TELECOM CHURN CASE STUDY

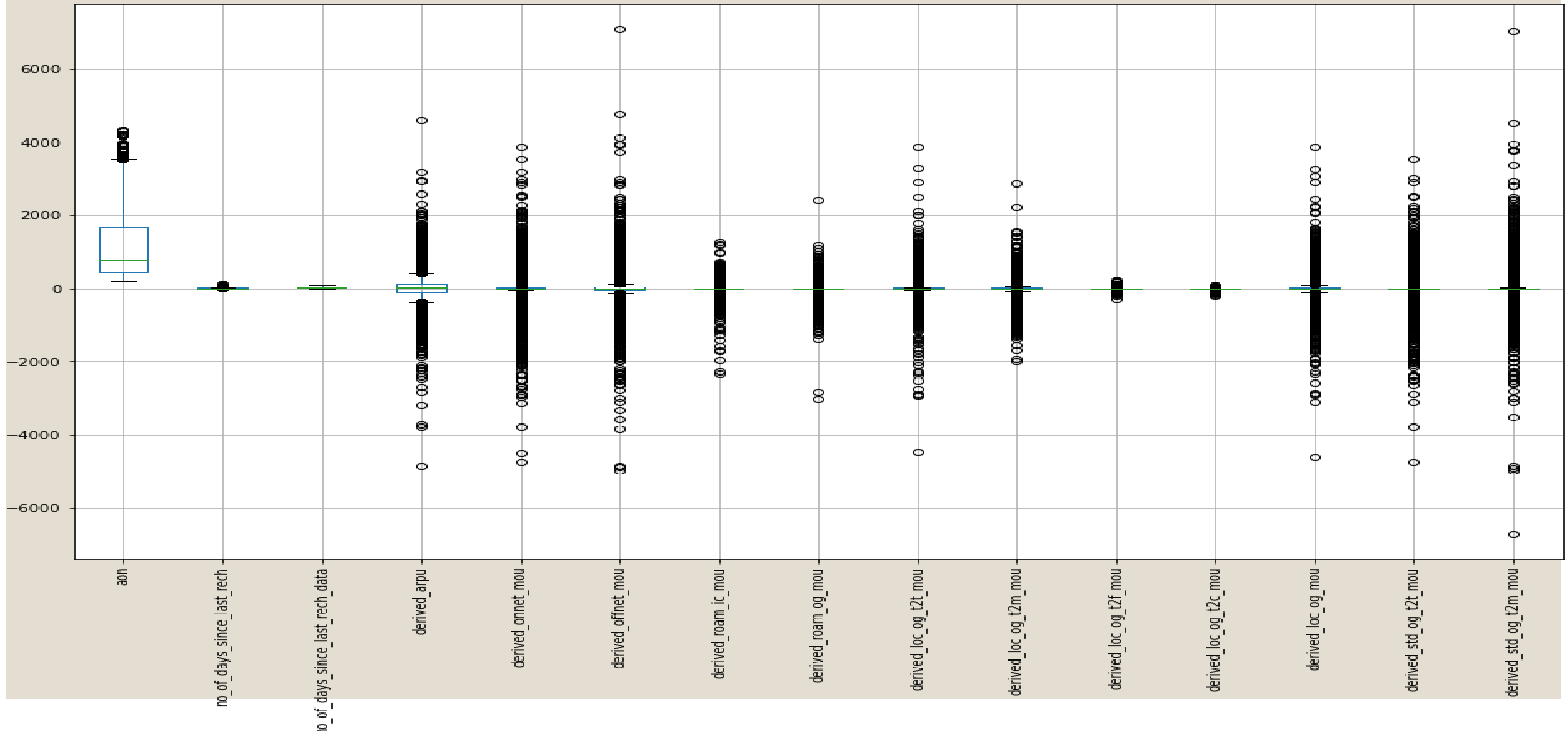
By:- Akshata P V, Abhilash, Nachiketa

Details

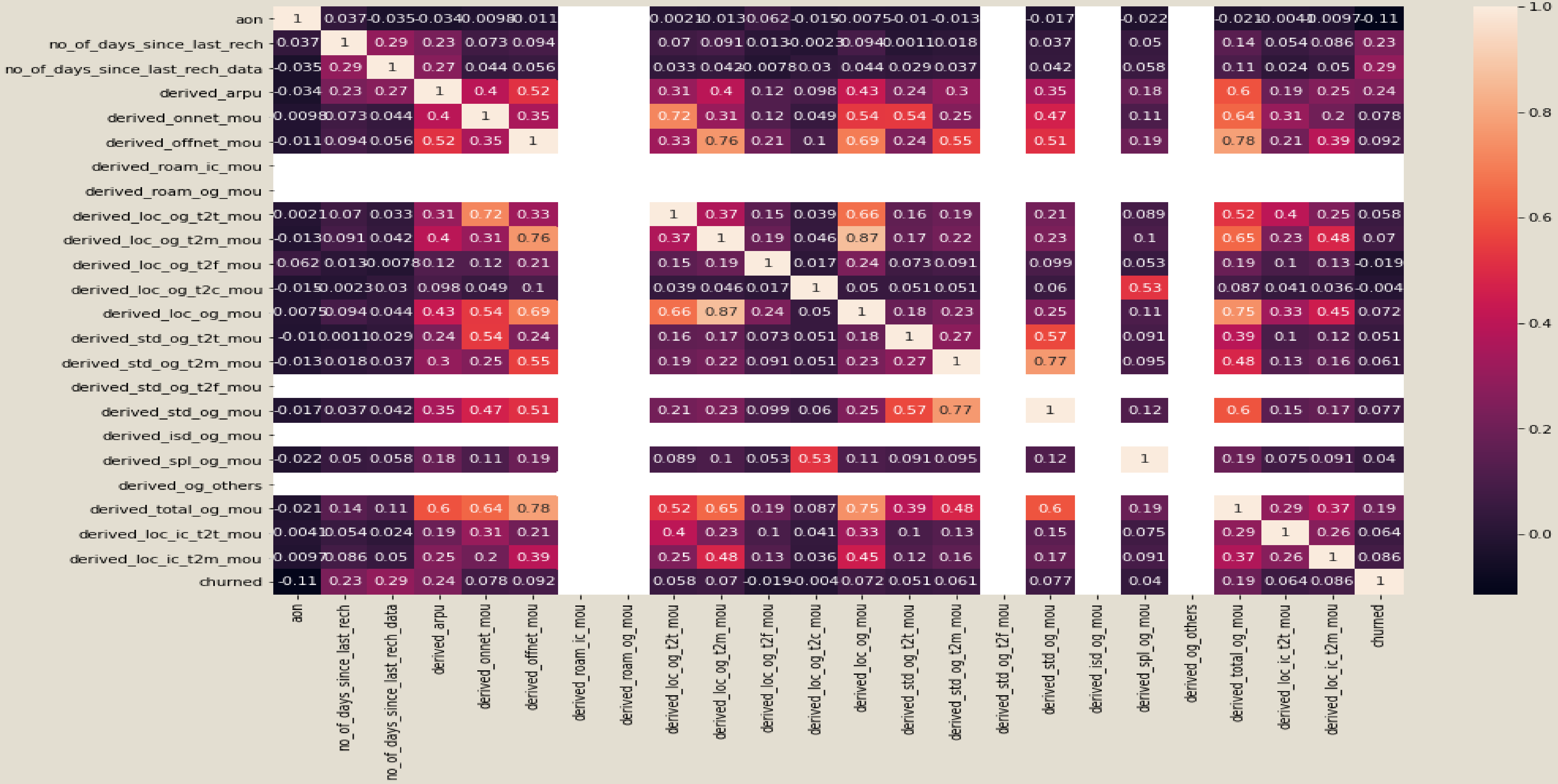
- In this project, we will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn, and identify the main indicators of churn.
- This project is based on the Indian and Southeast Asian market.
- The dataset contains customer-level information 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. Understanding the typical customer behaviour during churn will be helpful to do this task well.

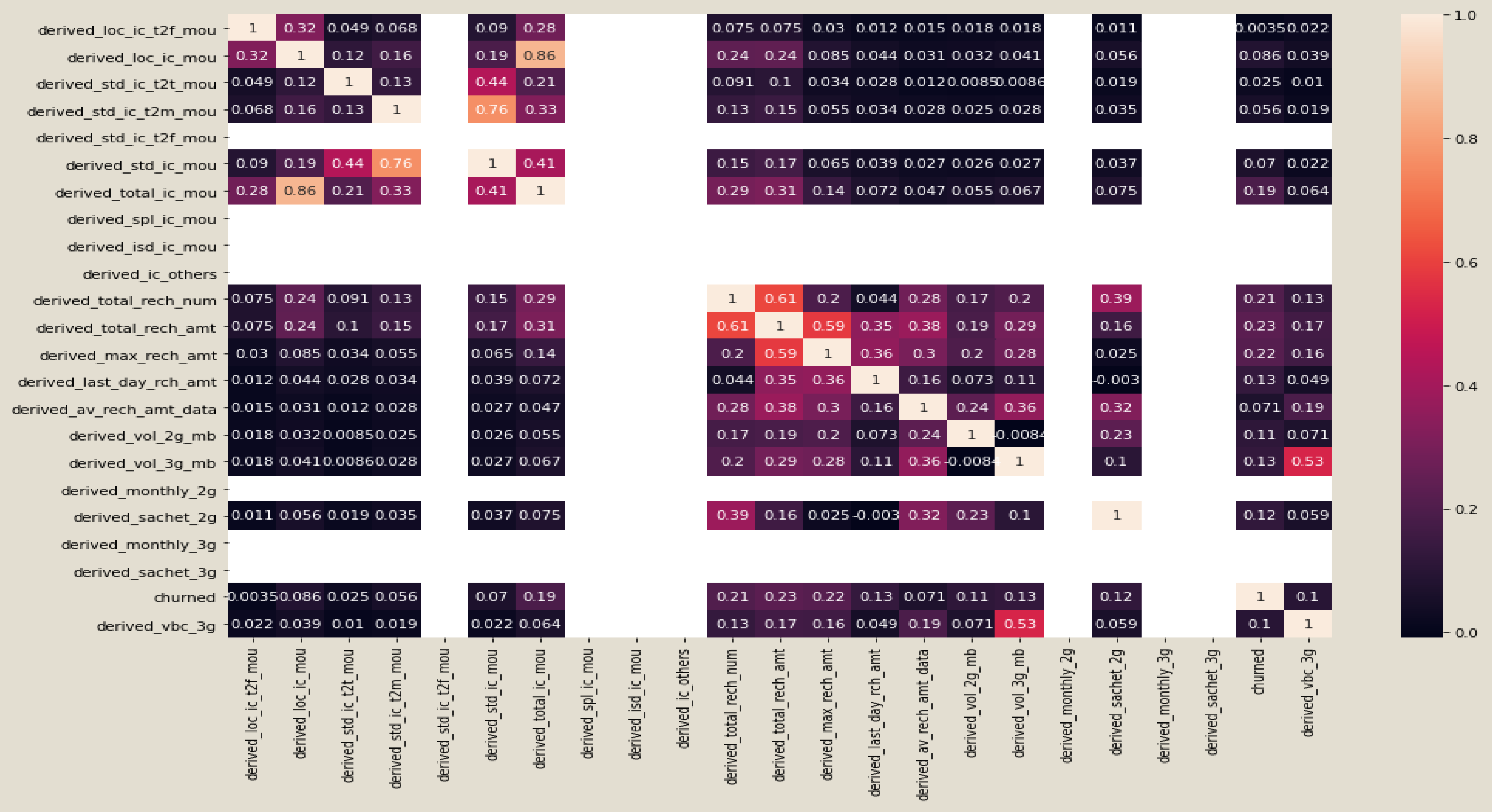
EDA

- **Outlier Treatment:** Most of the features have a high number of outliers, dropping the outliers will cause a loss of information. Hence capping and flooring the extreme values.



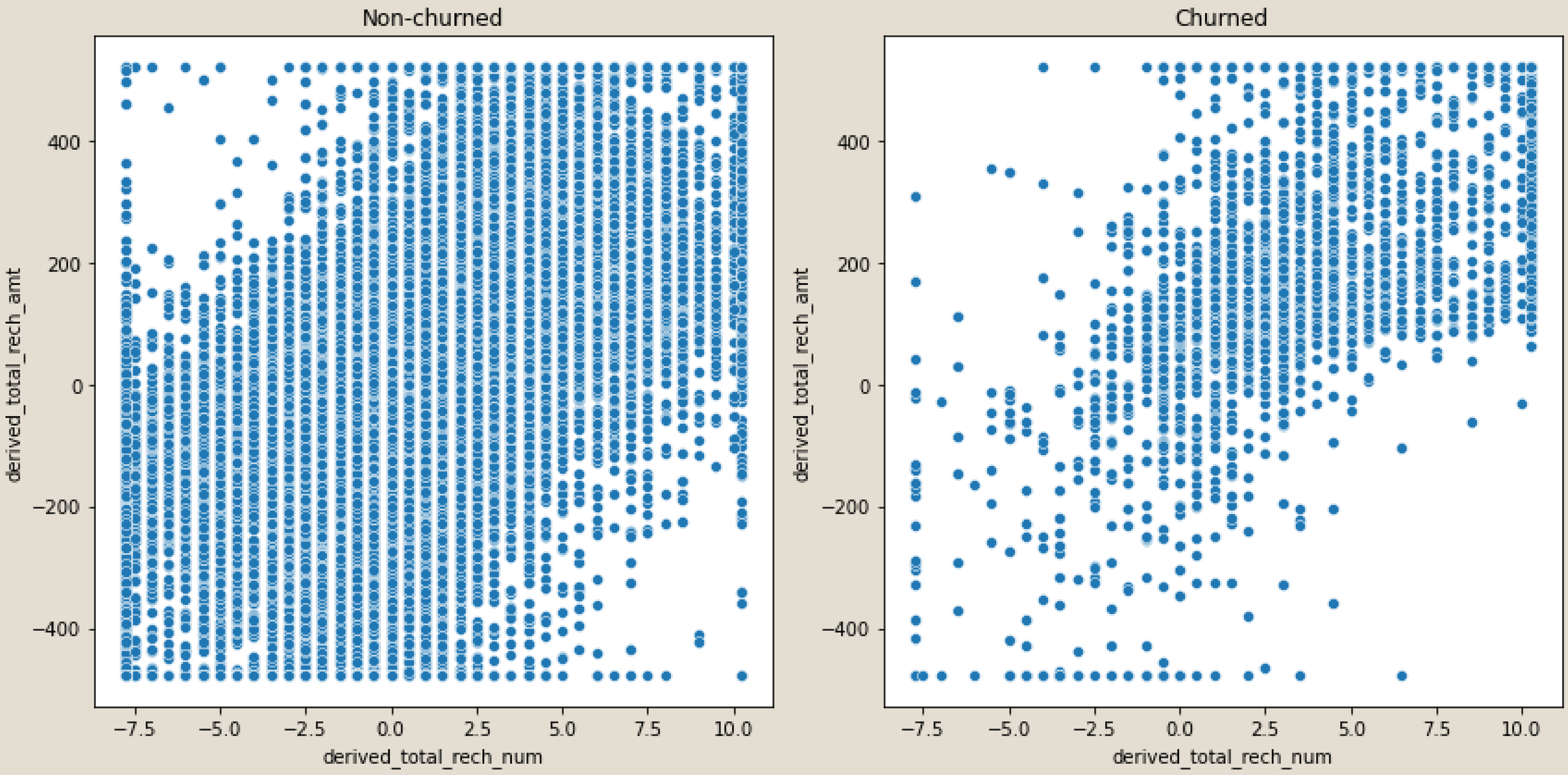
Checking Correlation





derived_loc_ic_t2f_mou 1 0.32 0.049 0.068 0.09 0.28 0.075 0.075 0.03 0.012 0.015 0.018 0.018 0.011 0.0035 0.022
derived_loc_ic_mou 0.32 1 0.12 0.16 0.19 0.86 0.24 0.24 0.085 0.044 0.031 0.032 0.041 0.056 0.086 0.039
derived_std_ic_t2t_mou 0.049 0.12 1 0.13 0.44 0.21 0.091 0.1 0.034 0.028 0.012 0.0085 0.0086 0.019 0.025 0.01
derived_std_ic_t2m_mou 0.068 0.16 0.13 1 0.76 0.33 0.13 0.15 0.055 0.034 0.028 0.025 0.028 0.035 0.056 0.019
derived_std_ic_t2f_mou
derived_std_ic_mou 0.09 0.19 0.44 0.76 1 0.41 0.15 0.17 0.065 0.039 0.027 0.026 0.027 0.037 0.07 0.022
derived_total_ic_mou 0.28 0.86 0.21 0.33 0.41 1 0.29 0.31 0.14 0.072 0.047 0.055 0.067 0.075 0.19 0.064
derived_spl_ic_mou
derived_isd_ic_mou
derived_ic_others
derived_total_rech_num 0.075 0.24 0.091 0.13 0.15 0.29 1 0.61 0.2 0.044 0.28 0.17 0.2 0.39 0.21 0.13
derived_total_rech_amt 0.075 0.24 0.1 0.15 0.17 0.31 0.61 1 0.59 0.35 0.38 0.19 0.29 0.16 0.23 0.17
derived_max_rech_amt 0.03 0.085 0.034 0.055 0.065 0.14 0.2 0.59 1 0.36 0.3 0.2 0.28 0.025 0.22 0.16
derived_last_day_rch_amt 0.012 0.044 0.028 0.034 0.039 0.072 0.044 0.35 0.36 1 0.16 0.073 0.11 -0.003 0.13 0.049
derived_av_rech_amt_data 0.015 0.031 0.012 0.028 0.027 0.047 0.28 0.38 0.3 0.16 1 0.24 0.36 0.32 0.071 0.19
derived_vol_2g_mb 0.018 0.032 0.0085 0.025 0.026 0.055 0.17 0.19 0.2 0.073 0.24 1 -0.0084 0.23 0.11 0.071
derived_vol_3g_mb 0.018 0.041 0.0086 0.028 0.027 0.067 0.2 0.29 0.28 0.11 0.36 -0.0084 1 0.1 0.13 0.53
derived_monthly_2g
derived_sachet_2g 0.011 0.056 0.019 0.035 0.037 0.075 0.39 0.16 0.025 -0.003 0.32 0.23 0.1 1 0.12 0.059
derived_monthly_3g
derived_sachet_3g
churned 0.0035 0.086 0.025 0.056 0.07 0.19 0.21 0.23 0.22 0.13 0.071 0.11 0.13 0.12 1 0.1
derived_vbc_3g 0.022 0.039 0.01 0.019 0.022 0.064 0.13 0.17 0.16 0.049 0.19 0.071 0.53 0.059 0.1 1
derived_loc_ic_t2f_mou
derived_loc_ic_mou
derived_std_ic_t2t_mou
derived_std_ic_t2m_mou
derived_std_ic_t2f_mou
derived_std_ic_mou
derived_total_ic_mou
derived_spl_ic_mou
derived_isd_ic_mou
derived_ic_others
derived_total_rech_num
derived_total_rech_amt
derived_max_rech_amt
derived_last_day_rch_amt
derived_av_rech_amt_data
derived_vol_2g_mb
derived_vol_3g_mb
derived_monthly_2g
derived_sachet_2g
derived_monthly_3g
derived_sachet_3g
churned
derived_vbc_3g
1.0
0.8
0.6
0.4
0.2
0.0

Bivariate and Multivariate Analysis

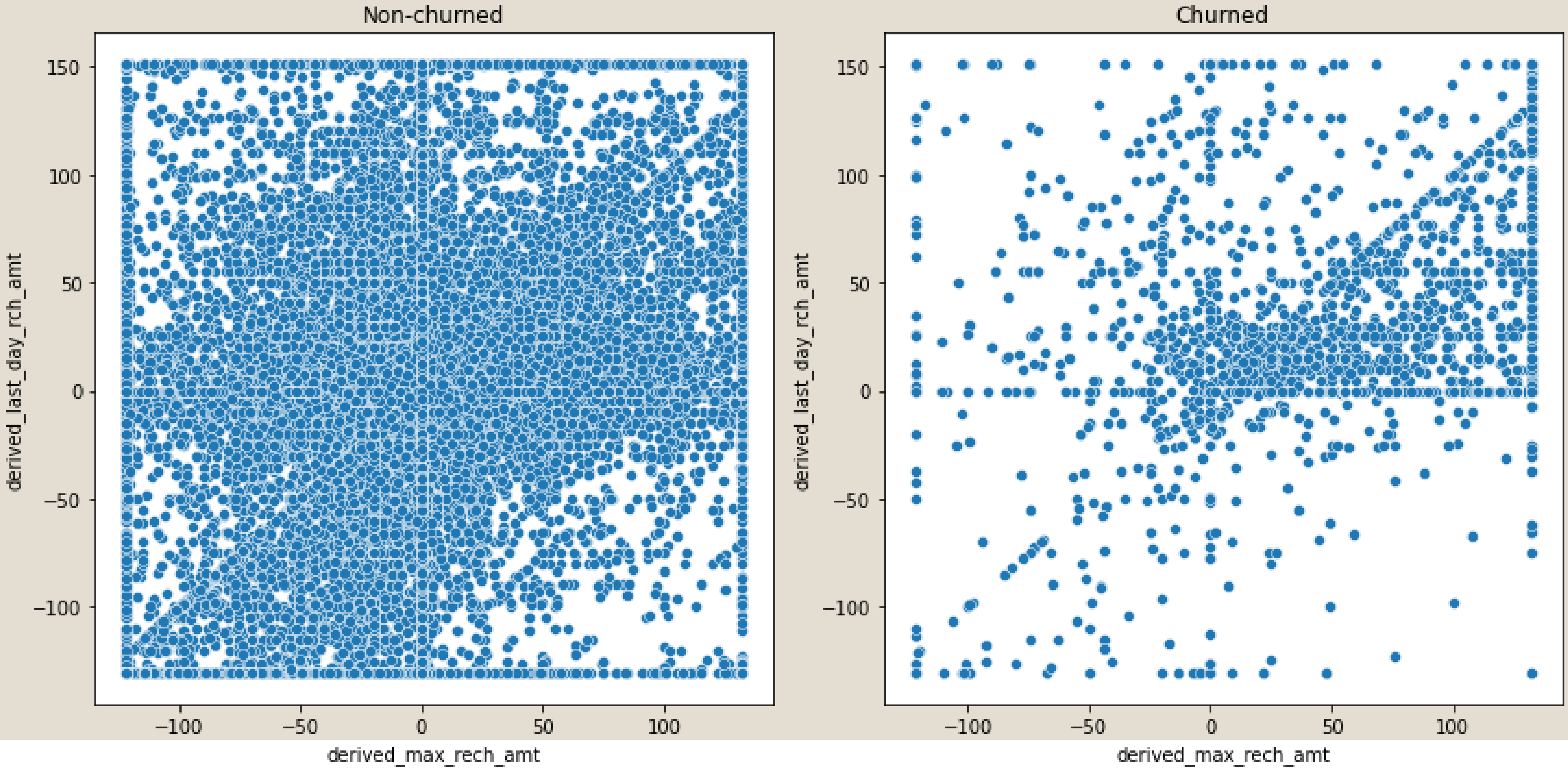


Observation:

- There is a positive correlation between total recharge amount and total number of recharges.
- 'derived_total_rech_amt' is derived by taking average of total recharge done in good phase (i.e. month 6 and 7) and deducting recharge of action phase (month 8) from it.
- A positive value of this feature means, recharge amount value dropped from good phase to action phase. The same applies for 'derived_total_rech_num' as well.
- A negative value of this feature means, recharge amount value increased from good phase to action phase. The same applies for 'derived_total_rech_num' as well.
- By comparing churned and non-churned customer we can see that, for churned customers both recharge number and recharge amount have dropped in action phase.

Observation:

- derived_max_rech_amt and derive_last_day_rch_amt are most likely to be positive for churned customers. This means, drop in maximum recharge amount and last day recharge amount in action phase can signify customer churn.



VIF data frame for all the variables

	Features	VIF
3	derived_arpu	2.47
8	derived_total_rech_num	1.98
5	derived_loc_og_mou	1.60
9	derived_max_rech_amt	1.55
11	derived_av_rech_amt_data	1.46
14	derived_sachet_2g	1.40
2	no_of_days_since_last_rech_data	1.34
7	derived_loc_ic_t2m_mou	1.28
13	derived_vol_3g_mb	1.28
1	no_of_days_since_last_rech	1.19
10	derived_last_day_rch_amt	1.18
12	derived_vol_2g_mb	1.17
6	derived_loc_ic_t2t_mou	1.15
4	derived_loc_og_t2f_mou	1.06
0	aon	1.02

VIF and p-value of the remaining features are in acceptable range. p-value < 0.05 (significance level)

Conclusion

- **Predictor Model**
- Since reducing customer churns with attractive offers is less expensive than making new customers, sensitivity score is considered for the building the final predictor model.
- Logistic Regression trained with 15 principal components has accuracy score of 0.74 and sensitivity score of 0.82 for Probability Cutoff = 0.42
- **The final model could predict 82% of the churned customers correctly out of all the churned customers. (sensitivity)**
- **Accuracy score 0.74 means 74% of the predictions made by the final model are correct out of all the predictions.**