

Business problem overview

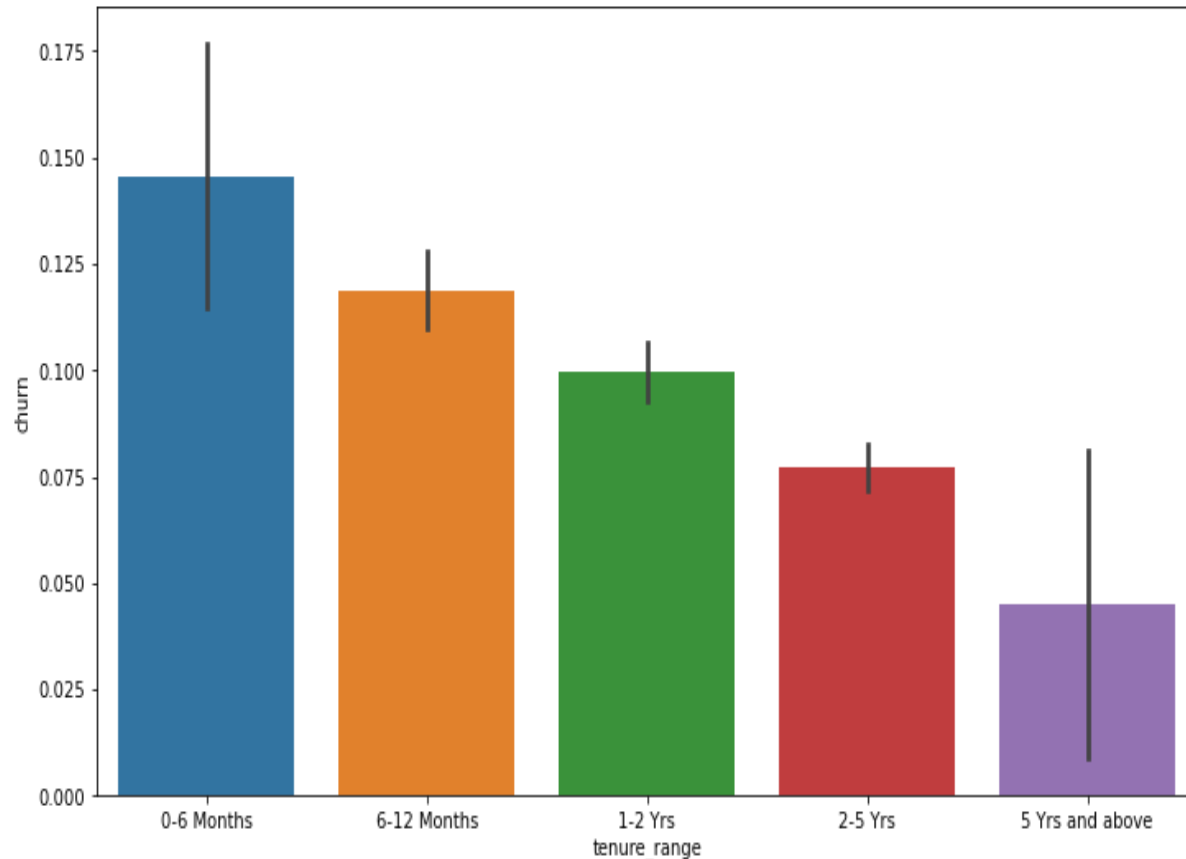
In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to 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, 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.

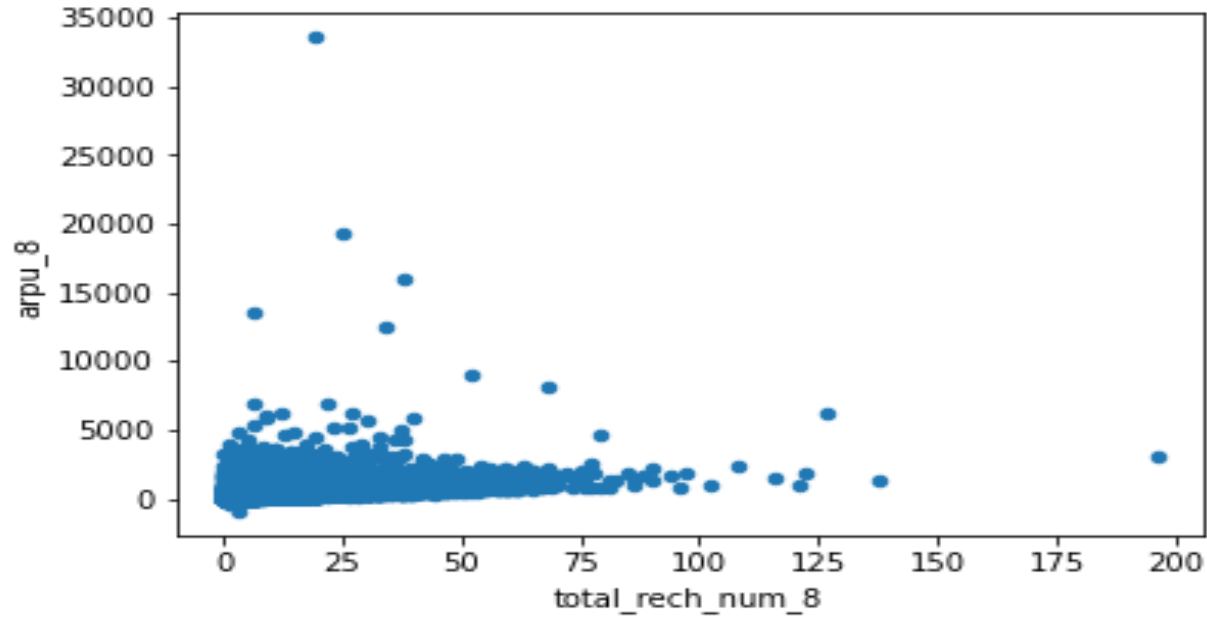
EDA Analysis



Analysis:

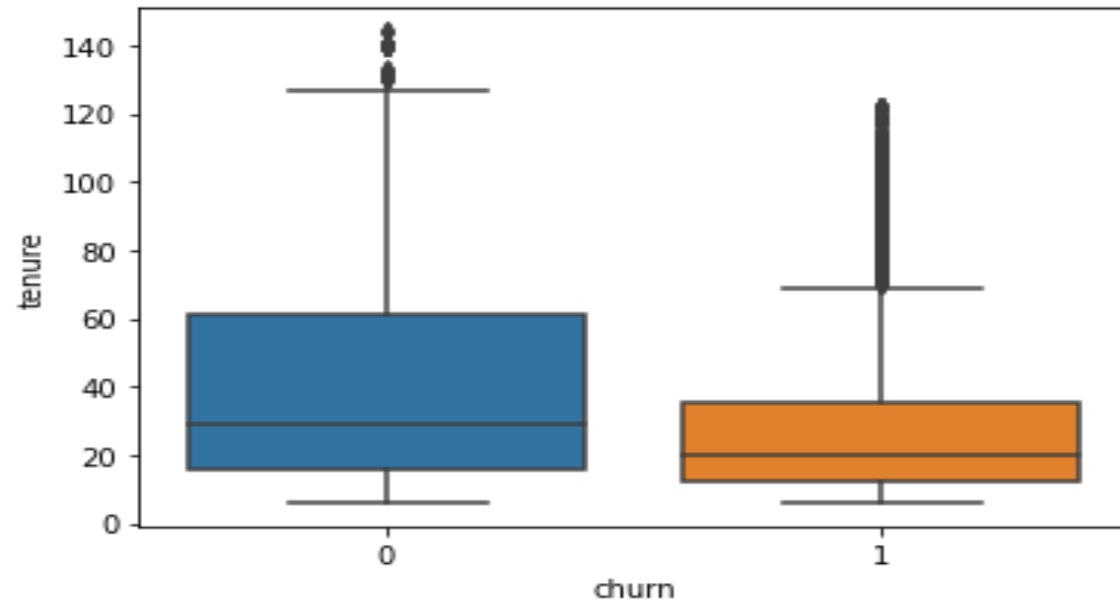
It can be seen that the maximum churn rate happens within 0-6 month, but it gradually decreases as the customer retains in the network.

The average revenue per user is good phase of customer is given by arpu_6 and arpu_7. since we have two separate averages, lets take an average to these two and drop the other columns.



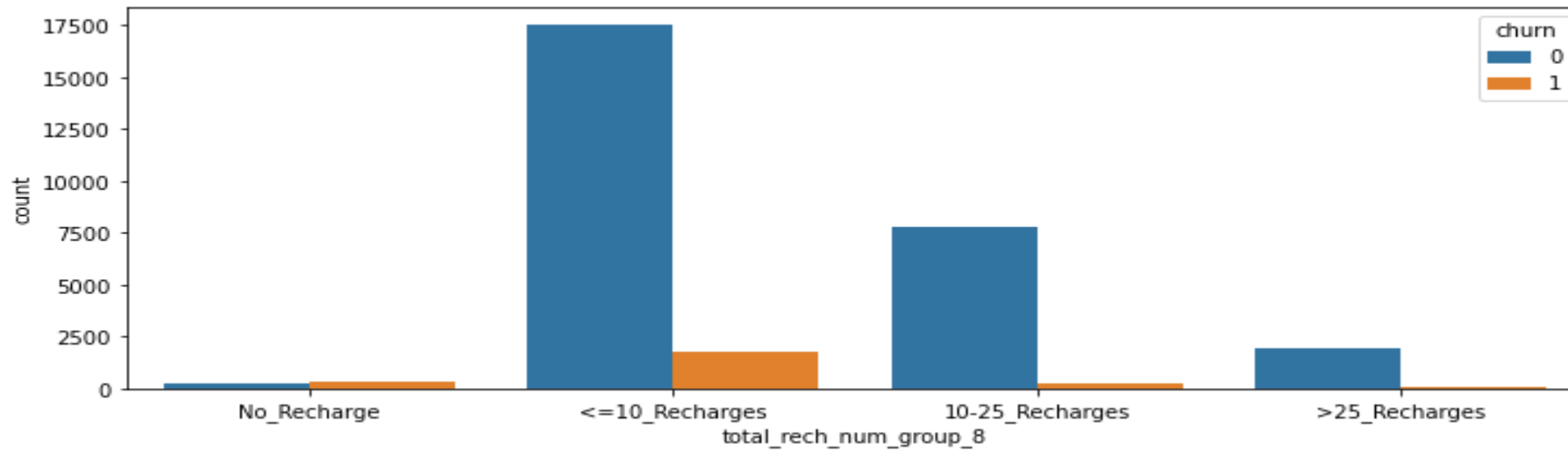
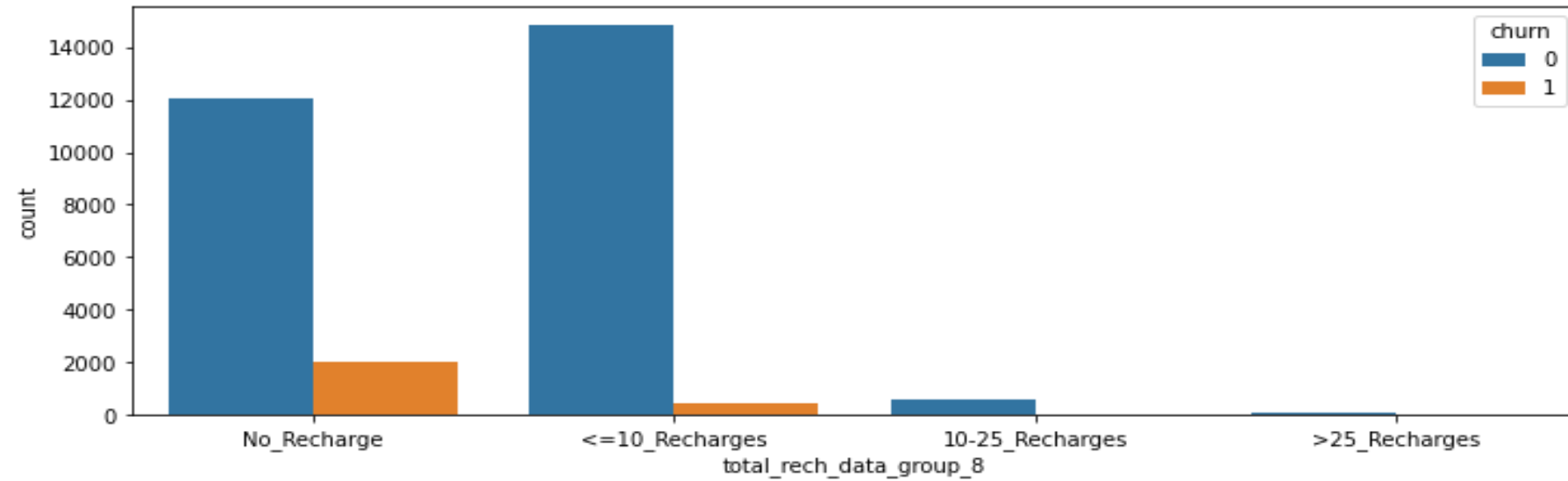
Analysis:

- Avg Outgoing Calls & calls on roaming for 6 & 7th months are positively correlated with churn.
- Avg Revenue, No. Of Recharge for 8th month has negative correlation with churn.



Analysis:

From the above plot , its clear tenured customers do no churn and they keep availing telecom services



As the number of recharge rate increases, the churn rate decreases clearly.

Conclusion:

- Given our business problem, to retain their customers, we need higher recall. As giving an offer to a user not going to churn will cost less as compared to losing a customer and bring new customer, we need to have high rate of correctly identifying the true positives, hence recall.
- When we compare the models trained we can see the tuned random forest and ada boost are performing the best, which is highest accuracy along with highest recall. So, we will go with random forest instead of adaboost as that is comparatively simpler model.

Strategies to Manage Customer Churn

The top 10 predictors are :

- loc_og_mou_8
- total_rech_num_8
- monthly_3g_8
- monthly_2g_8
- gd_ph_loc_og_mou
- gd_ph_total_rech_num
- last_day_rch_amt_8
- std_ic_t2t_mou_8
- sachet_2g_8
- aon

We can see most of the top predictors are from the action phase, as the drop in engagement is prominent in that phase

Some of the factors we noticed while performing EDA which can be clubbed with these insights are:

1. Users whose maximum recharge amount is less than 200 even in the good phase, should have a tag and re-evaluated time to time as they are more likely to churn
2. Users that have been with the network less than 4 years, should be monitored time to time, as from data we can see that users who have been associated with the network for less than 4 years tend to churn more
3. MOU is one of the major factors, but data especially VBC if the user is not using a data pack if another factor to look out