

TELECOM CUSTOMER CHURN PREDICTION



&



International
Institute of Information
Technology Bangalore

Data Science Program – August 2022

Domain Oriented **Case Study**

By -

1. Shekhar More
2. Kaveri Deotkar
3. Avik Kumar Saha

PROBLEM STATEMENT

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.



BUSINESS OBJECTIVES

In this project, our business objective is to 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. Thus, our focus would be on

- I. Retaining high profitable customers.
- II. Predicting which customers are at high risk of churn (in order to devise customer retention strategies accordingly).



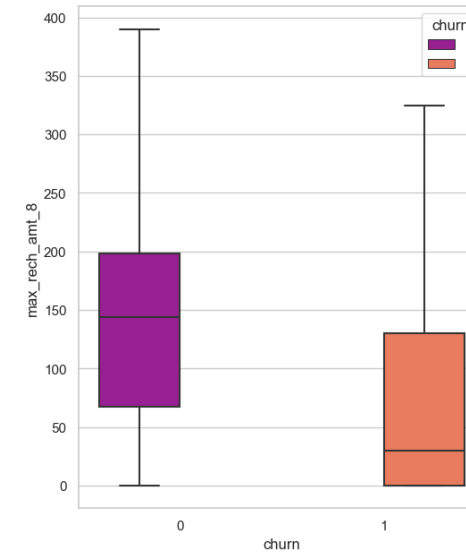
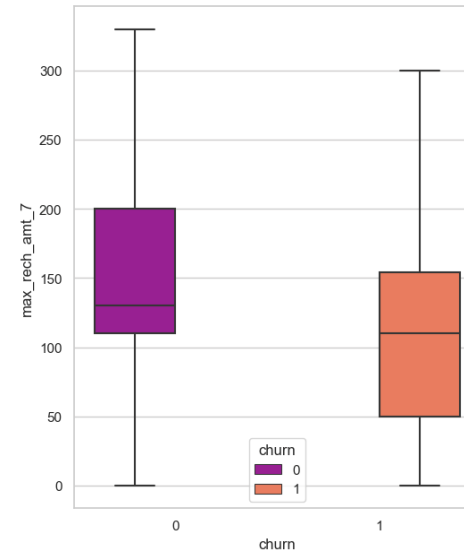
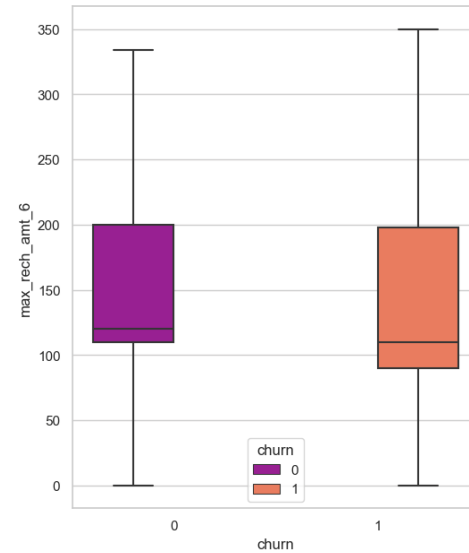
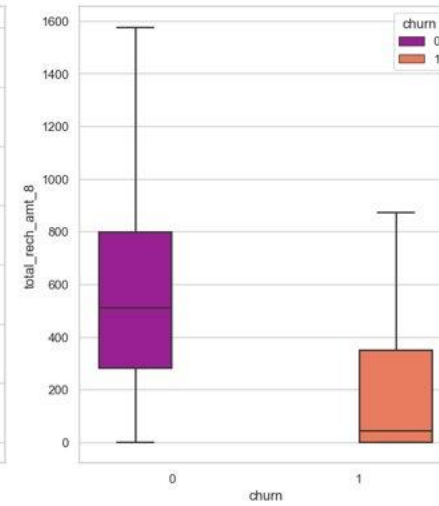
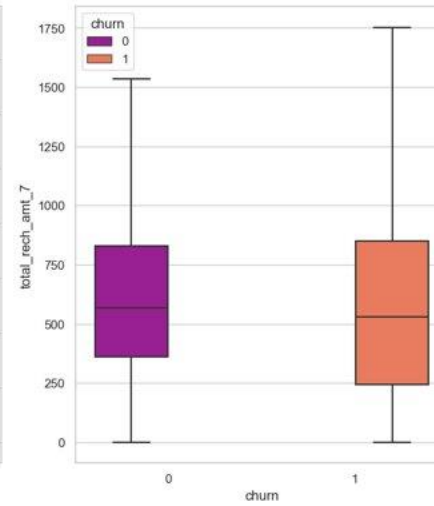
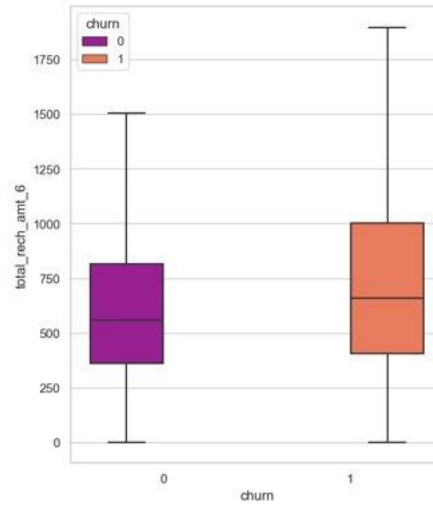
STEPS TO REACH OBJECTIVES

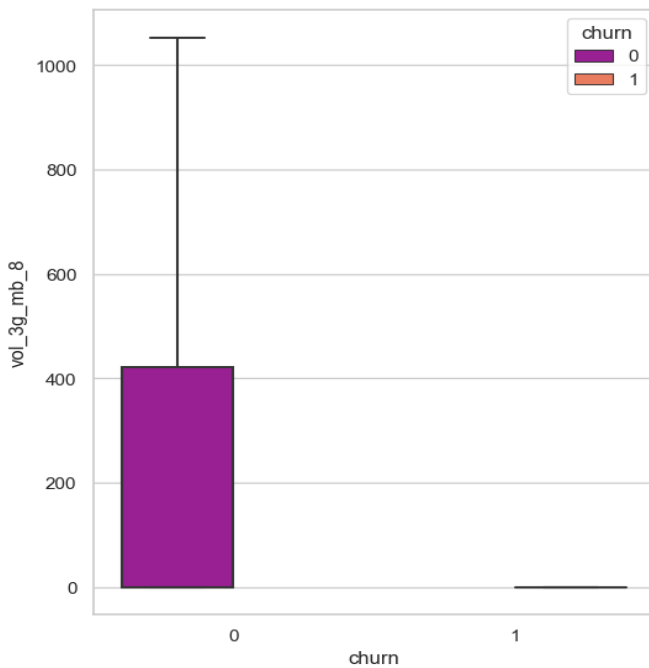
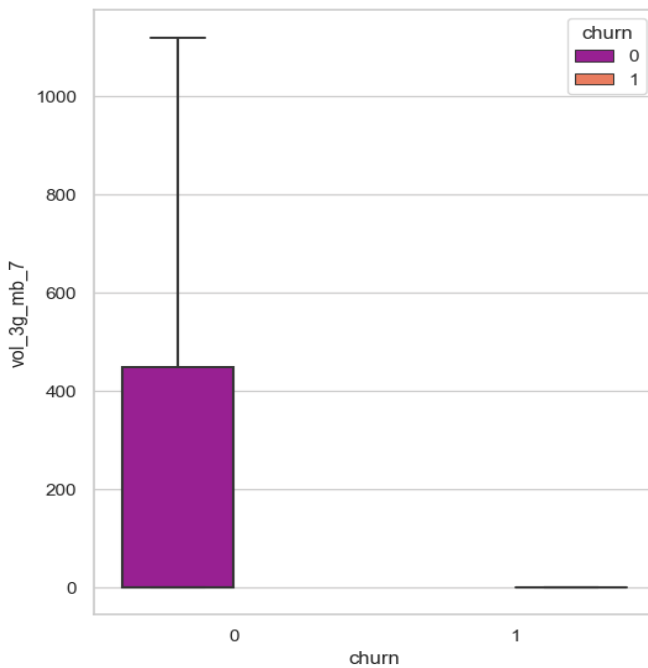
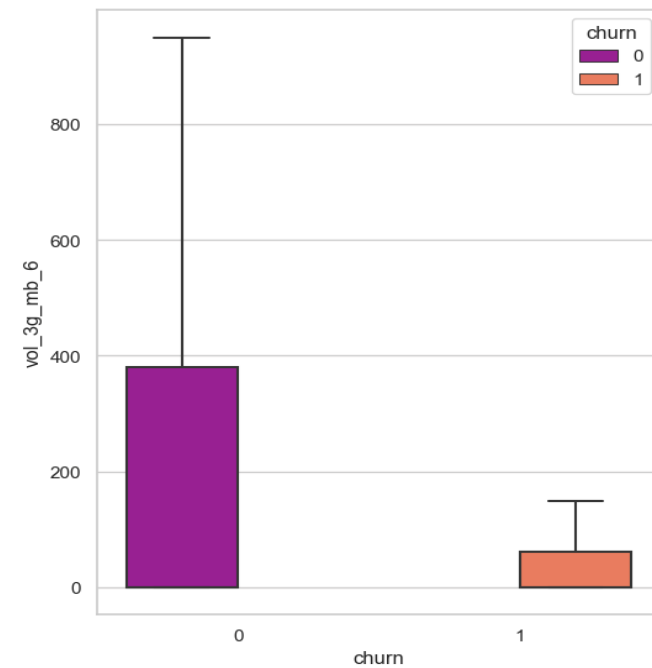
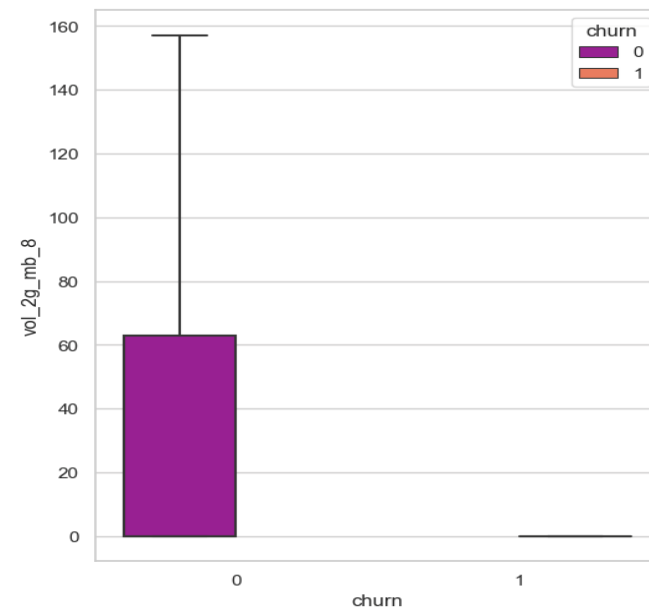
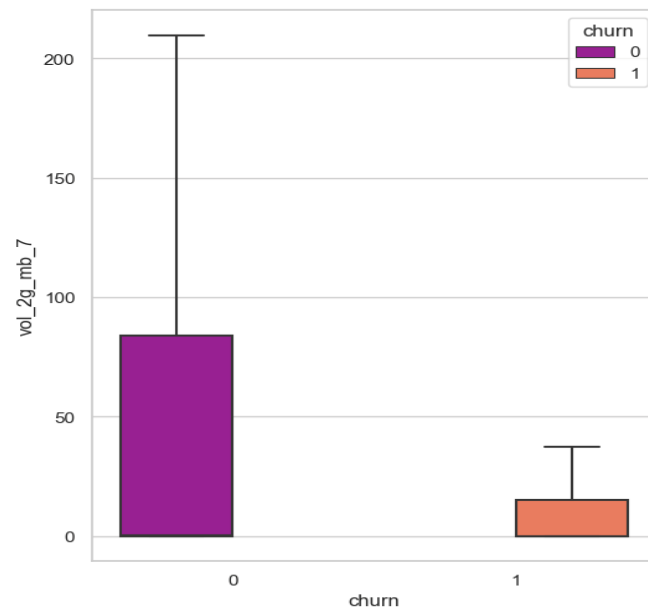
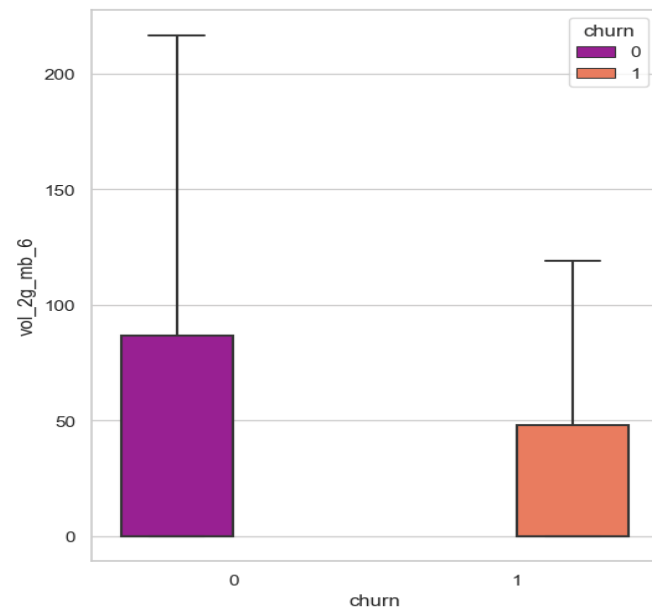
- ❑ Importing and inspecting data
- ❑ EDA
- ❑ Dummy variables creation
- ❑ Feature scaling
- ❑ Dealing with class imbalance
- ❑ Model Building & Evaluation
- ❑ Conclusion & Recommendations

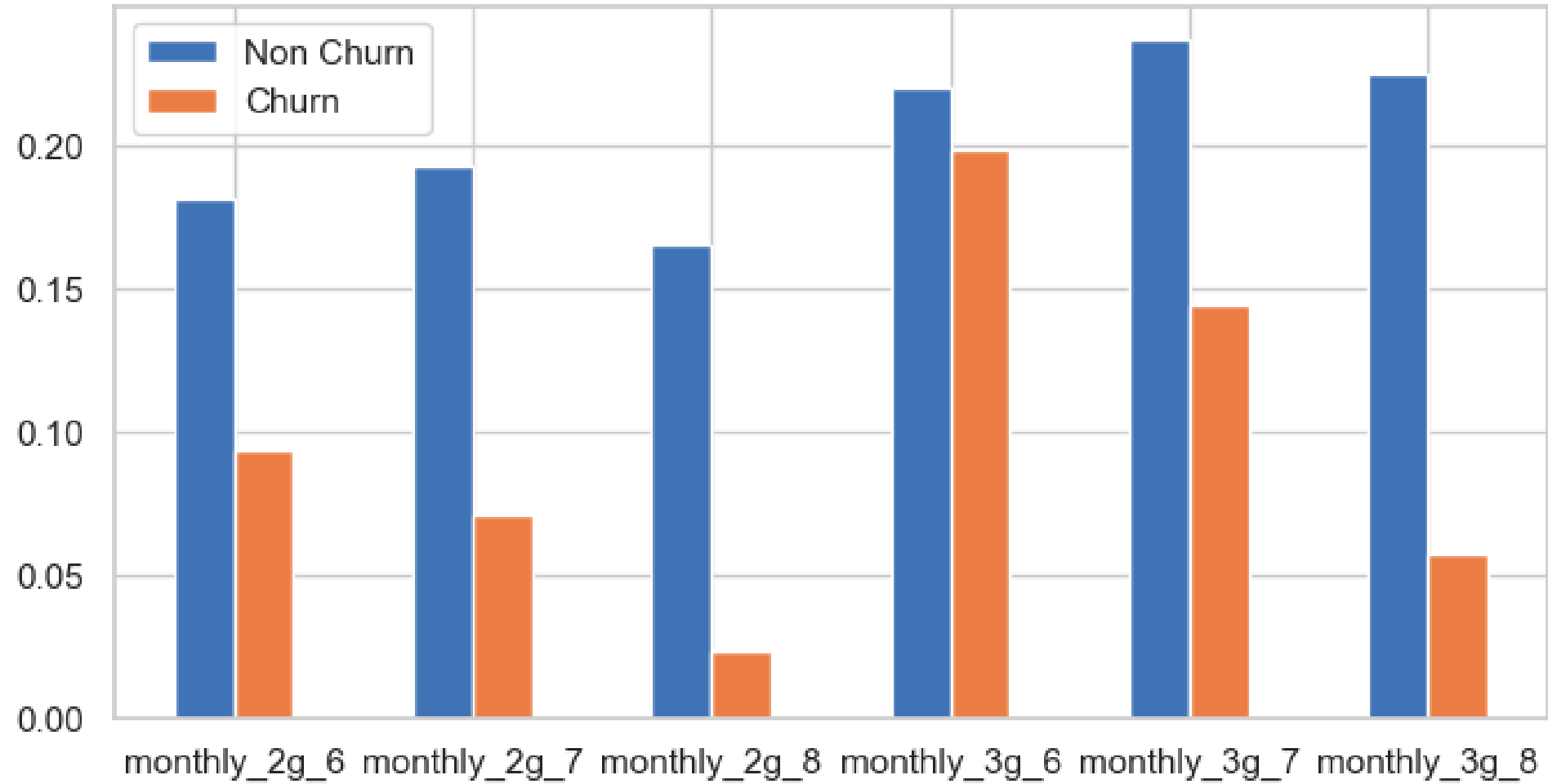


EDA (UNIVARIATE ANALYSIS)

Outliers

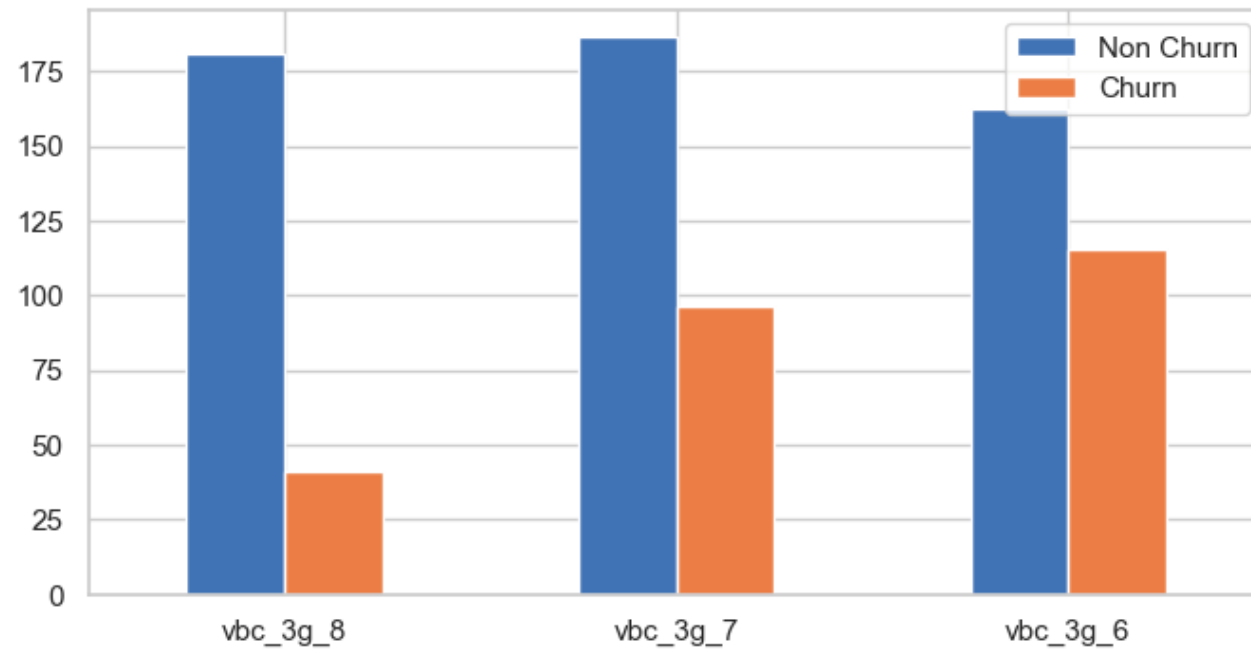
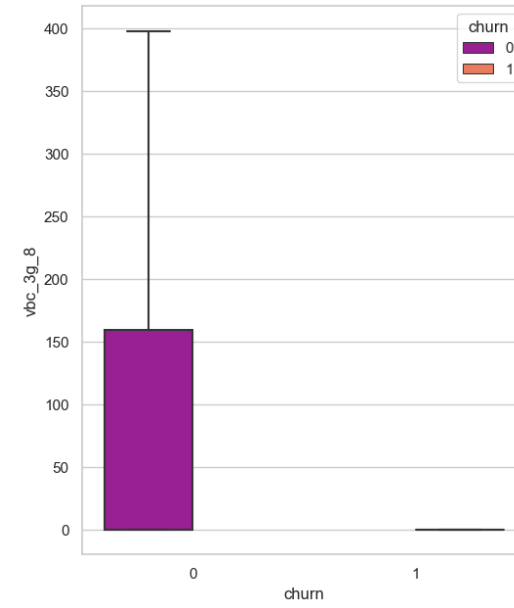
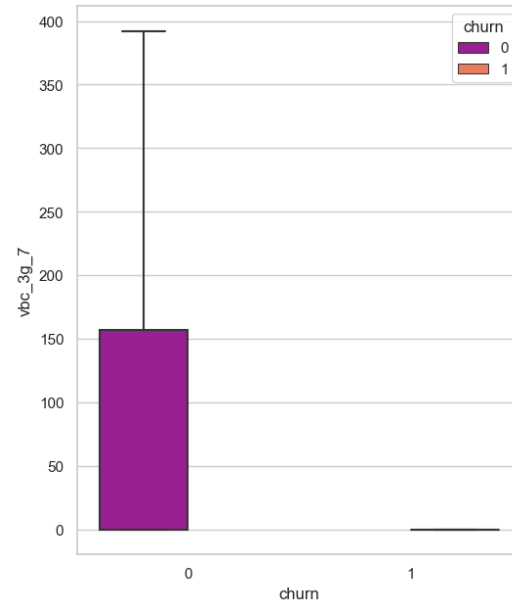
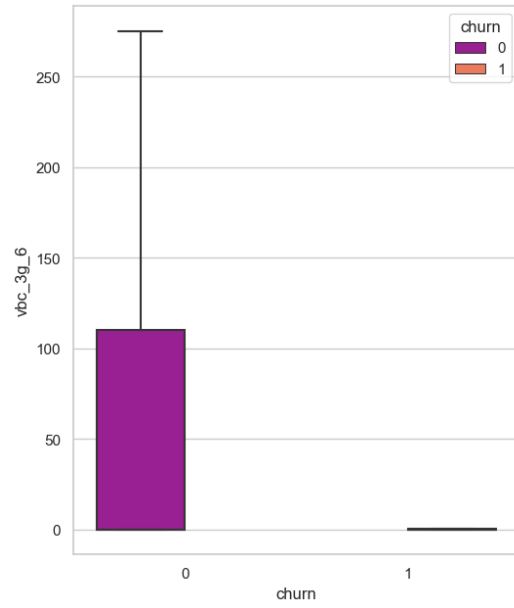


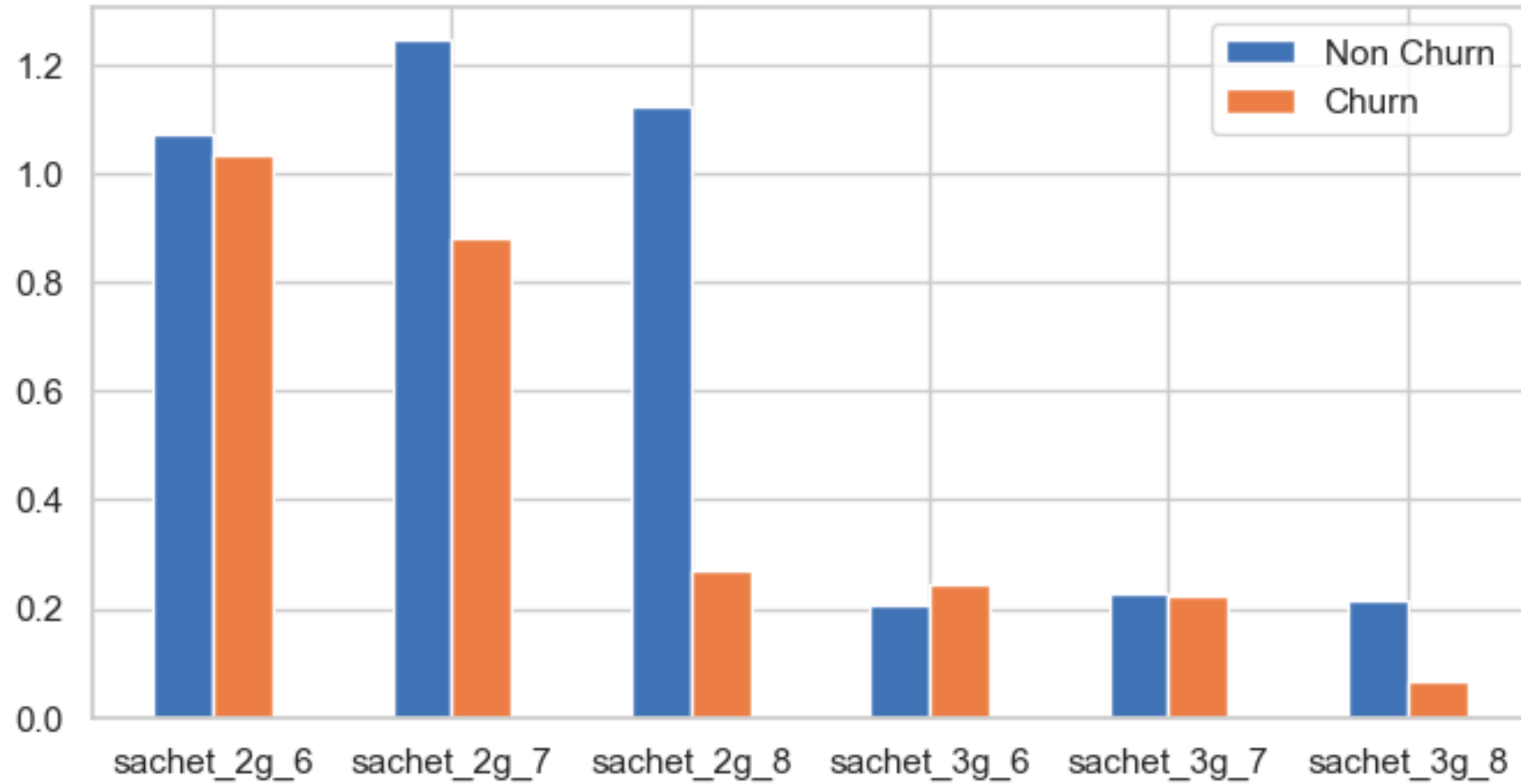




- There is a drop in monthly subscription for churned customers in 8th Month.





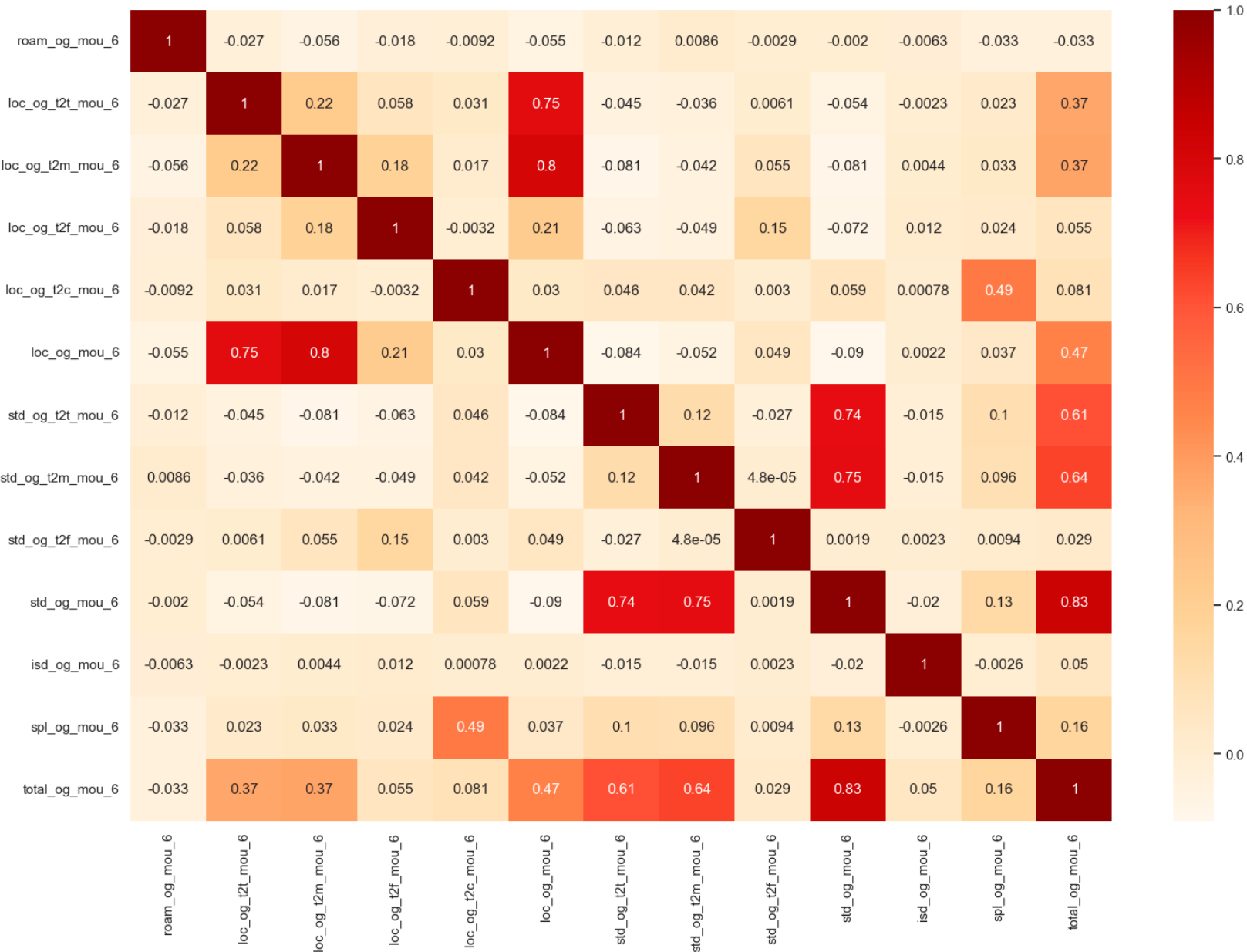


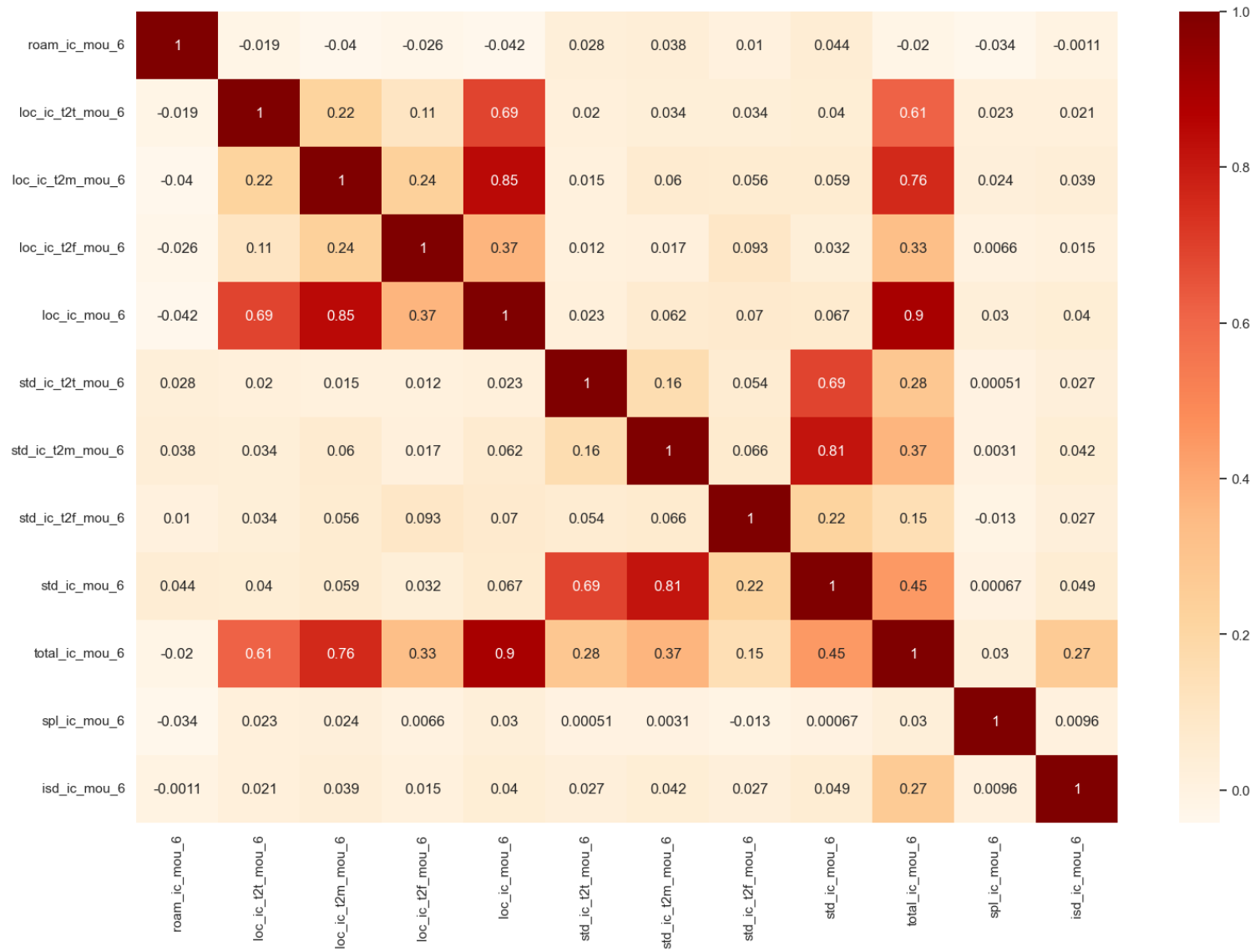
- Drop in sachet services in 8th month for churned customers



MULTIVARIATE ANALYSIS

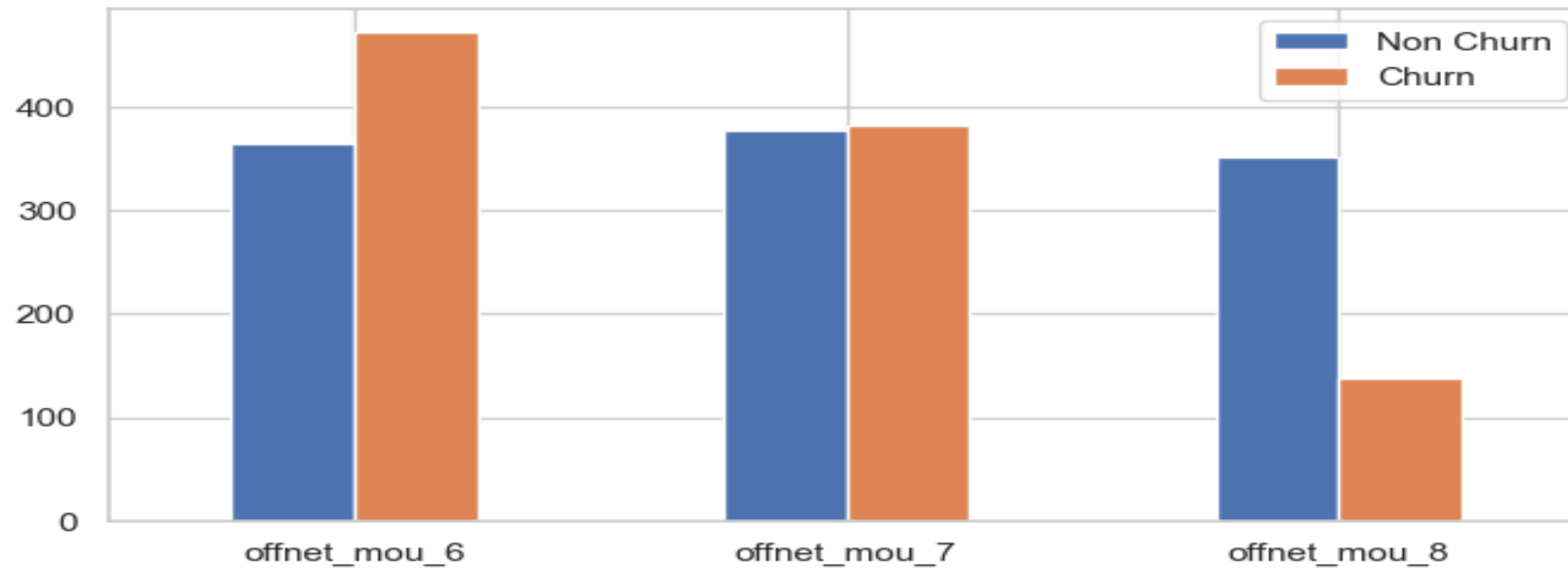
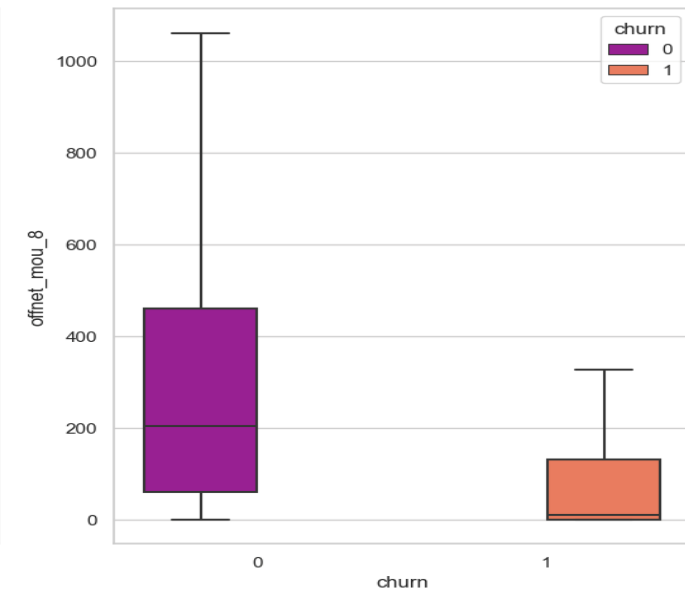
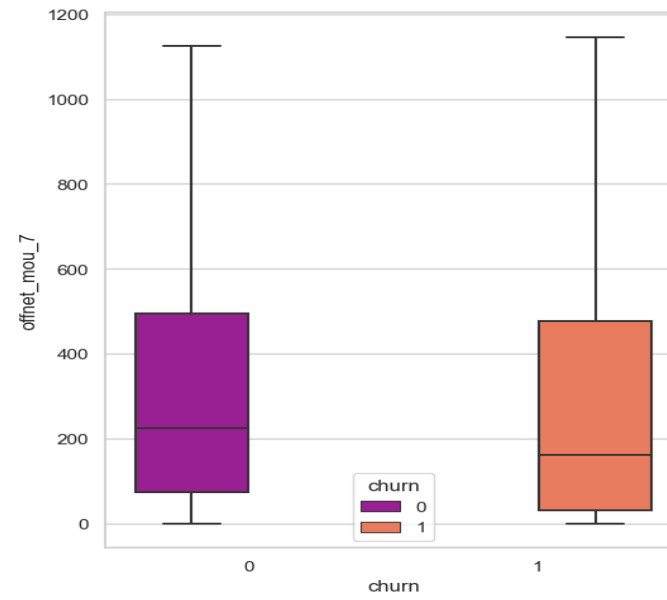
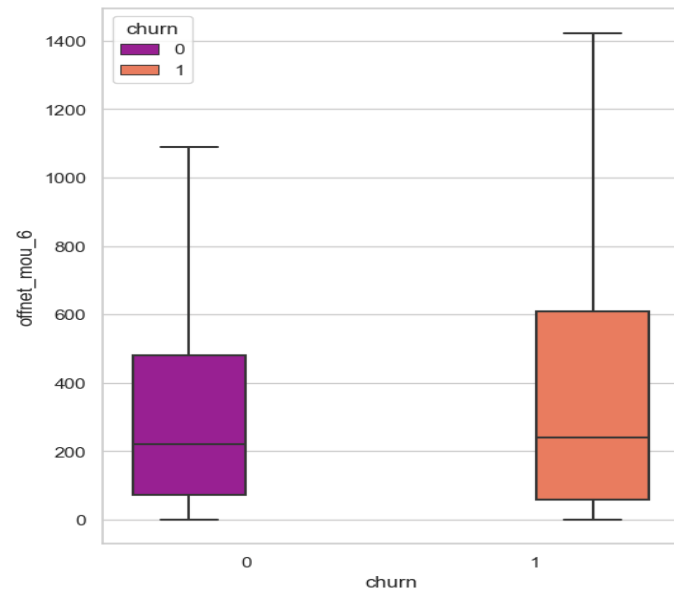
- The total_og_mou_6, std_og_mou_6 and loc_og_mou_6 have strong correlation with other fields and they need to be inspected to avoid any multicollinearity issues.





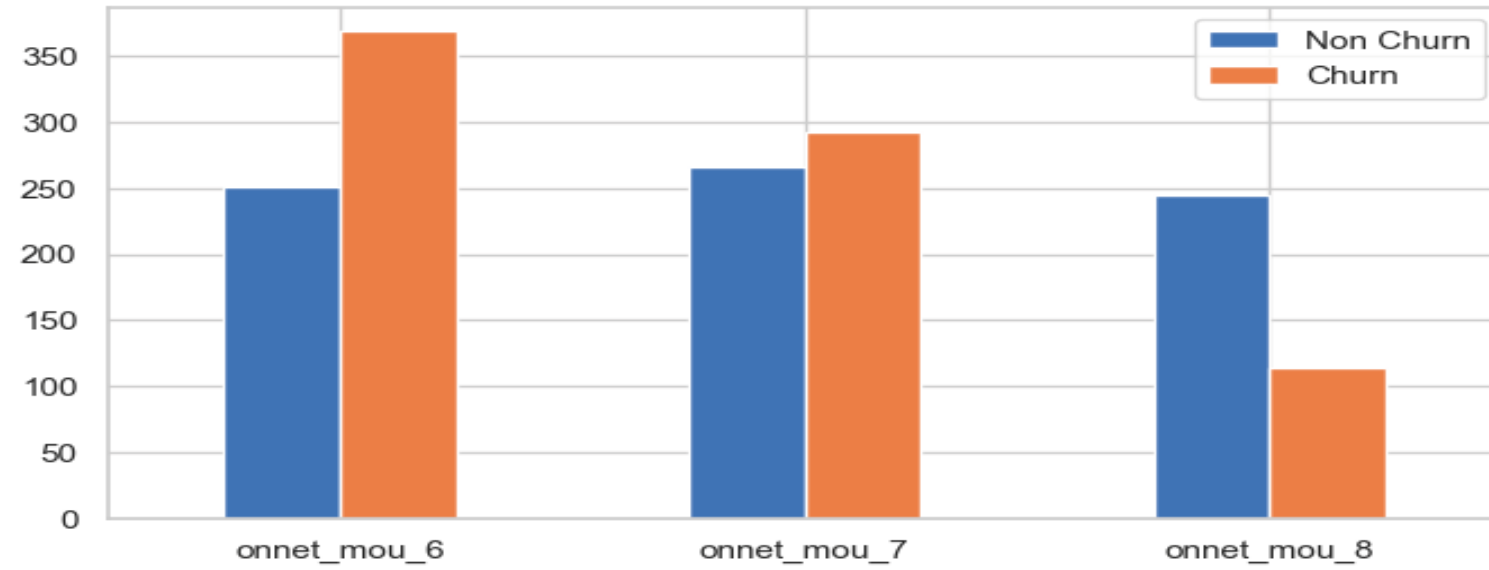
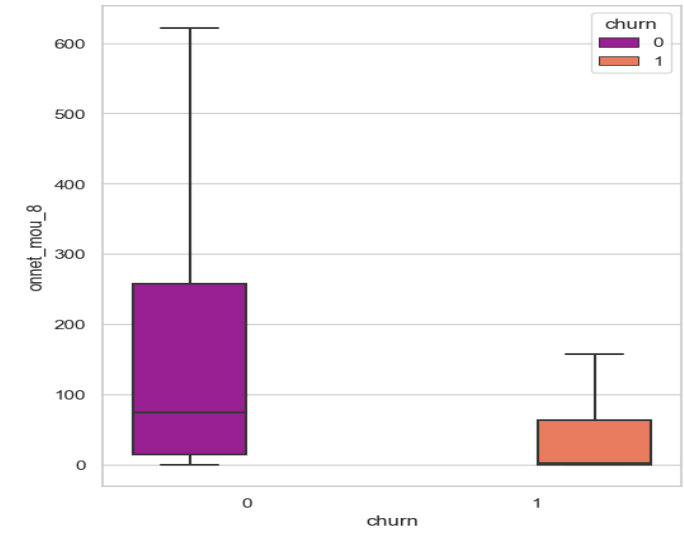
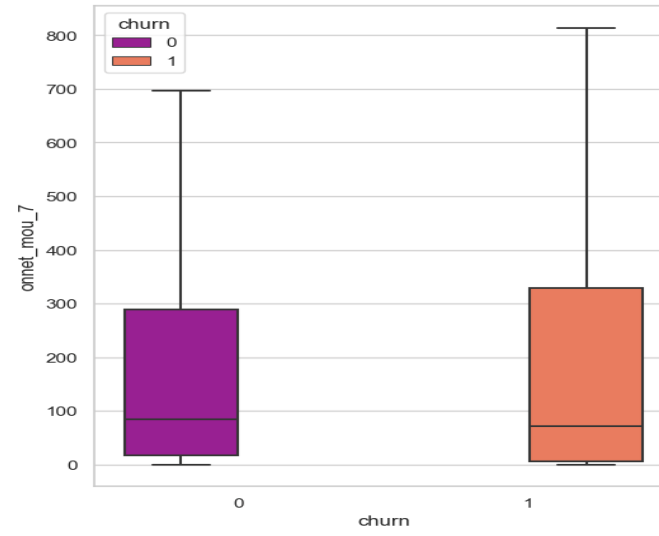
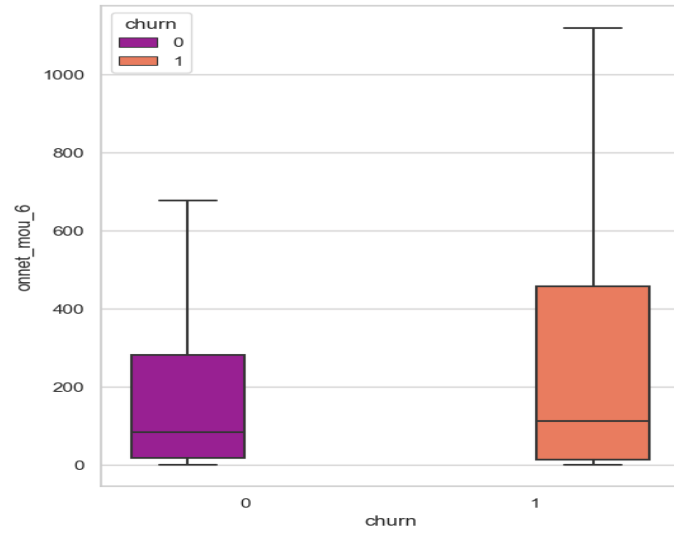
- The total_ic_mou_6, std_ic_mou_6 and loc_ic_mou_6 seems to have strong correlation with other fields and they needs to be inspected to avoid multicollinearity issues.

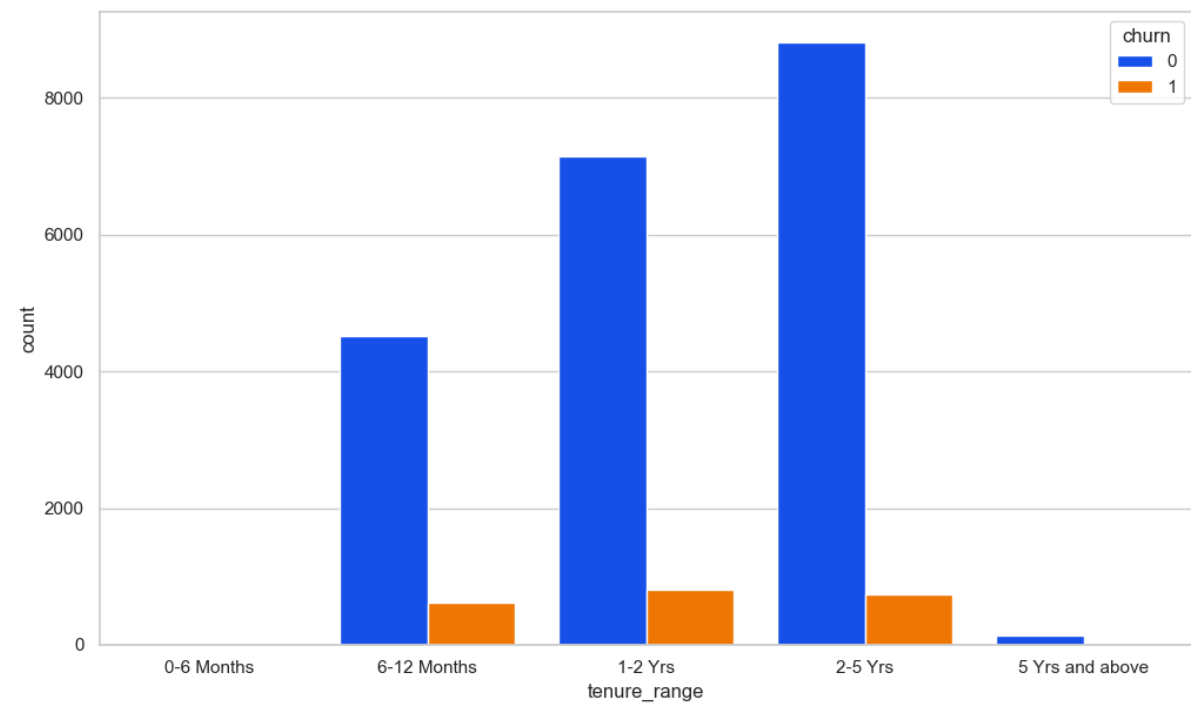
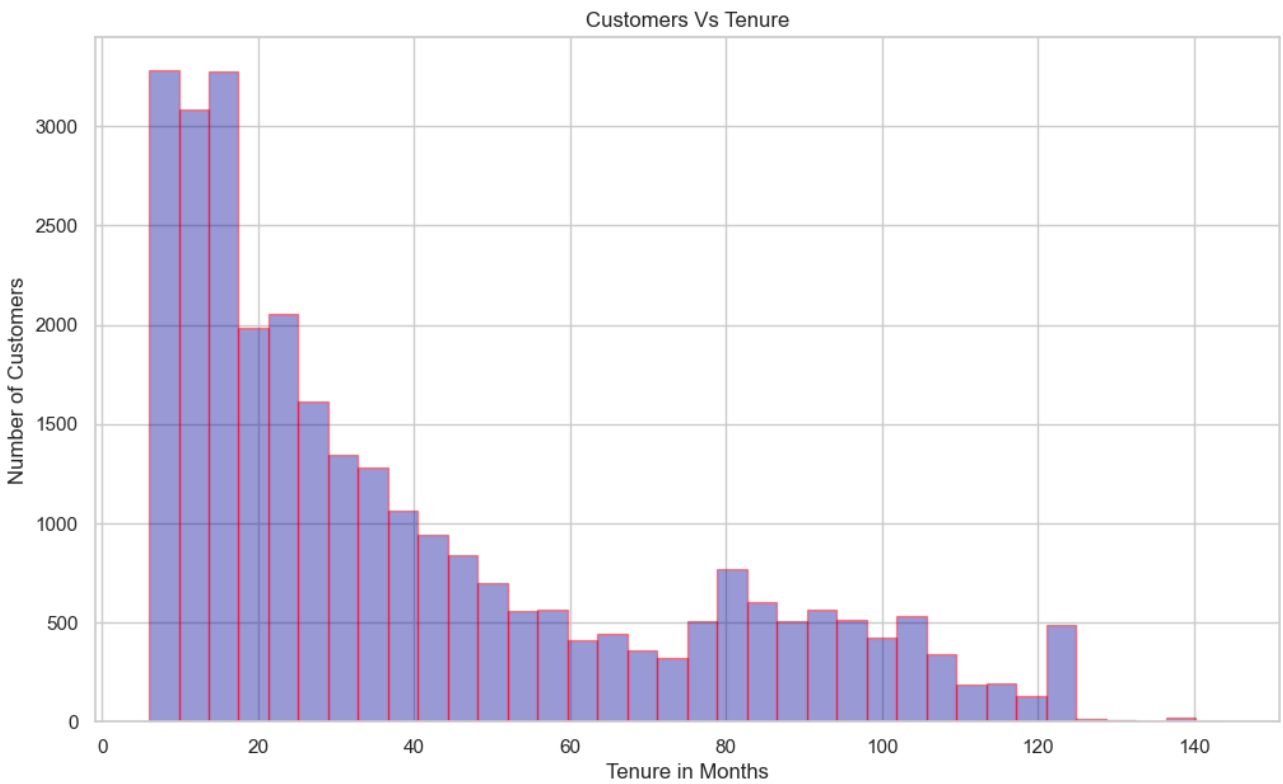




- There is a drop for off net mou services in the 8th month







EDA - INFERENCES

- huge drop in total recharge number also in the 8th month
- drop in maximum recharge for data also in the 8th month
- huge drop in 8th month recharge amount for churned customers
- drop in monthly subscription for churned customers in 8th Month.
- sachet services in 8th month for churned customers
- there is a drops for Arpu in 8th month for churned customers
- The total_og_mou_6, std_og_mou_6 and loc_og_mou_6 have strong correlation with other fields and they needs to be inspected to avoid any multicollinearity issues.
- The total_ic_mou_6, std_ic_mou_6 and loc_ic_mou_6 seems to have strong correlation with other fields and they needs to be inspected to avoid multicollinearity issues.



Handling Imbalanced Dataset:

- Churn rate is less than 10% of the overall data available. This indicates that we would need to handle the class imbalance in this classification problem.
- We can handle imbalanced classes by balancing the classes by increasing minority or decreasing majority. These are following few techniques.
- Random Under-Sampling
- Random Over-Sampling
- SMOTE - Synthetic Minority Oversampling Technique
- we ran three different algorithm (Logistic Regression, Decision tree, Random forest)with above techniques to handle imbalanced classes.
- We pick up **Logistic Regression with Over-Sampling method as best performing model** (based on Highest Recall value).



Model Building

- Splitting data into train and test set (70:30 ratio)
- Use RFE to choose top 25 feature variables for model building
- Repeated model building by removing variable whose p-value is more than 0.05
 - and VIF value greater than 5
- Balancing all metrics, 0.55 is taken as optimal cut off point
- With 0.55 as cut off point, below are performance evaluation metrics for train &
 - test data set.
- Final model is built and with probability threshold value of 0.55

Train set

- Accuracy = 0.84
- Sensitivity = 0.86
- Specificity = 0.82

Test set

- Accuracy = 0.82
- Sensitivity = 0.82
- Specificity = 0.82



Feature Analysis:

- Minutes of usage is one of the most important feature in churn prediction in action phase(i.e 8th month). Minutes of usage (Voice) in the action month is a very good indicator for customer churn. As the MOU in 8th month decreases, the chances of the customer churning out increases.
- Drop in Last day recharge amount in the action month(8th month) is an important indicator for churn of high value customers.
- Churn customers seems to have a relatively High ARPU (Average Revenue Per User) in the good phase, indicating that a sudden downturn in the ARPU from good to action phase is a major indicator of the customer churn.



Conclusions

- Less number of high value customer are churning but for last 6 months no new high valued customer 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 schemes to that group.
- **Average revenue per user** seems to be most important feature in determining churn prediction.
- **Incoming and Outgoing Calls on roaming** for 8th month are strong indicators of churn behavior
- **Local Outgoing calls** made to landline , fixed line , mobile and call center provides a strong indicator of churn behavior.
- Better 2G/3G area coverage where 2G/3G services are not good, it's strong indicator of churn behavior.



Business Recommendations:

- ☐ To recently joined customers (Tenure: 0-2 years) can be provided add on incentives for certain period.
- ☐ Provide incentives on recharge of data or voice to high value customers in the action phase to increase customer retention.
- ☐ Incentives based on usage can be provided to Customer having high ARPU in good phase to drive up the ARPU in action phase and retain the customer.
- ☐ Can provide free or discounted local on net and mobile usage voice minutes during the action phase.
- ☐ To retain customers, we need higher recall. As giving an offer to an user not going to churn will cost less as compared to loosing a customer and bring new customer, we need to have high rate of correctly identifying the true positives, hence recall.



