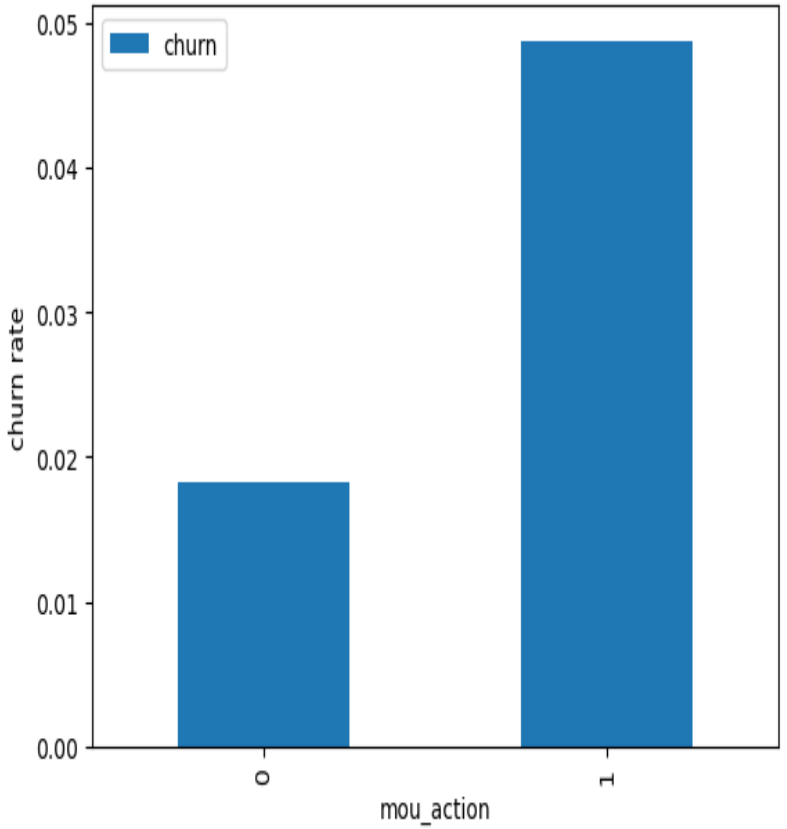
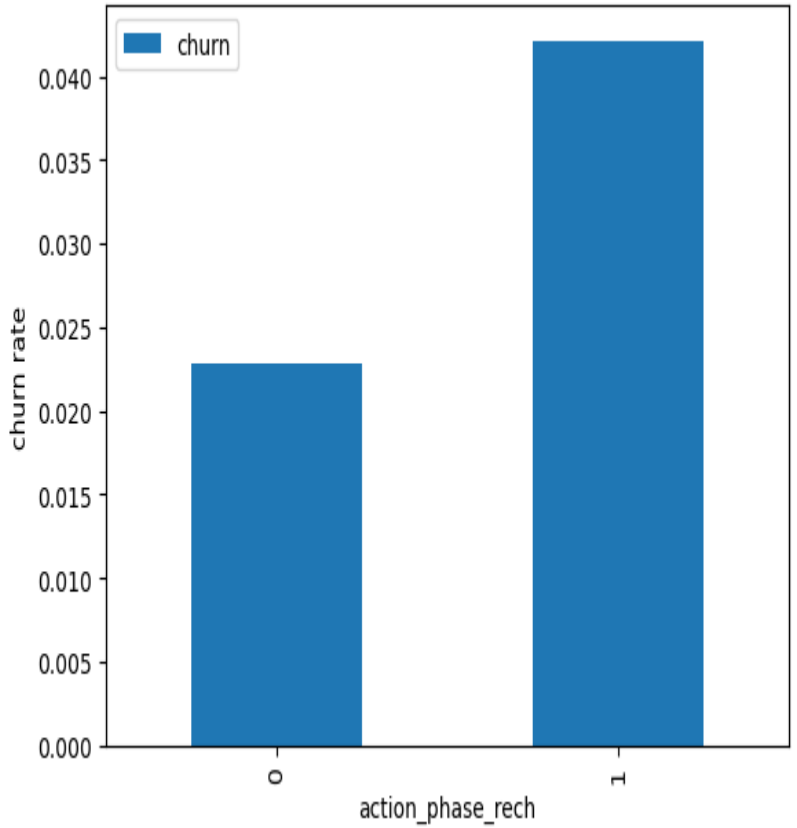
Telecom Churn Case Study

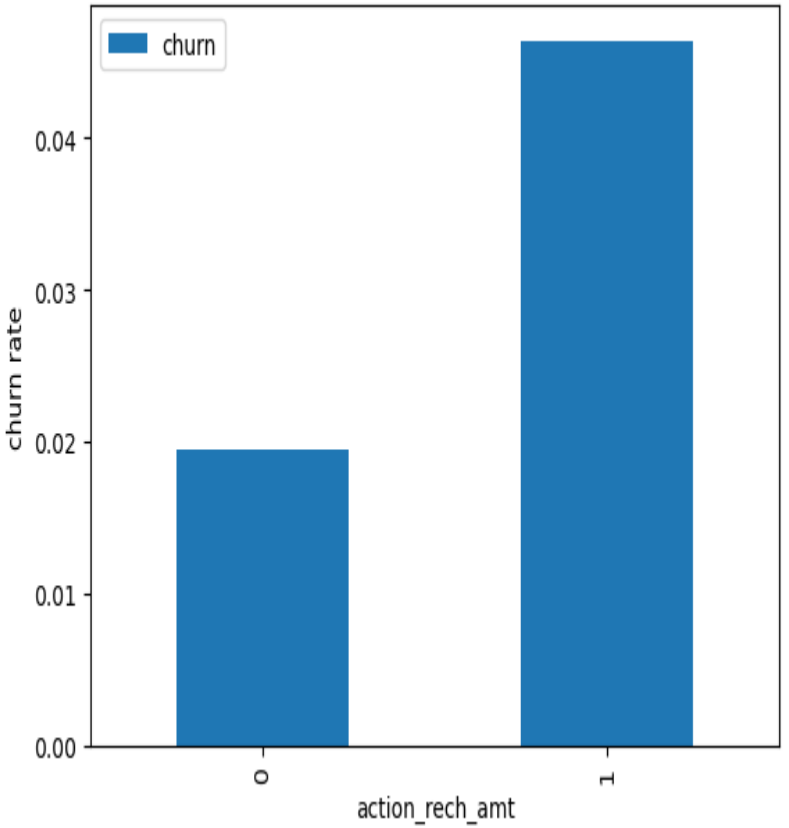
By: Suraj Das, Ruchir Rastogi, Robinson Kotamraju

Exploring Initial Data

- We can see that the churn rate is more for the customers, whose minutes of usage(mou) decreased in the action phase than the good. (Img 1)
- As expected, the churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase. (Img 2)
- Also we see the same behavior. The churn rate is more for the customers, whose amount of recharge in the action phase is lesser than the amount in good phase. (Img 3)

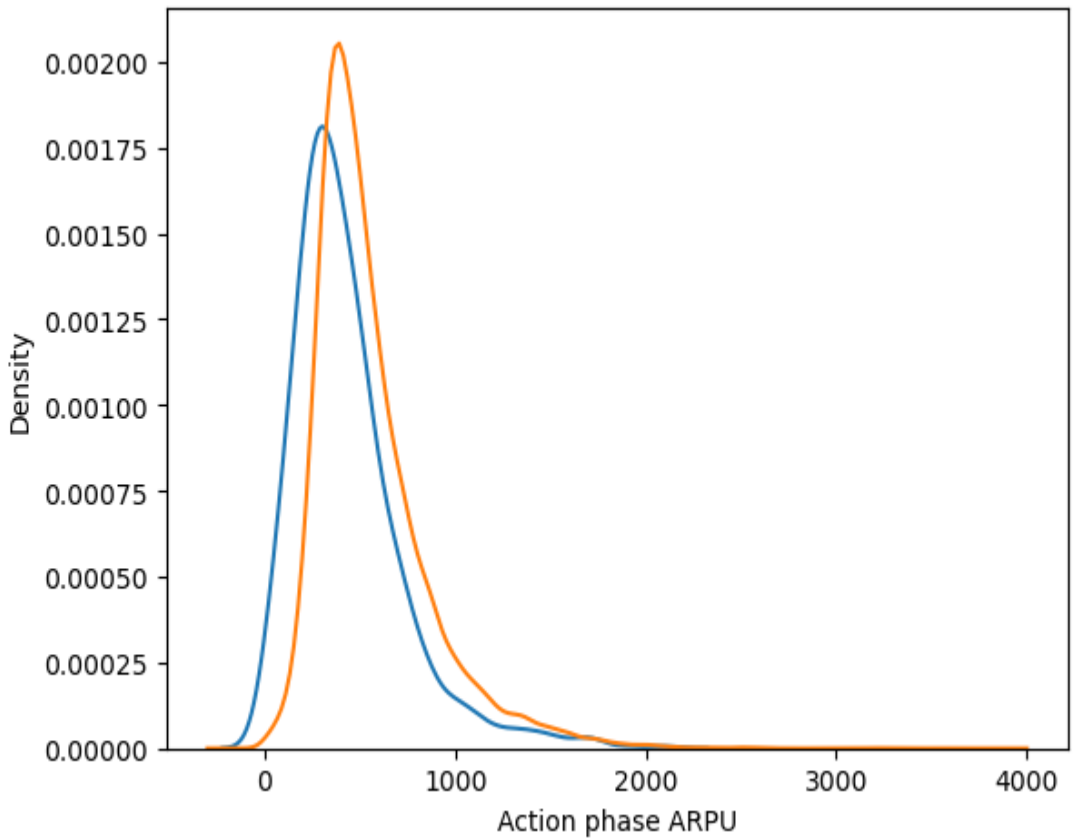


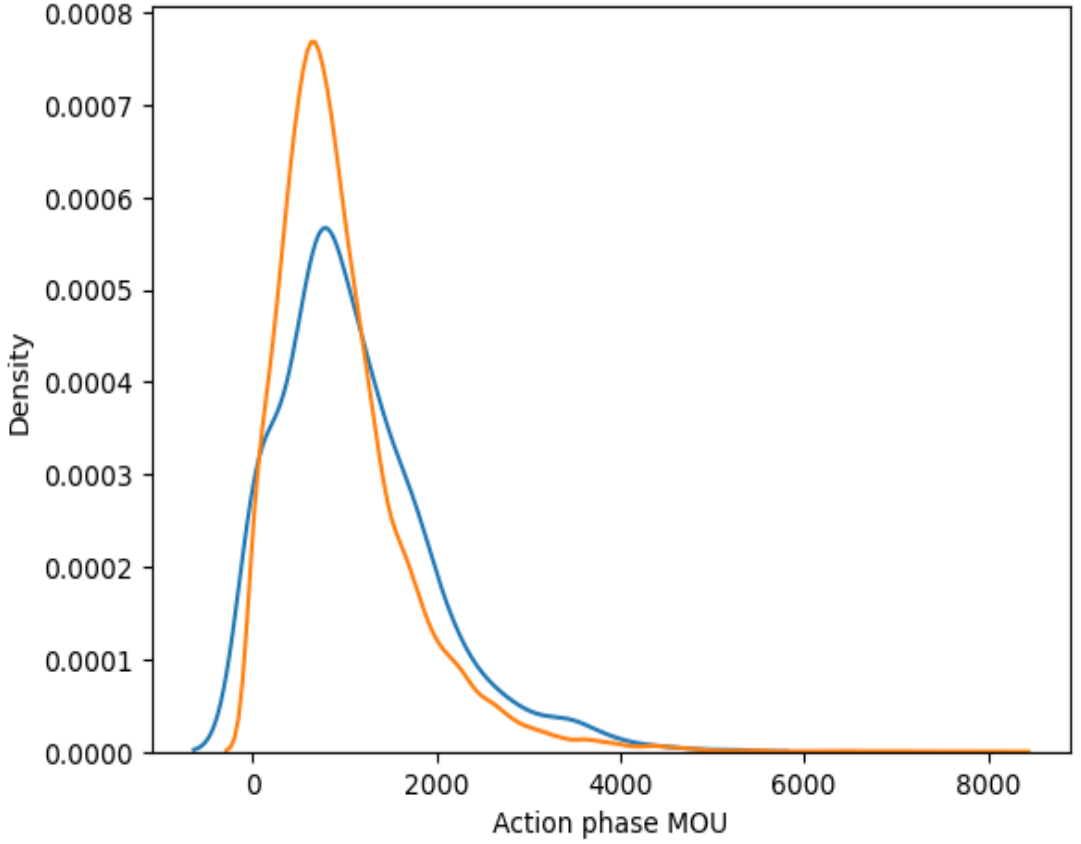




Exploring Initial Data

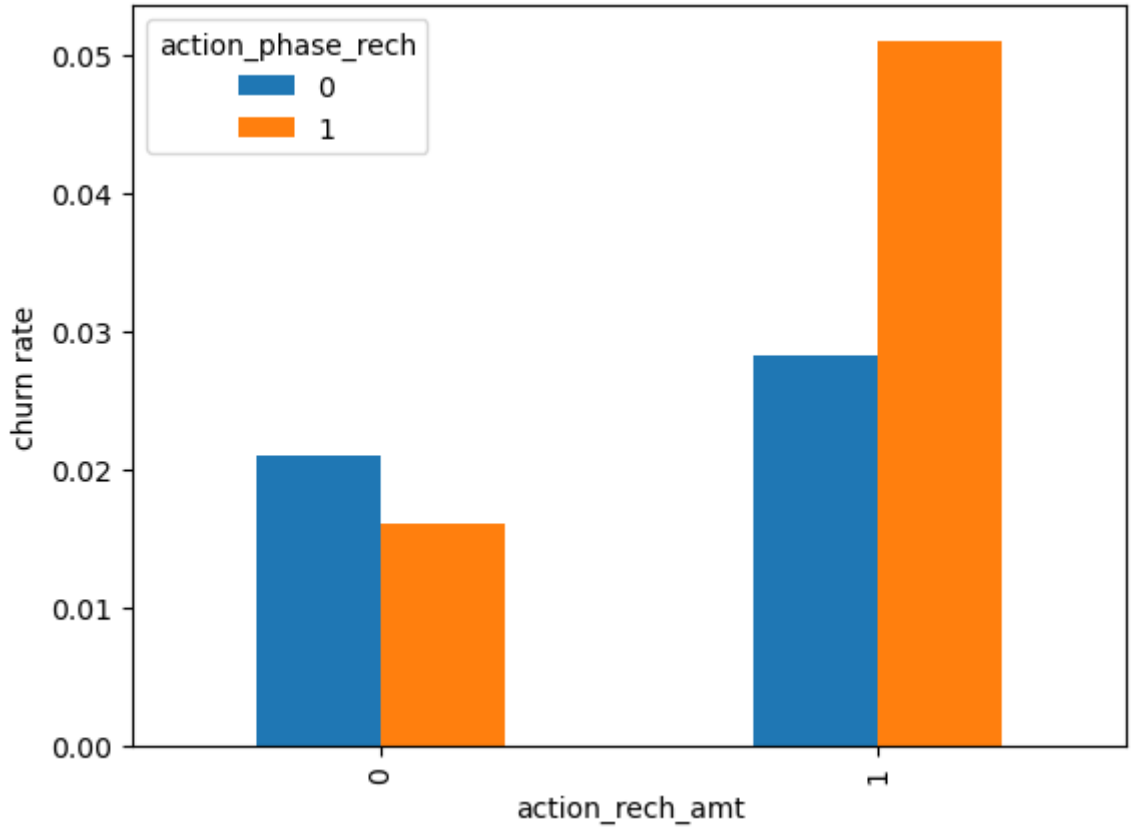
- Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned. ARPU for the not churned customers is mostly densed on the 0 to 1000.
- Minutes of usage(MOU) of the churn customers is mostly populated on the 0 to 2500 range. Higher the MOU, lesser the churn probability.





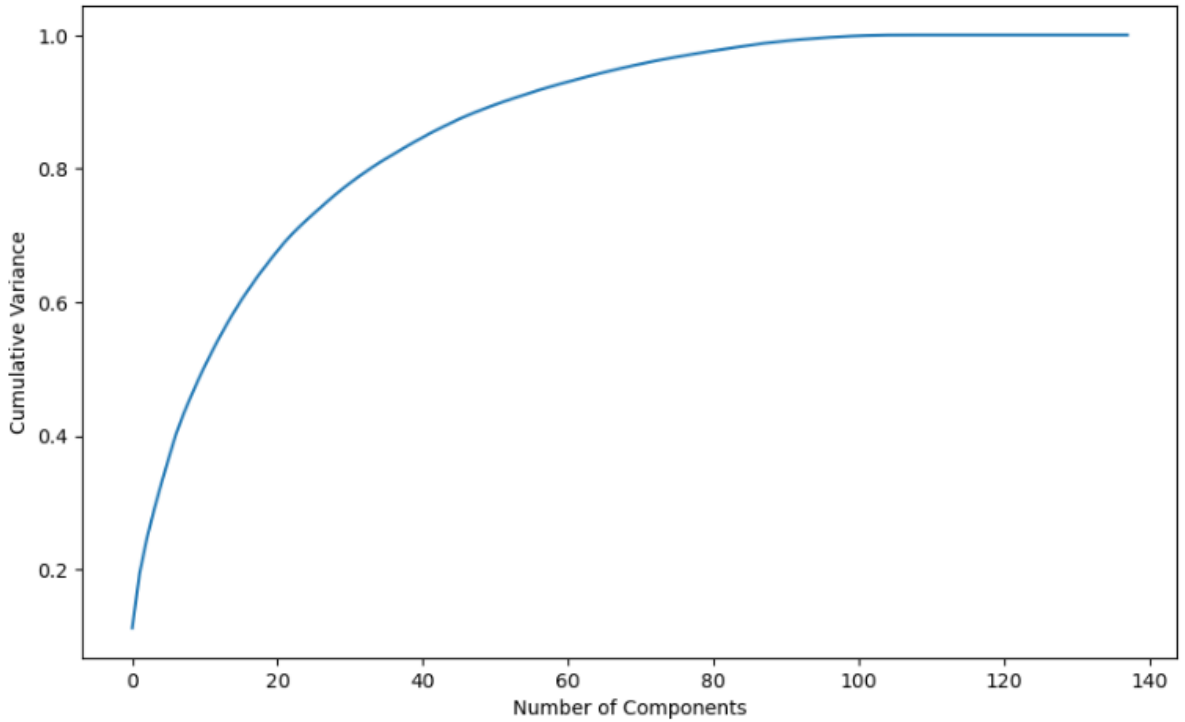
Bivariant Analysis

- We can see from the above plot, that the churn rate is more for the customers, whose recharge amount as well as number of recharge have decreased in the action phase than the good phase.



LR Model with PCA

- We can see that 60 components explain almost more than 90% variance of the data. So, we will perform PCA with 60 components.



Models with PCA

- Two different models were tried and we can see that for achieving the best sensitivity, which was our ultimate goal, the classic Logistic Regression model performed well. For both the models the sensitivity was averaged at 75%. And a good accuracy averaged at 85%.

Model	Prediction Train Data Set			Prediction Test Data Set		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Logistic Regresstion With PCA	0.86	0.89	0.83	0.83	0.81	0.83
Decision Tree With PCA	0.9	0.91	0.88	0.86	0.69	0.86

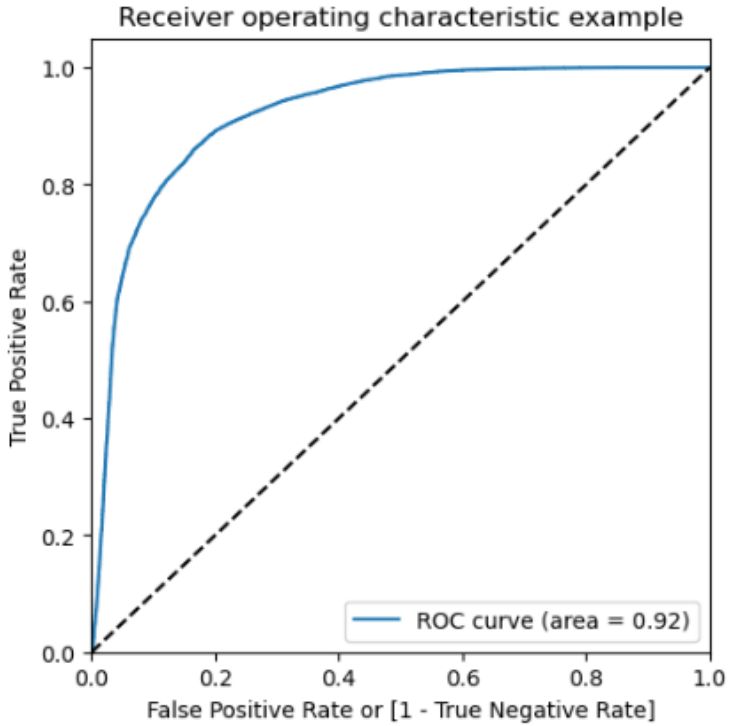
LR Model w/o PCA

- The LR model we build seems to be a best fit with Accuracy averaged at 81%, Sensitivity averaged at 85% and Specificity averaged at 79%

Model	Prediction Train Data Set			Prediction Test Data Set		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Logistic Regresstion W/o PCA	0.84	0.89	0.79	0.78	0.82	0.78

ROC Curve

- The AUC score is 92% which is further a testament of how excellent the model is and also indicates that it was worth the test.



Recommendations Based on Predictive Analysis

- Customers with less usage of local incoming and ISD outgoing calls in the action phase should be targeted
- Provide special offers to customers having their charges for value added services go up. They are most likely to churn
- Target customers with monthly 3g recharge go up during the action phase
- Customers with lesser monthly 2g usage during the action phase are likely to churn
- Customers whose roaming outgoing minutes have increased will most likely churn (given its coefficient is positive)

Thank you!