By: Anirudh Swami

Telecom Churn Case Study

Conclusion with No PCA

- ***Model summary***
- Train set
  - Accuracy = 0.84
  - Sensitivity = 0.81
  - Specificity = 0.83
- Test set
  - Accuracy = 0.78
  - Sensitivity = 0.82
  - Specificity = 0.78

- Overall, the model is performing well in the test set, what it had learnt from the train set.

Final conclusion with no PCA

- We can see that the logistic model with no PCA has good sensitivity and accuracy, which are comparable to the models with PCA. So, we can go for the more simplistic model such as logistic regression with PCA as it explains the important predictor variables as well as the significance of each variable. The model also helps us to identify the variables which should be act upon for making the decision of the to be churned customers. Hence, the model is more relevant in terms of explaining to the business.

Business recomendation

- **Top predictors**
- Below are few top variables selected in the logistic regression model.

Variables	Coefficients
loc_ic_mou_8 -3.3287	-3.3287
og_others_7	-2.4711
ic_others_8	-1.5131
isd_og_mou_8	-1.3811
decrease_vbc_action	-1.3293
monthly_3g_8	-1.0943

We can see most of the top variables have negative coefficients. That means, the variables are inversely correlated with the churn probability.

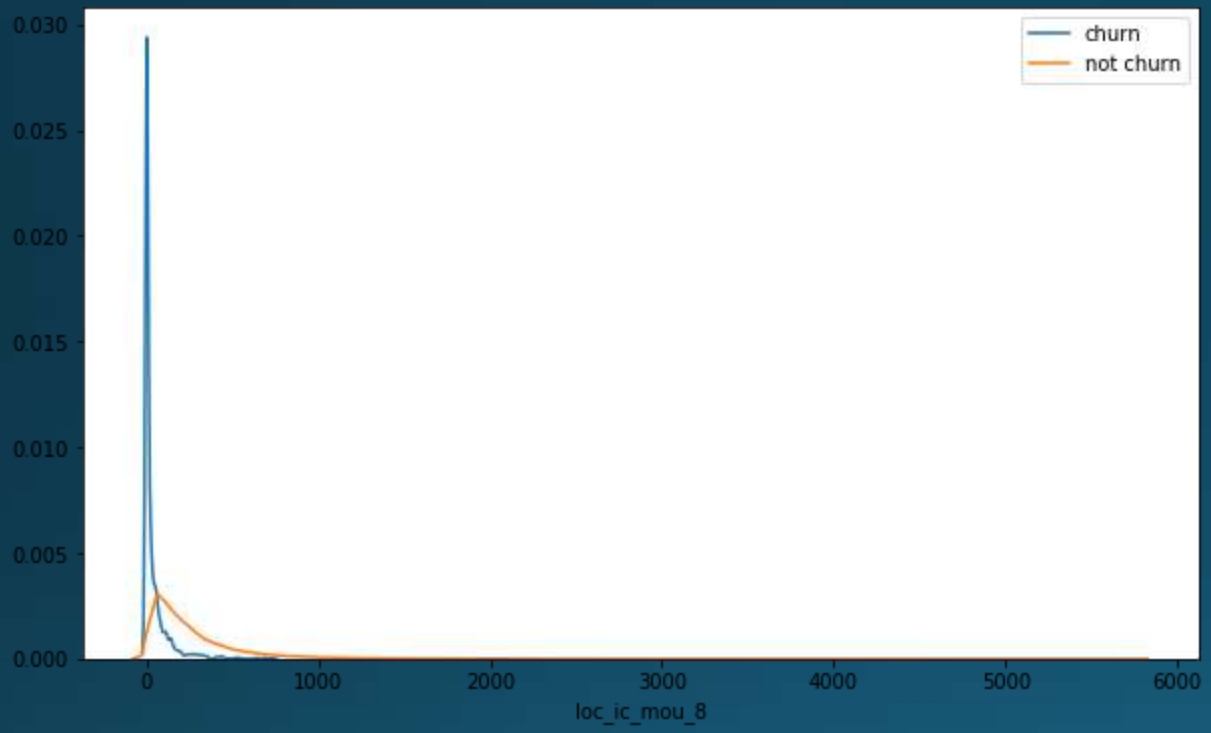
E.g.:-

If the local incoming minutes of usage (loc_ic_mou_8) is lesser in the month of August than any other month, then there is a higher chance that the customer is likely to churn.

Recomendations

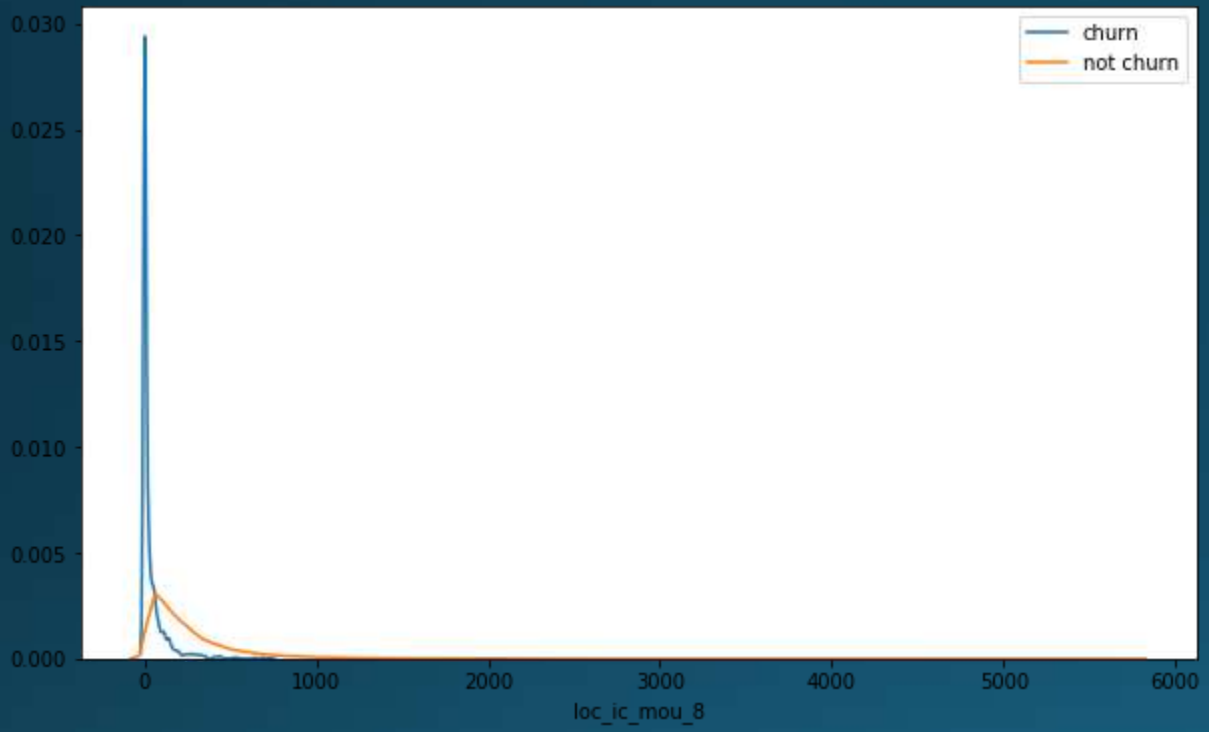
- Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
- Target the customers, whose outgoing others charge in July and incoming others on August are less.
- Also, the customers having value based cost in the action phase increased are more likely to churn than the other customers. Hence, these customers may be a good target to provide offer.
- Customers, whose monthly 3G recharge in August is more, are likely to be churned.
- Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
- Customers decreasing monthly 2g usage for August are most probable to churn.
- Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.
- roam_og_mou_8 variables have positive coefficients (0.7135). That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.

Plots of important predictors for churn and non churn customers



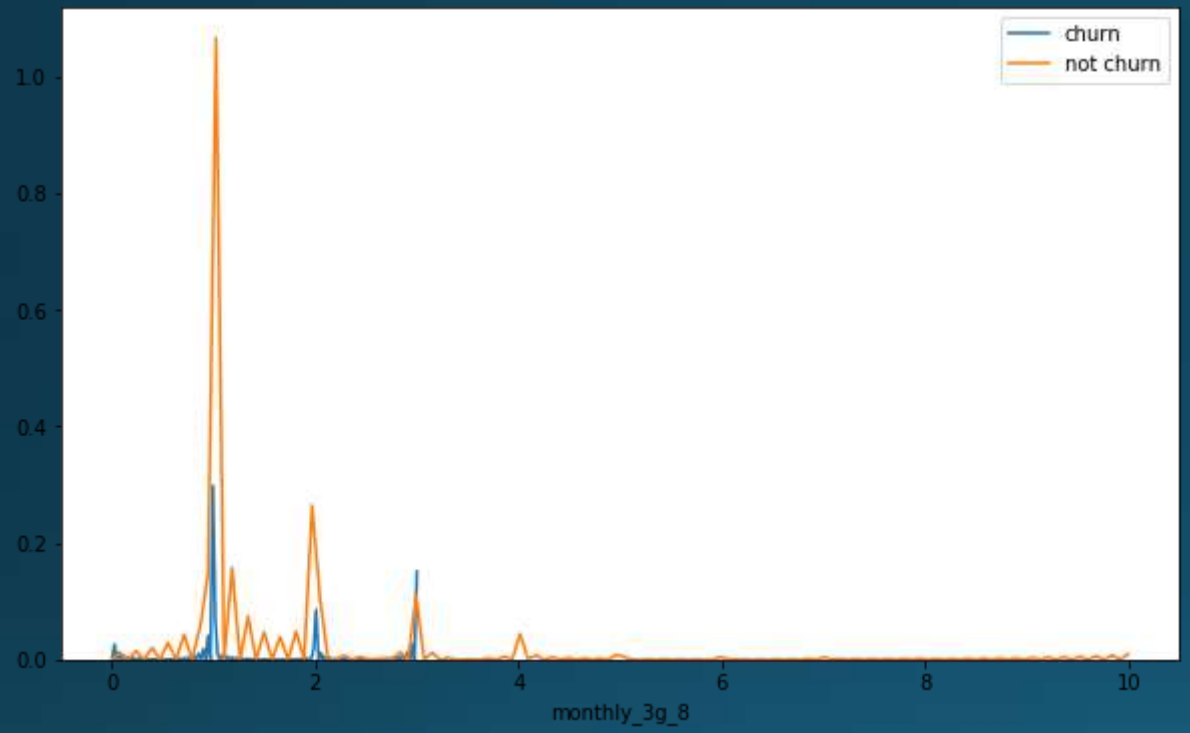
We can see that for the churn customers the minutes of usage for the month of August is mostly populated on the lower side than the non churn customers.

Plots of important predictors for churn and non churn customers



We can see that for the churn customers the minutes of usage for the month of August is mostly populated on the lower side than the non churn customers.

Plots of important predictors for churn and non churn customers



The number of monthly 3g data for August for the churn customers are very much populated aroud 1, whereas of non churn customers it spreader across various numbers.
Similarly we can plot each variables, which have higher coefficients, churn distribution.