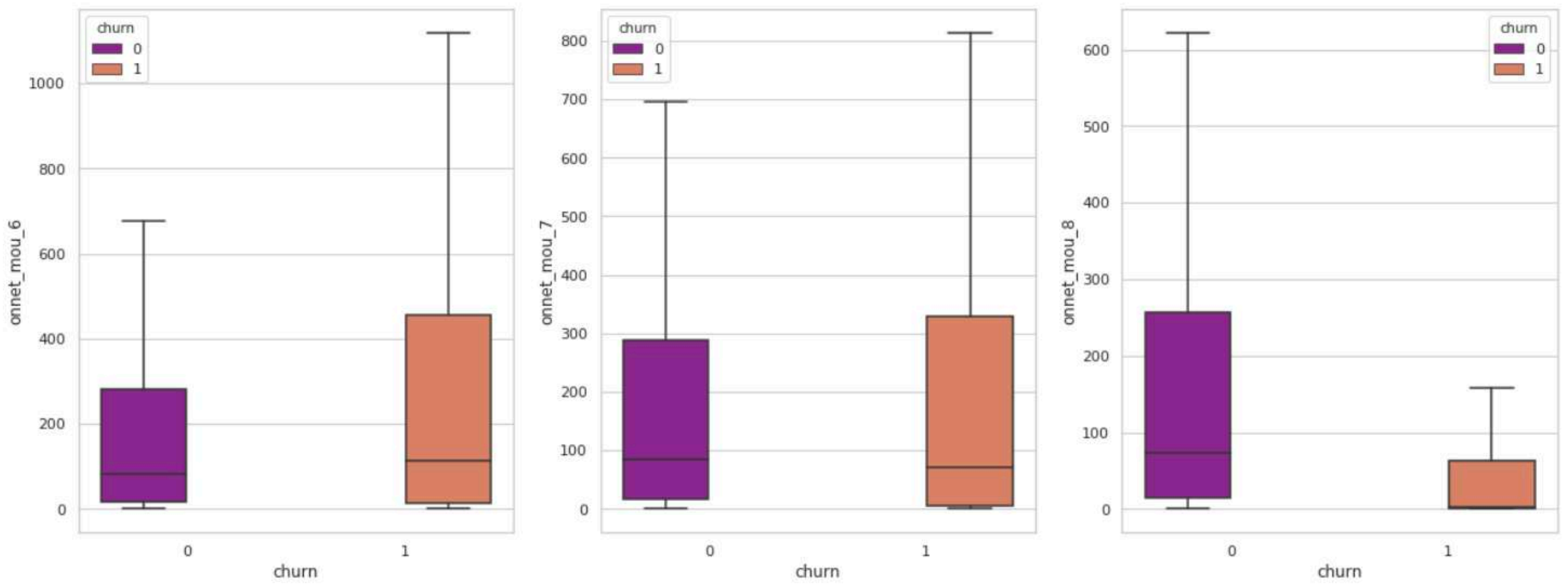
```
plot_mean_bar_chart(telecom_df_high_val_cust, offnet_usage_service_col)
```



	offnet_mou_6	offnet_mou_7	offnet_mou_8
Non Churn	365.12	377.88	352.50
Churn	471.95	382.28	138.52

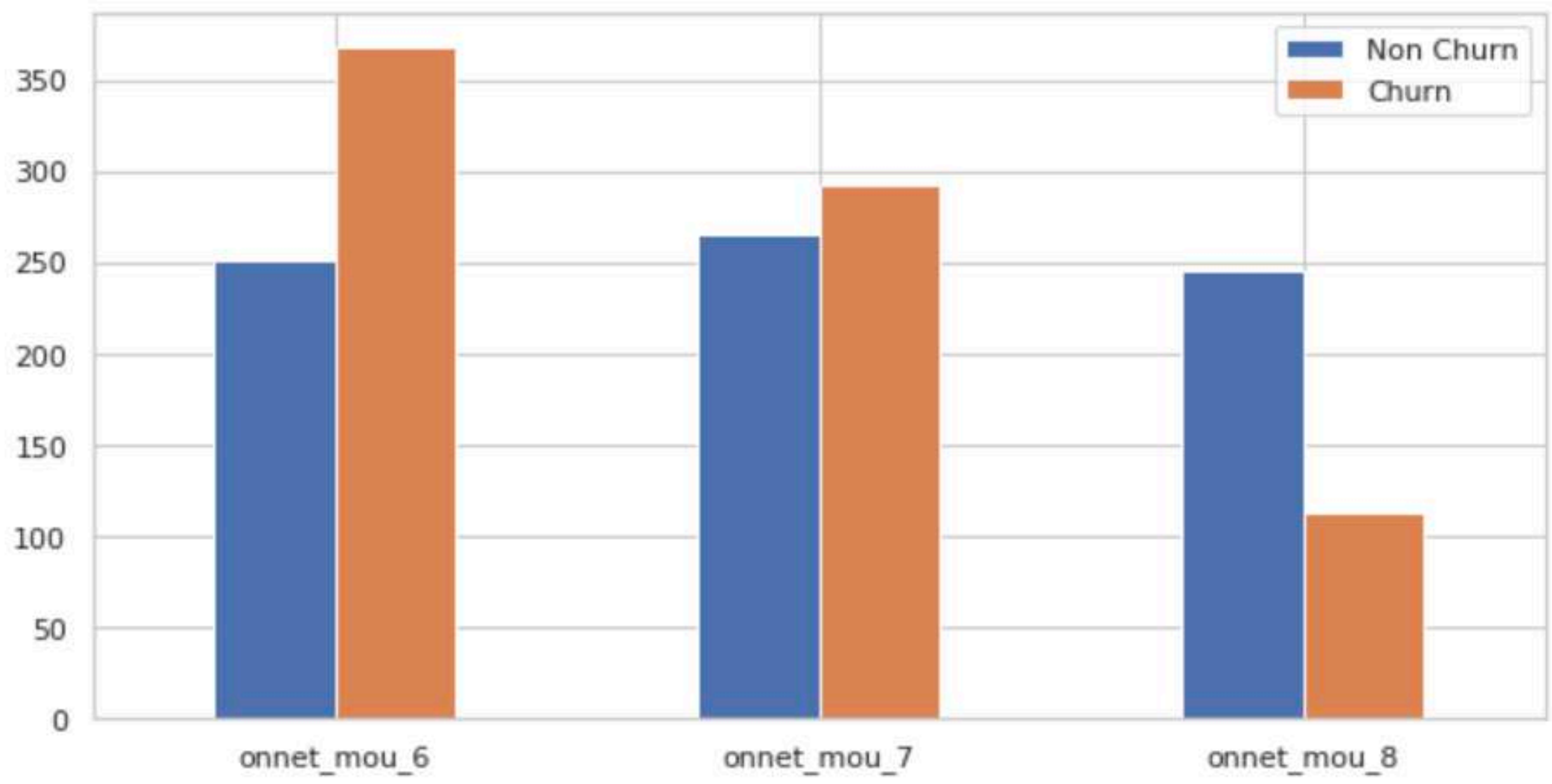
Onnet Usage

We also see that there is a drop in Onnet usage in the 8th month for churned customers



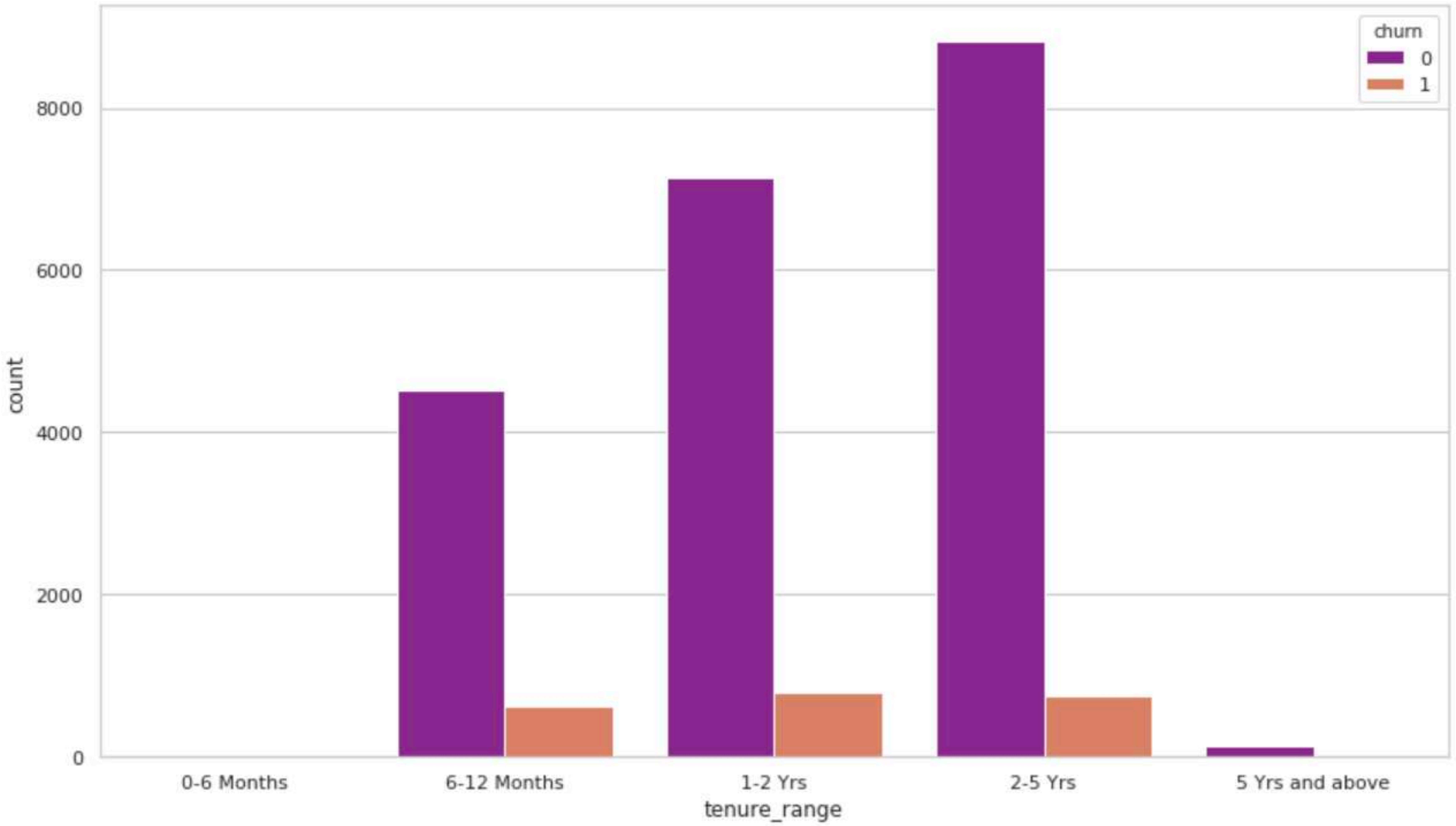
Slide Type

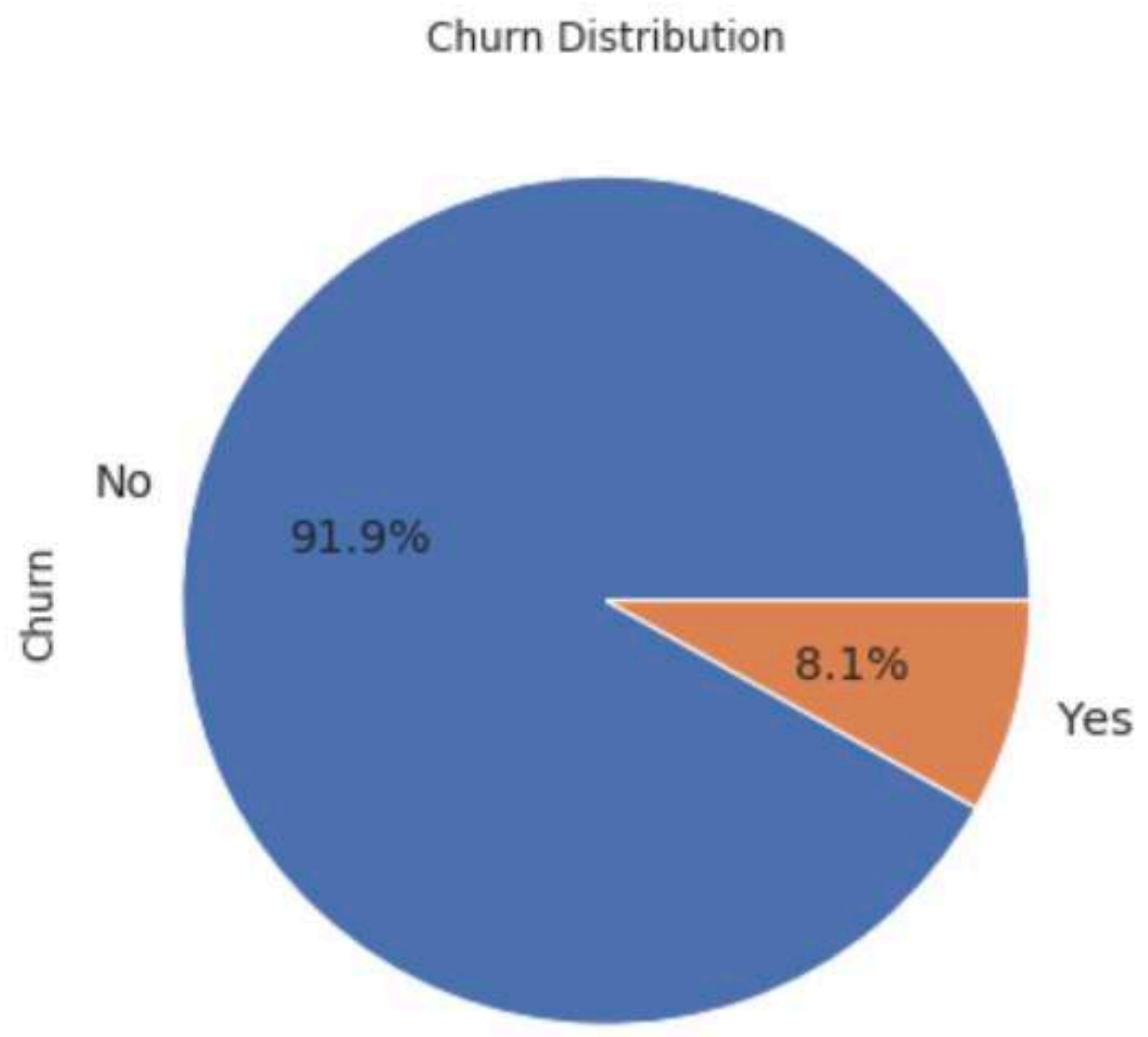
```
plot_mean_bar_chart(telecom_df_high_val_cust, onnet_usage_service)
```

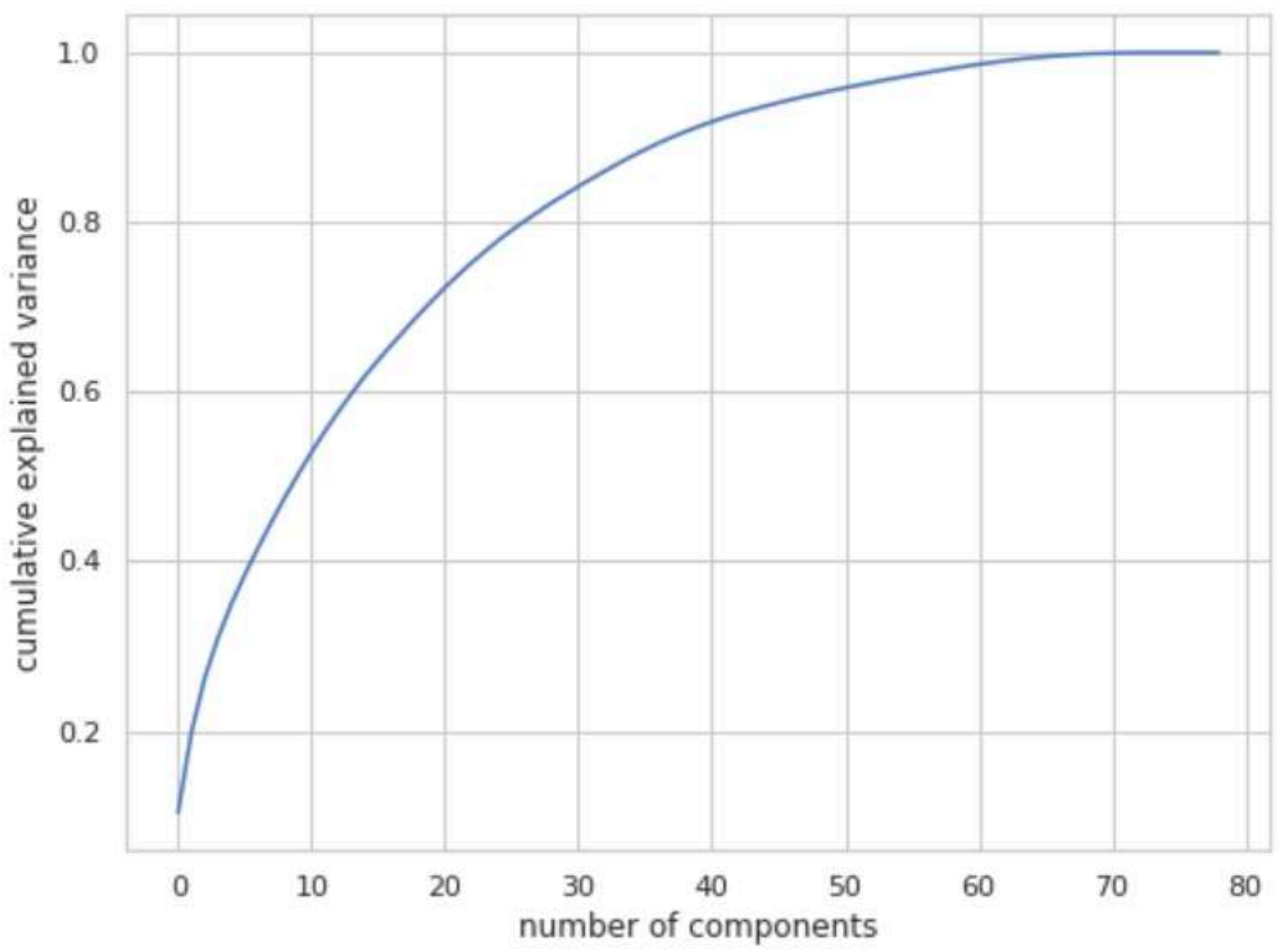


	onnet_mou_6	onnet_mou_7	onnet_mou_8
Non Churn	251.37	265.86	245.03
Churn	368.66	292.85	113.48

Tenure Analysis







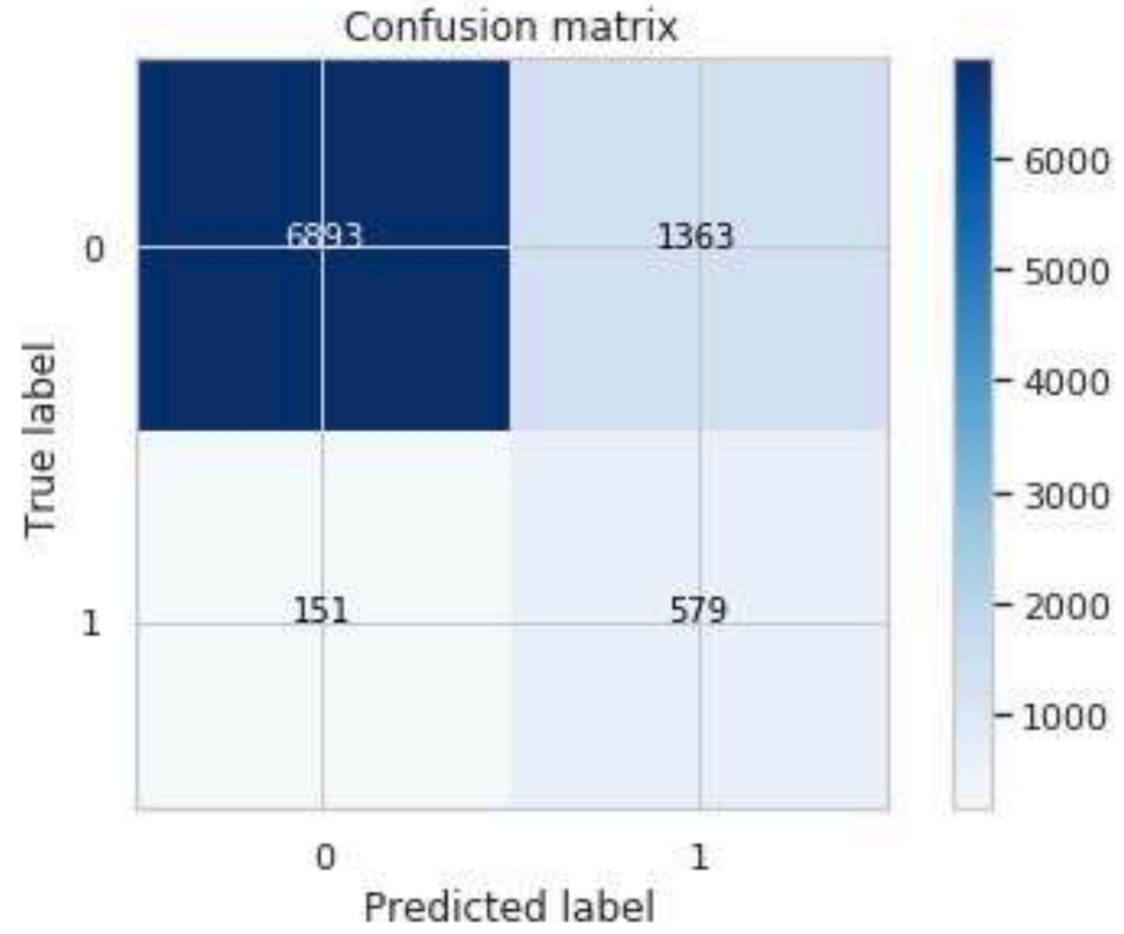
.1]:

	PC1	PC2	PC3	Feature
0	0.00	0.29	0.15	arpu_8
1	-0.11	0.11	0.15	onnet_mou_8
2	-0.10	0.20	0.11	offnet_mou_8
3	-0.01	0.05	0.01	roam_ic_mou_8
4	-0.02	0.07	0.03	roam_og_mou_8
5	-0.03	0.10	0.02	loc_og_t2t_mou_8
6	-0.03	0.18	-0.03	loc_og_t2m_mou_8
7	0.01	0.11	-0.09	loc_og_t2f_mou_8
8	-0.01	0.02	0.08	loc_og_t2c_mou_8
9	-0.11	0.07	0.16	std_og_t2t_mou_8

SMV Regression Modelling

Confusion Matrix

The non-linear model gives approx. 87% accuracy. Thus, going forward, let's choose hyperparameters corresponding to non-linear models.

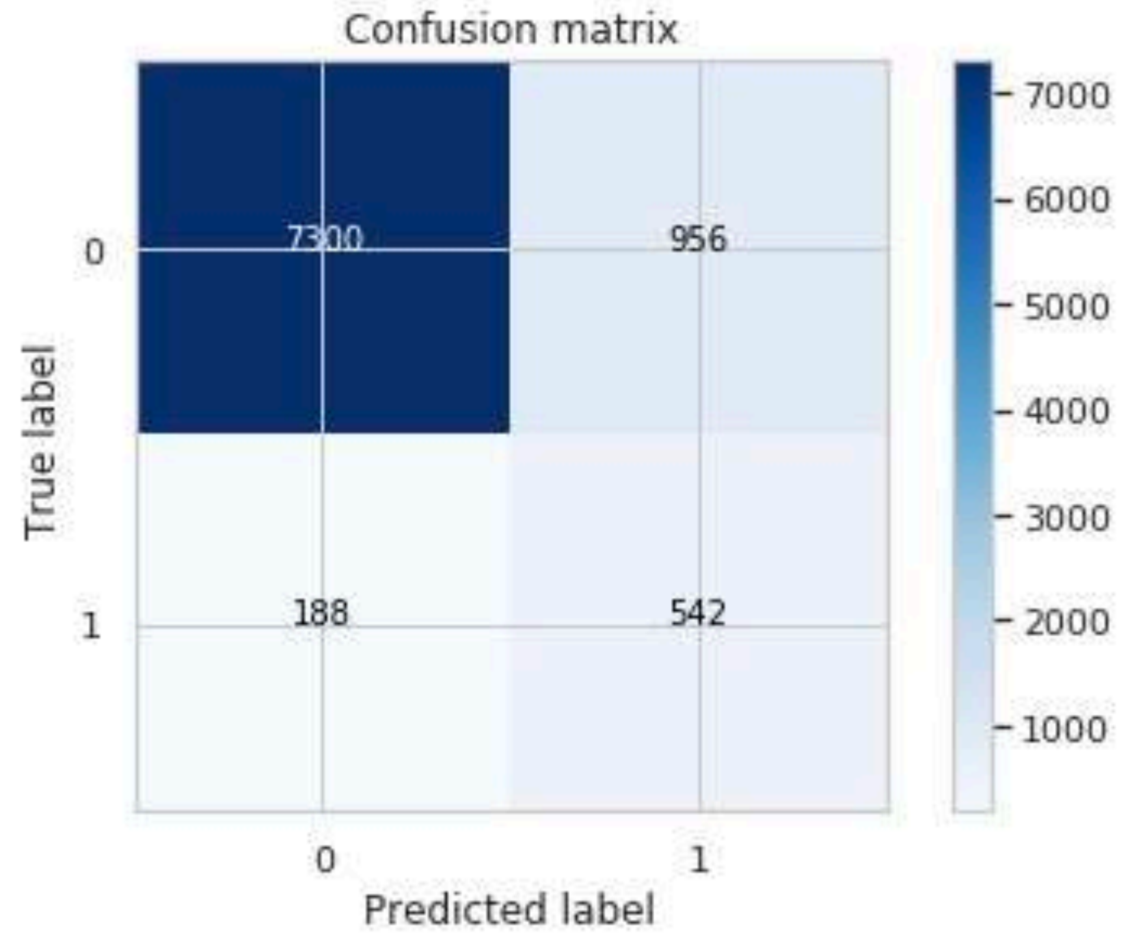



```
get_svm_model_stats(x_train,x_test, y_train, y_test,"rbf")
```

```
<IPython.core.display.Markdown object>

<IPython.core.display.Markdown object>

<IPython.core.display.Markdown object>
```



Recommendations

Business Insights

Less number of high value customer are churing but for last 6 month no new high valued cusotmer is onboarded which is concerning and company should concentrate on that aspect.

Customers with less than 4 years of tenure are more likely to churn and company should concentrate more on that segment by rolling out new schems to that group.

Average revenue per user seems to be most important feature in determining churn prediction.

Incoming and Outgoing Calls on romaing for 8th month are strong indicators of churn behaviour

Local Outgoing calls made to landline , fixedline , mobile and call center provides a strong indicator of churn behaviour.

Better 2G/3G area coverage where 2G/3G services are not good, it's strong indicator of churn behaviour.