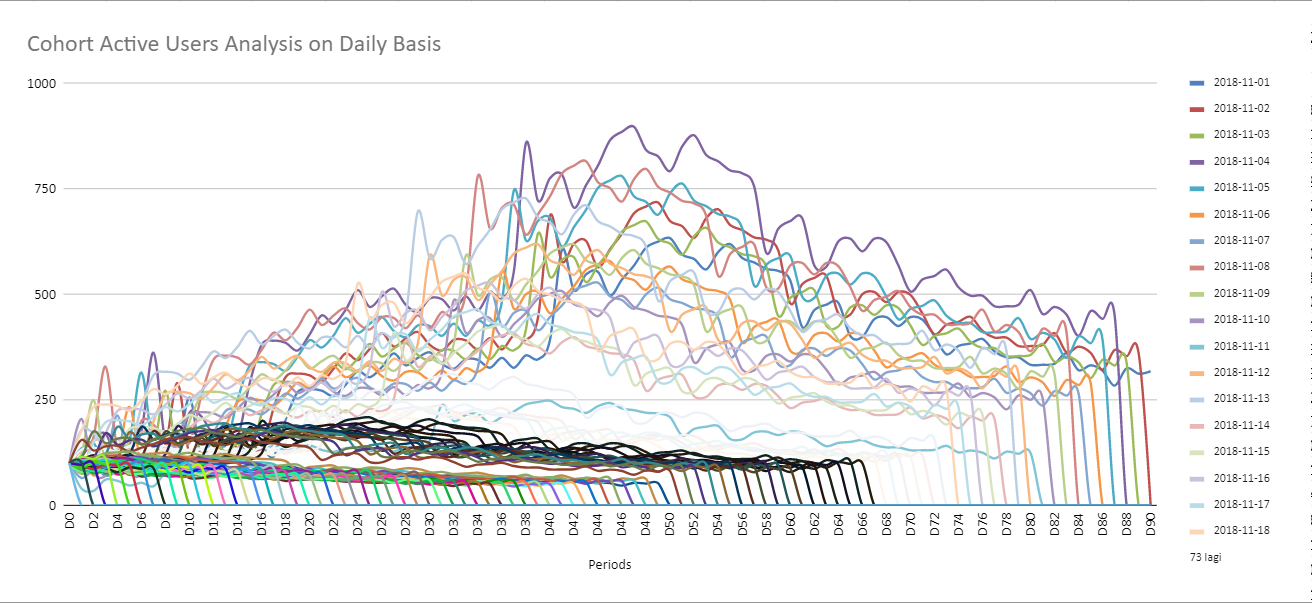
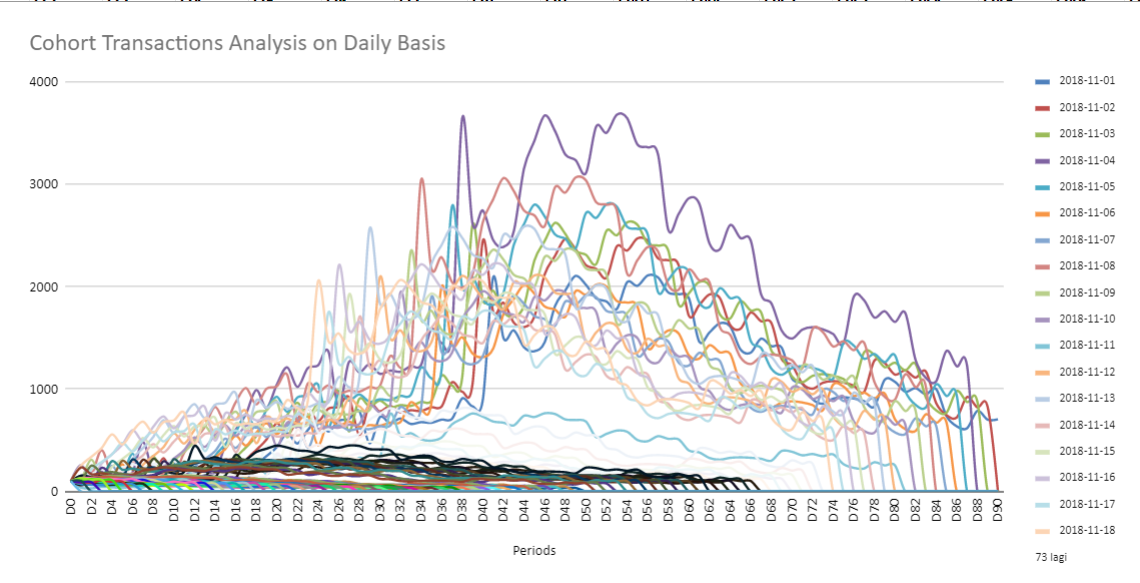
**Report of Dana’s Data Science Assignments**

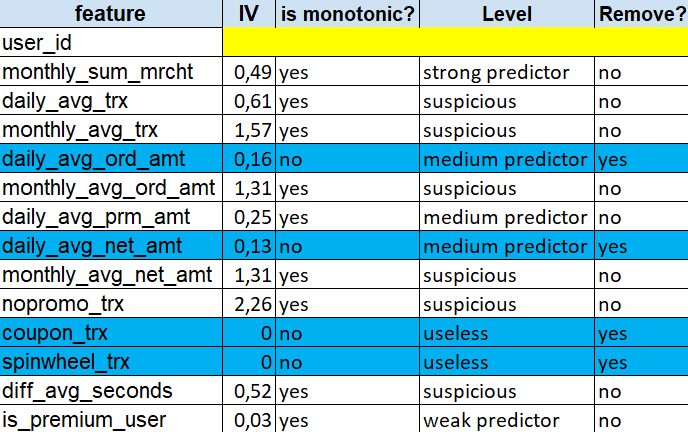
The problem on this assignment is a company wants minimizing reviving customer which are already churn or in dormant segment because it will be spending more cost than new customer, so for preventing on that situation, company wants to predict a customer who are going to churn specifically on 1st march 2019. In the assignment, flag churn is 20 days if a customer didn’t do transaction for more than 20 days since last transactions. I wonder, does 20 days flag churn is already suitable for a company or we need to redefine number of days to flag as a churn, in this analysis, I already provided the analysis with retention cohort analysis by using cohort transactions analysis and cohort active user analysis. The following graph of cohort analysis be provided below.



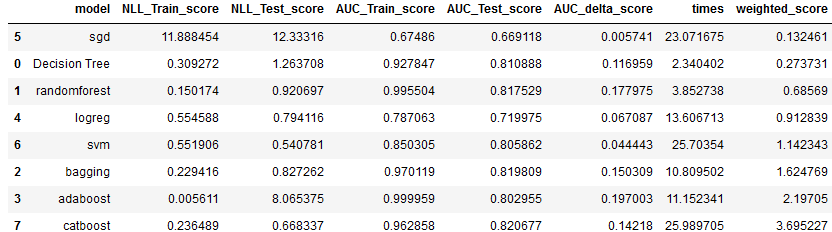


Based on cohort graph above, we can gather insight is that the retention transactions and active users have increasing trends for 40 to 50 days after their last transactions, it started to decreasing from days 51. It means that 20 days is short to flag a customer churn as we can see on the graph the retention transactions and active user will be dropped stars from days 51, I think based on graph above, we could redefine flag churn if a customers didn’t do transaction for more than 51 days.

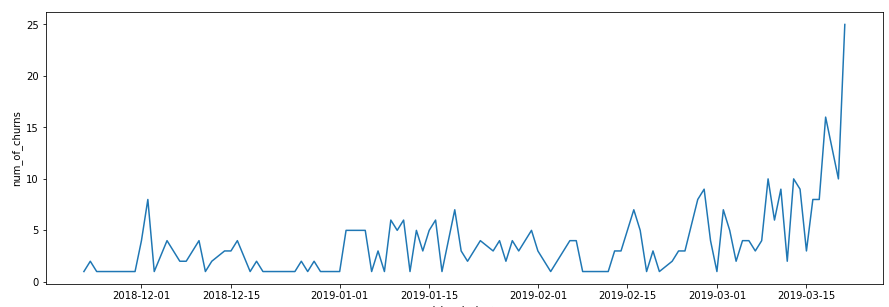
Moving on to the request of the assignment is predict a customer who are going to churn specifically on 1st march 2019. There were several steps to predict customer churn by building models. First, aggregate transactions data into user level to create new features or we called it feature engineering (already provided on the notebook). Second, Create flag churn in training and testing dataset, training dataset flag churns based on last\_created\_time on order\_data minus last\_date on January means 2019-01-31, while testing dataset using last\_trx\_data on test\_data minus February 28, then we got a flag churn both dataset (training and testing). Third, do some feature selection by applying optimum bining, to find which features can distinctive population of churn and non-churners, here is the result analysis for every features as follow:



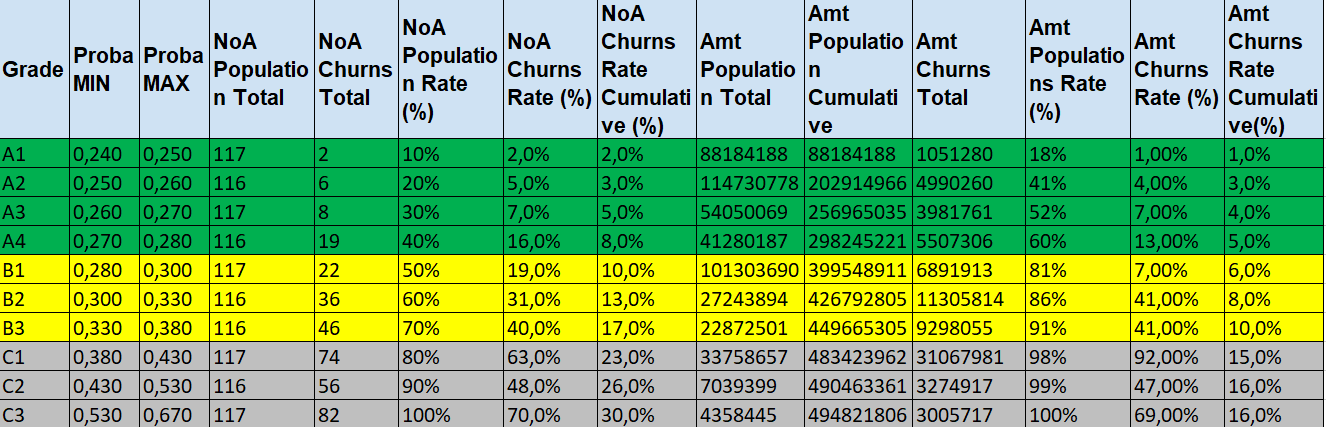
The Blue colors on left table above denoted features that should be removed when build a model because based on optimum bining, it is no monotonic (decreasing or increasing) and Information Value is small. Forth, Build some modeling for choosing which one is the best, here is the summary model as follow:



I select svm model as final model because it has consistency score in both evaluation score ( NLL and ROC). Fifth, Predict date flag churn for each customer, here is the graph of date churn for each customer as follow:



For overall predicted churn on testing data around 98 customers which divided for each date, but for specifically churn on 1st march, it will be only 1 customer who are going to churn. Lastly, it is additional analysis to which segment that should be retain or not by applying risk rank table as follow:



Based on table above, The green one, we won’t offer them with retained product, but we can give them a product for upgrade level. The grey one, we are not going to offer this segment as we need to customized product for each customer on this segment to make customer retain, but it will be spending more money. I select the yellow ones to get offered of retain product because churns rate for each segment is still less than 50%.