Customer Churn Prediction

Data Science *in collaboration with* Product Management

We reactivate less than 30% users churning every week

Need to identify customers at the risk of churn and understand reasons why they disengage

	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
WAU or MAU	2.85M	2.6M	2.66M	2.8M	2.89M	2.79M	2,911,077
DAU/WAU or DAU/MAU	20.82%	23.51%	23.32%	23.90%	23.91%	23.60%	24.51%
Inactive Users							
30D	792K	772K	822K	886K	918K	762K	809K
90D	351K	338K	371K	391K	406K	386K	453K
180D (Lapsed)	271K	270K	263K	258K	270K	264K	288K
Reactivated Users							
30D	537K	466K	499K	531K	582K	526K	550K
90D	197K	171K	178K	186K	198K	180K	189K
180D	79K	72K	74K	79K	84K	77K	81K

AGENDA

Churn Prediction Algorithm

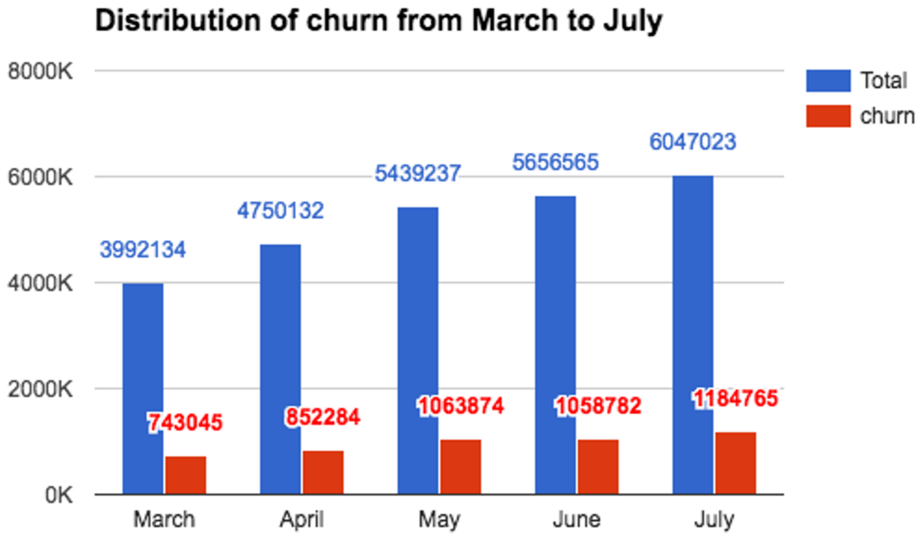
Goal : Understand reasons why customers churn, identify customers who are at the risk of churn and identify potential levers which we can use to reduce churn

Approach :

1. Associate a churn probability score for each customer
2. For churned users, find a set of reasons for their churn / disengagement
3. For top reasons that lead to churn, work across teams to provide preferential treatment to users at risk of churn

How big is the Churn problem?

A customer is considered to have churned if no booking is completed in the last 180 Days



Ride Frequency	% Share of Churners	% Users churning (LF/HF/MF group)
LF (1-2)	78%	26.5%
MF (3-8)	19%	11%
HF (>=9)	3%	**5.3%**

Consistently around 20% of our monthly user base is churning out, **5.3% of the HF base** in March and July churned out (no completed booking in the next 180 days)

Low frequency users who churned in May (9L users)

Observations

These users might not have any need of regular travel and because of low # of rides taken it is difficult to attribute them as churned. These users have thus been excluded from the analysis.

Bookings Distribution (Pan India)

These percentages are taken at 4 months avg (May, June, July, Aug 2016)

Category	%Total Users	% Bookings
LF (1-3)	57%	37%
MF (4-8)	32%	20%
HF (>=9)	11%	**43%**

HF Users: Disengaged Vs. Churned

HF users in *Aug* : **5.4L**

Churn: 180D lapsed (0 bookings in 180 Days) : **30.2K**

Disengaged : Engagement* falls by >= 80%

Category	# Users	% Users
Engaged	401240	73.4%
Disengaged	**144830**	**26.5%**
Churned	30230	5.5%

Percentile users	Fall in engagement(180D)
0%	0
10%	0
20%	0.0
30%	0.24
40%	0.41
50%	0.55
60%	0.67
70%	0.77
73%	0.80
80%	0.87
90%	0.93
100%	1

* Engagement is calculated as average monthly bookings in next 6 months

Model Features

Customer profile

1. Time_since_signup
2. # Completed rides
3. Preferred category
4. % Category rides
5. Avg. ATA-ETA
6. Avg_Distance per ride_3mnths

Availability

1. Stockouts

Classification Model

Pricing / ATS

1. Blended Rate_3mnths
2. ATS_3mnths
3. Last ride peak multiplier
4. % Peak rides taken

Cancellation

1. % Driver cancellations
2. % DDD
3. % Customer cancellations

Customer Feedback

1. Avg. Ride Rating
2. Calls to customer care

Prediction Model(s)

Model V0

Regularised Logistic Regression	
Train Data	: Total 203172
Test Data	: Total 163821
Precision	: 62%
Recall	: 51%

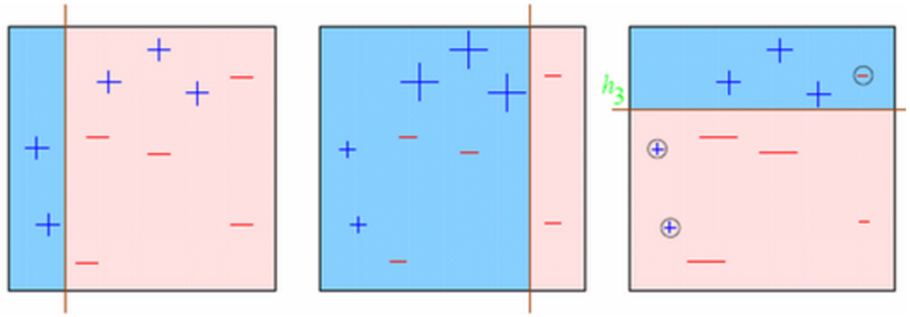
Model V1

Gradient Boosting (XgBoost)	
Train Data	: Total 203172
Test Data	: Total 163821
Precision	: 70%
Recall	: 51%

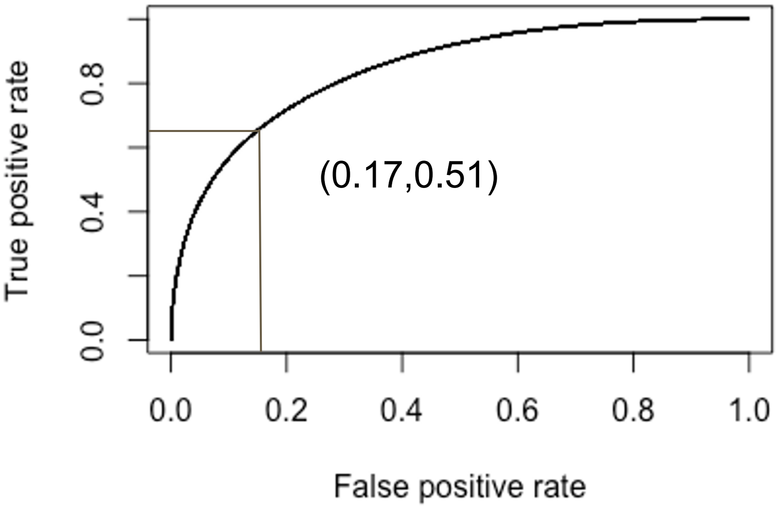
To remove high class imbalance we under sample engaged users so that we get Engaged and Disengaged in 1:1 ratio.

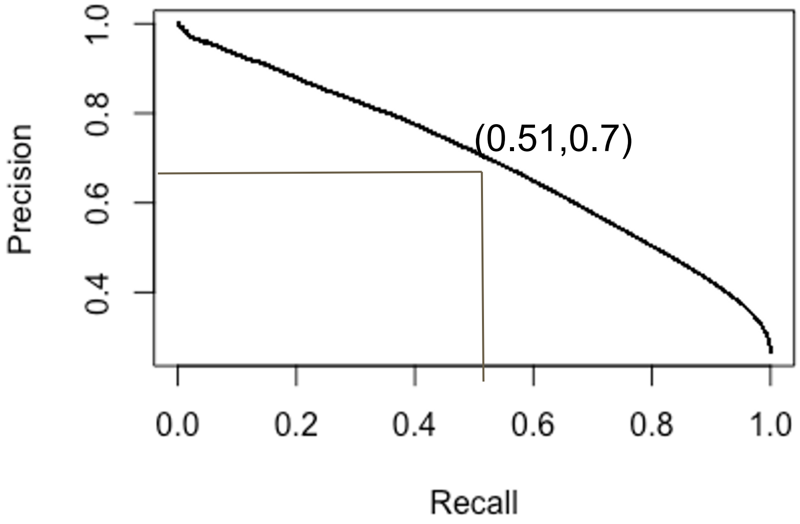
Why XgBoost?

- Sequential technique which works on the principle of ensemble. It combines a set of **weak learners** and delivers improved prediction accuracy.
- Memory optimisation to make it run faster.
- Works better with higher dimension data.



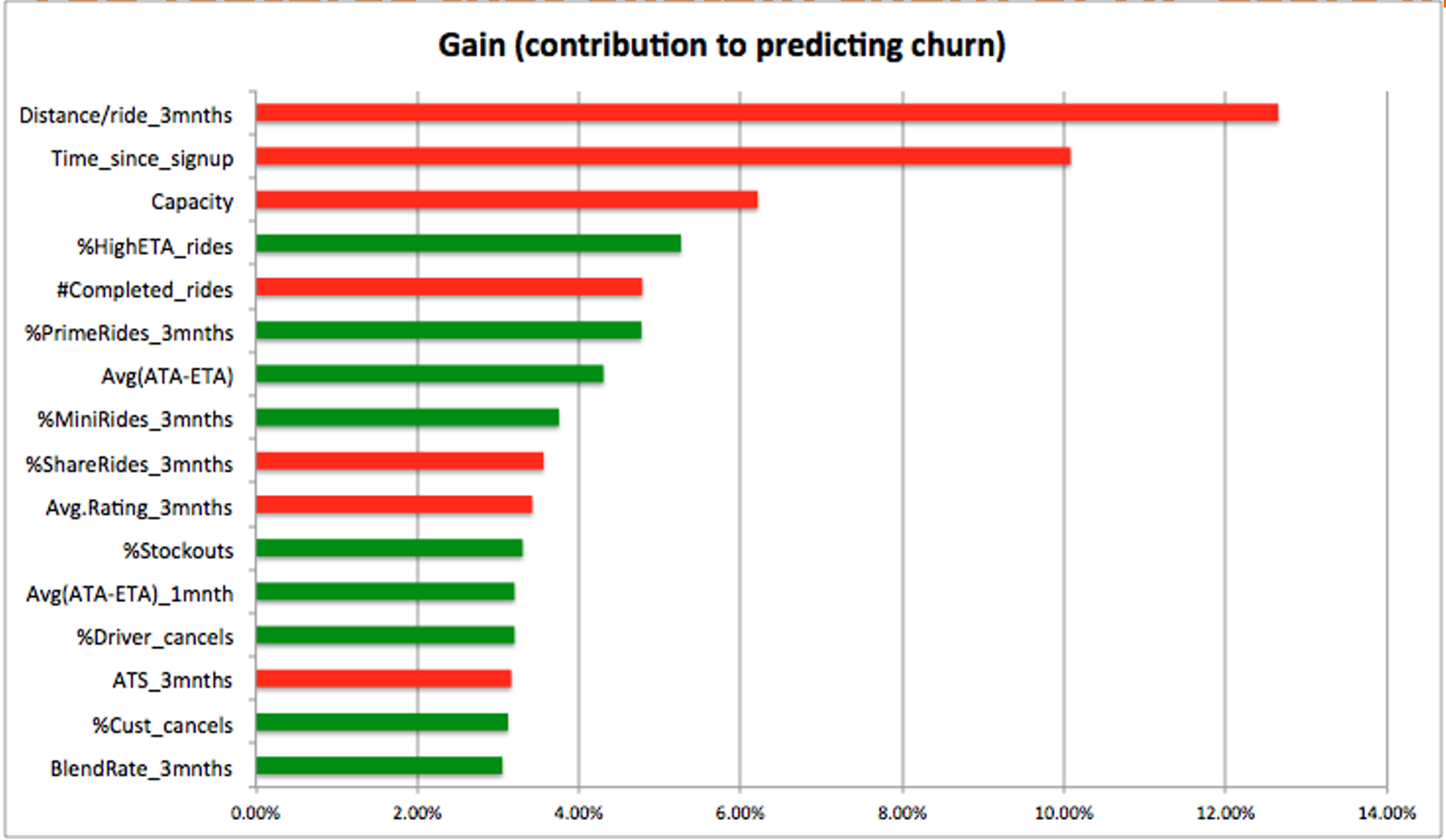
ROC curve and PR curve of XG boost Model





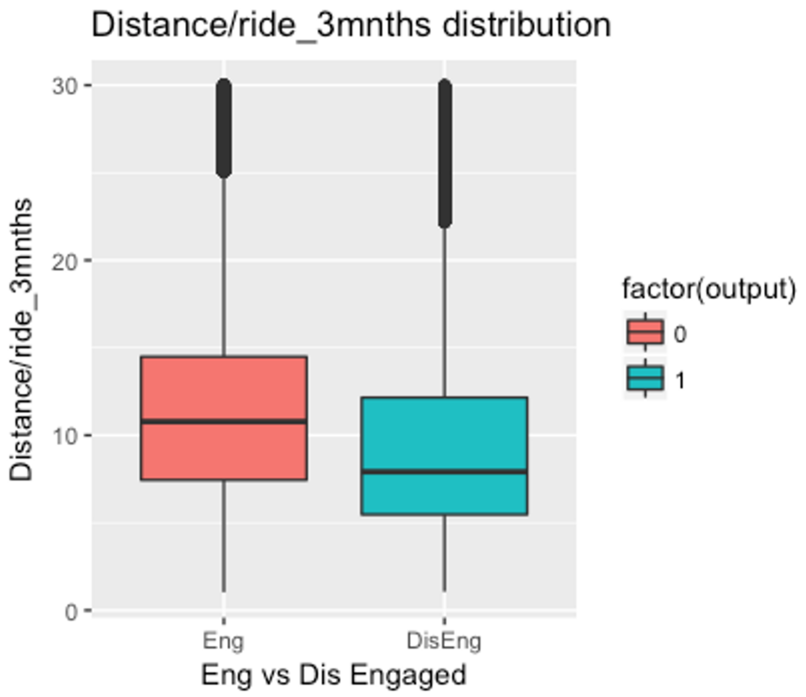
***Precision** is the fraction of retrieved instances that are relevant, while **Recall** (also known as sensitivity) is the fraction of relevant instances that are retrieved.

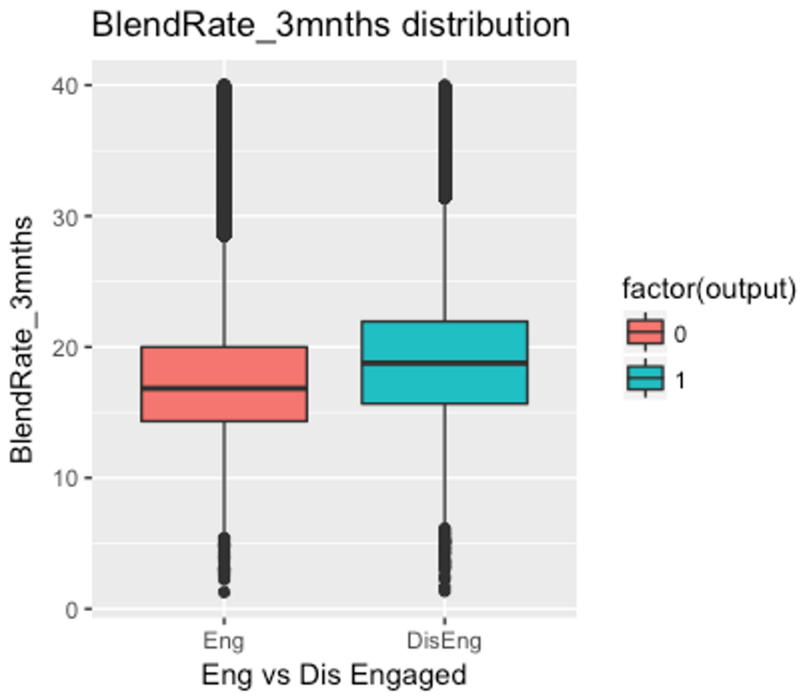
Top features that explain churn of HF users in Aug



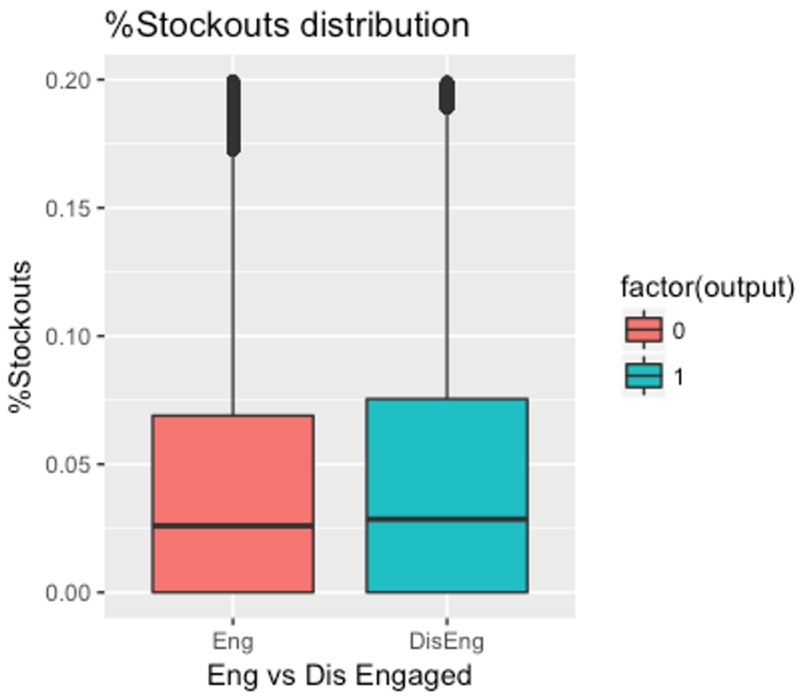
Feature distribution Analysis

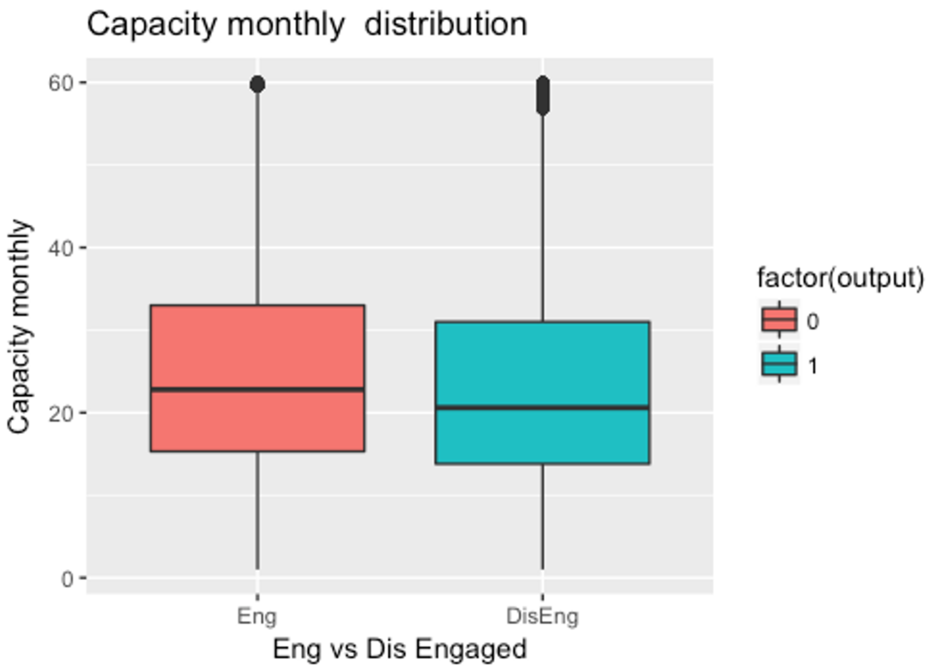
HF Users who churn have higher # of total bookings and lower cancelled / completed %



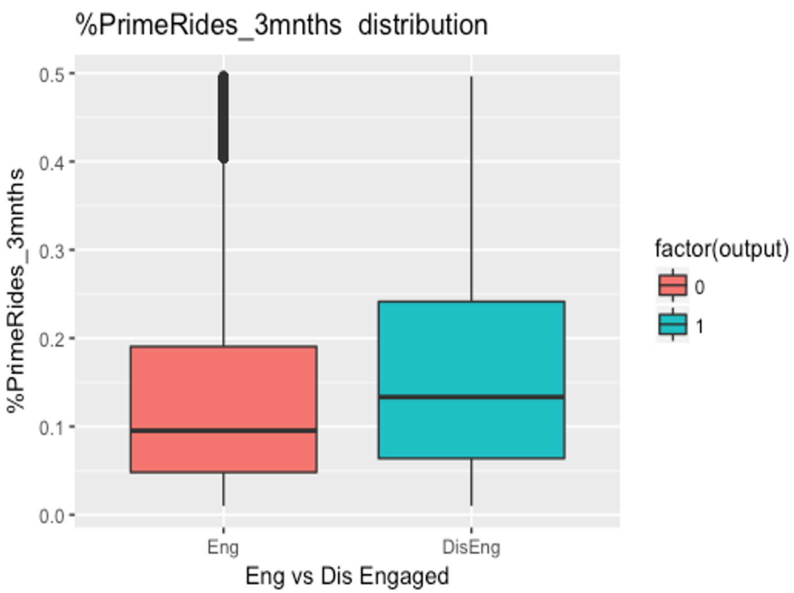


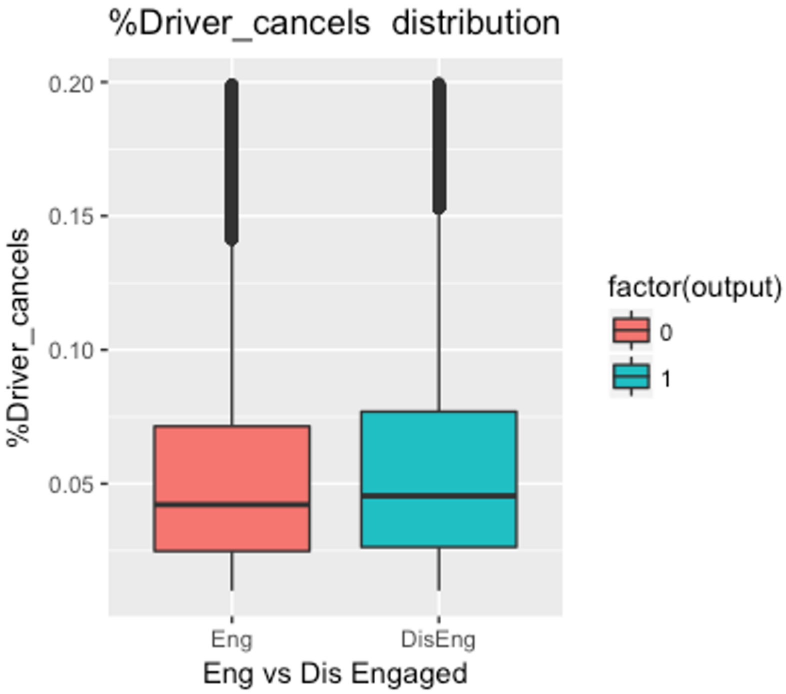
Feature distribution Analysis cont.



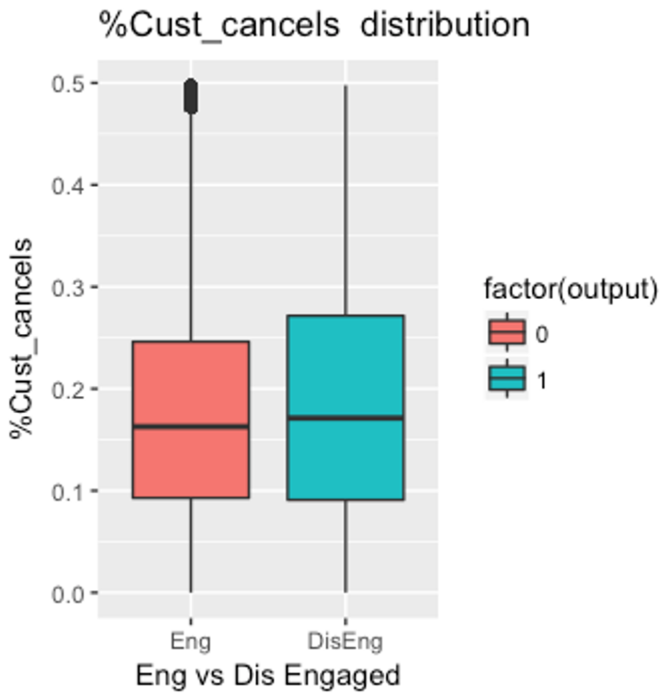


Feature distribution Analysis cont.





Feature distribution Analysis cont.



July month feature ranking	
Feature	**Gain**
Time_since_signup	9.26%
Distance/ride_3mnths	9.17%
Capacity	5.16%
ATS_3mnths	4.96%
#Completed_rides	4.90%
Avg(ATA-ETA)	4.54%
%MicroRides_3mnths	4.27%
%HighETA_rides	4.05%
%MiniRides_3mnths	3.92%
%PrimeRides_3mnths	3.71%
%Driver_cancels	3.58%
BlendRate_3mnths	3.51%
%Cust_cancels	3.50%
Avg(ATA-ETA)_1mnth	3.32%
%ShareRides_3mnths	3.30%
%Stockouts	3.28%
%Cust_cancels_1mnth	3.16%
%Canceled/Completed	3.16%
%Rides_Rating<3	3.06%
Avg.Rating_3mnths	2.90%
%Driver_cancels_1mnth	2.76%
%Stockouts_1mnth	2.51%
Calls / ride	2.39%
%DDD_1mnth	1.93%
Calls/ride_1mnth	1.27%

Aug month feature ranking	
Feature	**Gain**
Distance/ride_3mnths	12.65%
Time_since_signup	10.08%
Capacity	6.21%
%HighETA_rides	5.26%
#Completed_rides	4.78%
%PrimeRides_3mnths	4.77%
Avg(ATA-ETA)	4.30%
%MiniRides_3mnths	3.75%
%ShareRides_3mnths	3.56%
Avg.Rating_3mnths	3.42%
%Stockouts	3.30%
%Driver_cancels	3.20%
Avg(ATA-ETA)_1mnth	3.20%
ATS_3mnths	3.16%
%Cust_cancels	3.12%
BlendRate_3mnths	3.05%
%Canceled/Completed	2.93%
%MicroRides_3mnths	2.91%
%Cust_cancels_1mnth	2.63%
%Driver_cancels_1mnth	2.42%
%Rides_Rating<3	2.37%
%Stockouts_1mnth	2.15%
%DDD_1mnth	1.95%
Calls / ride	1.77%
Calls/ride_1mnth	1.01%

May month feature ranking	
Feature	**Gain**
Time_since_signup	13.57%
Distance/ride_3mnths	8.70%
Capacity	5.63%
ATS_3mnths	4.88%
BlendRate_3mnths	4.44%
%PrimeRides_3mnths	4.30%
Avg(ATA-ETA)_1mnth	4.13%
%HighETA_rides	4.01%
%MiniRides_3mnths	3.80%
%MicroRides_3mnths	3.77%
%Canceled/Completed	3.70%
%Cust_cancels_1mnth	3.70%
#Completed_rides	3.67%
%Cust_cancels	3.45%
Avg(ATA-ETA)	3.43%
Avg.Rating_3mnths	3.42%
%Driver_cancels	3.03%
%Rides_Rating<3	2.96%
%ShareRides_3mnths	2.78%
%Driver_cancels_1mnth	2.70%
%Stockouts	1.89%
Calls / ride	1.82%
#Peakrides_last9rides	1.81%
%Stockouts_1mnth	1.74%
Multiplier_lastpeakride	1.33%

Revenue & Bookings Impact

Average Monthly Revenue(June,July,Aug) of users who got Disengaged (Aug month users) = ~ Rs 27.4 cr

Expected in next 6 months = 27 * 6 = ~ Rs 164.4 cr

Revenue received in next 6 months = Rs 15.2 cr

Revenue Lost in next 6 months = ~ Rs 149.2 cr

Potential revenue can save in six months* = 0.5 (model accuracy) * 149.2
= ~ Rs 74.6 cr

Average Monthly Bookings(June,July,Aug) of users who got Disengaged (Aug month users) = 17.24 L

Expected in next 6 months = 17.24 * 6 = ~ 1cr

Bookings received in next 6 months = 8.5 L

Bookings Lost in next 6 months = ~ 92 L

Potential Bookings can save in six months* = 0.5 (model accuracy) * 92 L = 46 L

*This is only for users we lost in the month of Aug 2016.

DS Next Steps

- Build models for different granularities (like 60D, 90D) to predict churn
- Work with engineering to productionalize this model
  - Allow cross-functional teams to access list of users at the risk of churn in realtime to provide preferential service quality

- Add more features (e.g. Driver Segments, Location based features)
- Scale this model for MF (3-8 bookings) users and try a different approach for LF users

Product / Biz Next steps

- **Short distance riders are churning faster,** one potential reason could be higher pricing Vs competition for short rides, we might need to reduce pricing for short rides and match competition fares

- **High (ATA - ETA)** is a strong reason for customer churn, for HF users we need to show the correct ETA on booking confirmation

- **Stockouts** lead to churn amongst HF users, these users ***must get*** prioritised allotment in the Demand queue

- **# HighETA_rides** contribute to churn, for HF users, we need to prioritise allocation with a goal to minimise post booking ETA for them

Product / Biz Next steps

- Users with more lifetime rides have a lower propensity to churn, we will identify Regular users with a good history of taking rides and provide preferential service (availability, allocation & better cars / drivers)

- Since **users with high capacity churn less**, such users that show a declining engagement will be automatically sent frequency boosting offers

- For HF user rides with low ratings (<=3), we need to maintain and review a weekly summary of the top reasons mentioned for the same, and consistently act on the feedback and compensate the customer when required

- Review **top reasons for Ola Care contacts** from HF customers and publish a weekly summary on the same

Q & A