

# Attrition Analysis

*Marketing Analytics*

*December 14, 2016*

- 1 Executive Summary
- 2 Data Exploration
- 3 Methodology
- 4 Leaving Members Modeling
  - 4.1 Initial Modeling Results
  - 4.2 Final Modeling Results
- 5 Churn Modeling
  - 5.1 Initial Modeling Results
  - 5.2 Final Modeling Results
- 6 Conclusions
- 7 Appendix A - Sql queries

## 1 Executive Summary

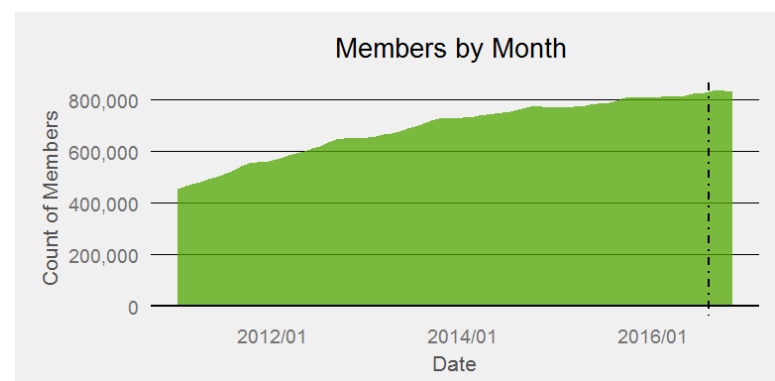
This is an analysis of attrition trends for Zipcar North America, ostensibly to provide context to whatever (if any) effect the introduction of online self-close has had on attrition. It is not intended to make any inference on the value of self-close - rather it can be used in conjunction with other data points and analyses to make a determination on the self-close initiative.

Both raw attrition and churn were examined. Time series models were fitted for each KPI and retroactively 'forecast' over the past half year to determine if either measure exhibited unexpected behavior that could potentially be associated with self close. August 1st is used as the timeframe to determine pre/post self-close behavior.

After reviewing the data it does not appear that self-close has significant causal impact on either the raw number of monthly attriting members or the churn rate. While there was a noticeable increase in both the month directly after implementation, the results are within the acceptable parameters of variance for what was expected. Additionally, subsequent months (September 2016-November 2016) show no indications of an above average trend and are actually trending under the forecast. I would be hesitant to make any judgements at this point, as there is quite a bit of noise in the attrition data in general - and would recommend gathering additional data points if a definitive opinion on self-close impact is warranted. If that is not an option, and this is the entirety of the data, I tend to lean towards August being more of an outlier than an indicator.

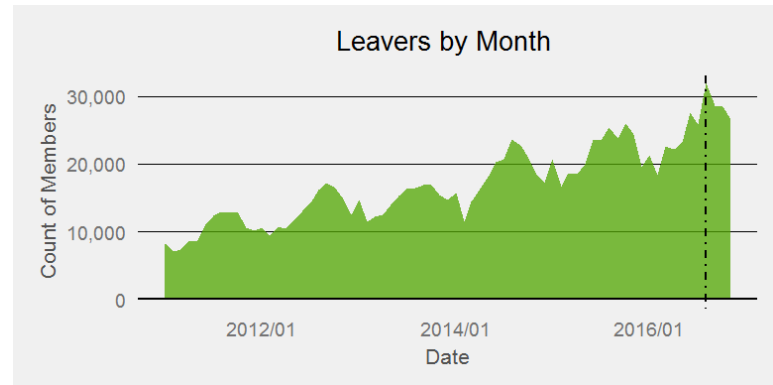
## 2 Data Exploration

The period covered in the analysis is January, 2011 through November 2016. Monthly counts of members were retrieved from the Datashed member month details view. Attriting members were derived from the Datashed account driver activity table. With the benefit of hindsight, only a member's final leave date qualified as a leaving member. As a result more recent months would be (potentially) penalized as 'habitual'



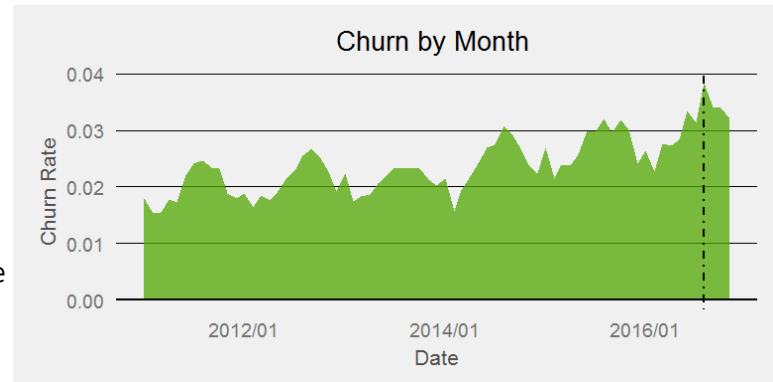
rejoiners would be discounted as attriting from earlier observational periods. If further analysis is warranted, each intermittent leave could be considered in it's relevant time frame. fig 1

Overall membership growth for North America is graphed in Figure 1. The trend line is linear in nature, although it has flattened out in recent years. There is a slight yearly seasonal trend that is generally masked by size of the established member base. The dotted black line bisecting the chart indicates August of 2016.



The number of monthly leaving members, and the monthly churn rate, are shown in the next two graphs (Figs. 2-3). Two things of note on these numbers:

1. The month to month variance (reflected in both measures) is much higher. Churn rate effectively mirrors Attriting members - this is unsurprising given the relative consistent nature of the denominator (Total members). You can see some of the variance tamped down out of the churn rate compared to leaving members as the total member base grows.
2. A definitive spike in August 2016 that matches the launch of self-close. While there are YoY growth as well as seasonal trends evident in both metrics, The increase at this point is particularly noticeable. fig 2



Given the trends indicated in the attrition data, a basic decomposition of the time series using moving averages was generated (fig 4). There are four parts to this chart

1. Observed - The time series chart of leaving members (same as fig. 2)
2. Trend - The yearly trend line for the data
3. Seasonal - The seasonal pattern of leaving members
4. Irregular - The difference between the observed data and what is accounted for by the Trend and Seasonal components. fig 3

In other words - if we had a simple two part moving average model, the fourth chart indicates how wrong the model is everyr month. Again, the largest (positive) variance is seen in August 2016. However, it is of note that for subsequent months the error rate is actually below expected totals.

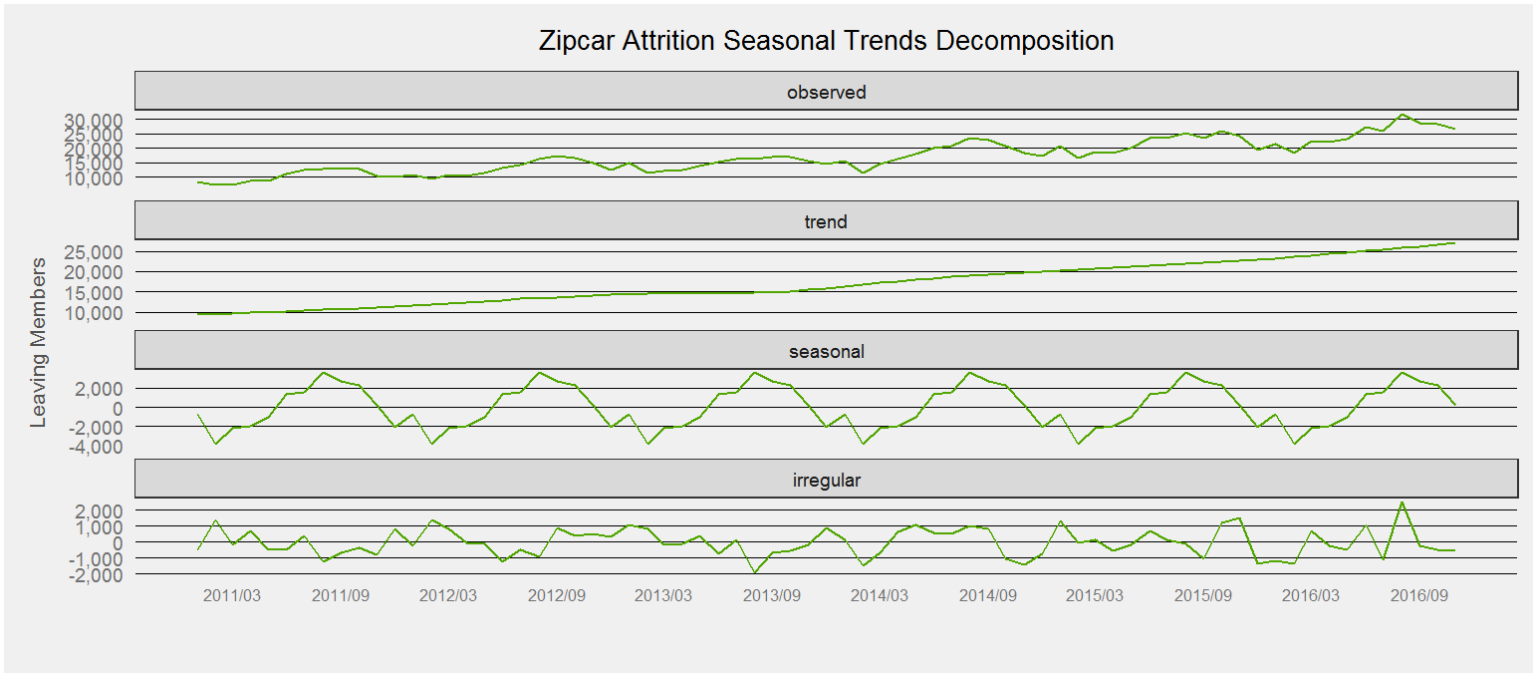


fig 4

### 3 Methodology

Given that the simple univariate moving average model explains a good portion of the variance, a series of more complex time series models are fit to see if there is a more complete model. Data from 2011 to 2014 is used to train the models, and fit is measured on the accuracy of the predictions for 2015. The best model type - using Mean Absolute Error - is chosen, and the winning model is retrained on data from 2011 through Jun of 2016. This model is then be used to predict July 2016 through November 2016, and measure actual results and error terms against prediction to look for outliers or unexpected variance. The same process will be used for Leaving members and churn rate.

### 4 Leaving Members Modeling

#### 4.1 Initial Modeling Results

The results from the initial modeling run can be seen in the table below (Fig 5). The Holt Winters Multiplicative model was the best fitting model, with a MAE of 616.7 on the training dataset and 1091.8 on the test dataset.

fig 5

Model Type	Mean Absolute Error
MAM	1091.819
MMM	1571.113
stl	1500.676
arima	1430.877

The fit of the winning model and associated confidence intervals for the test data set can be seen in Figure 6.

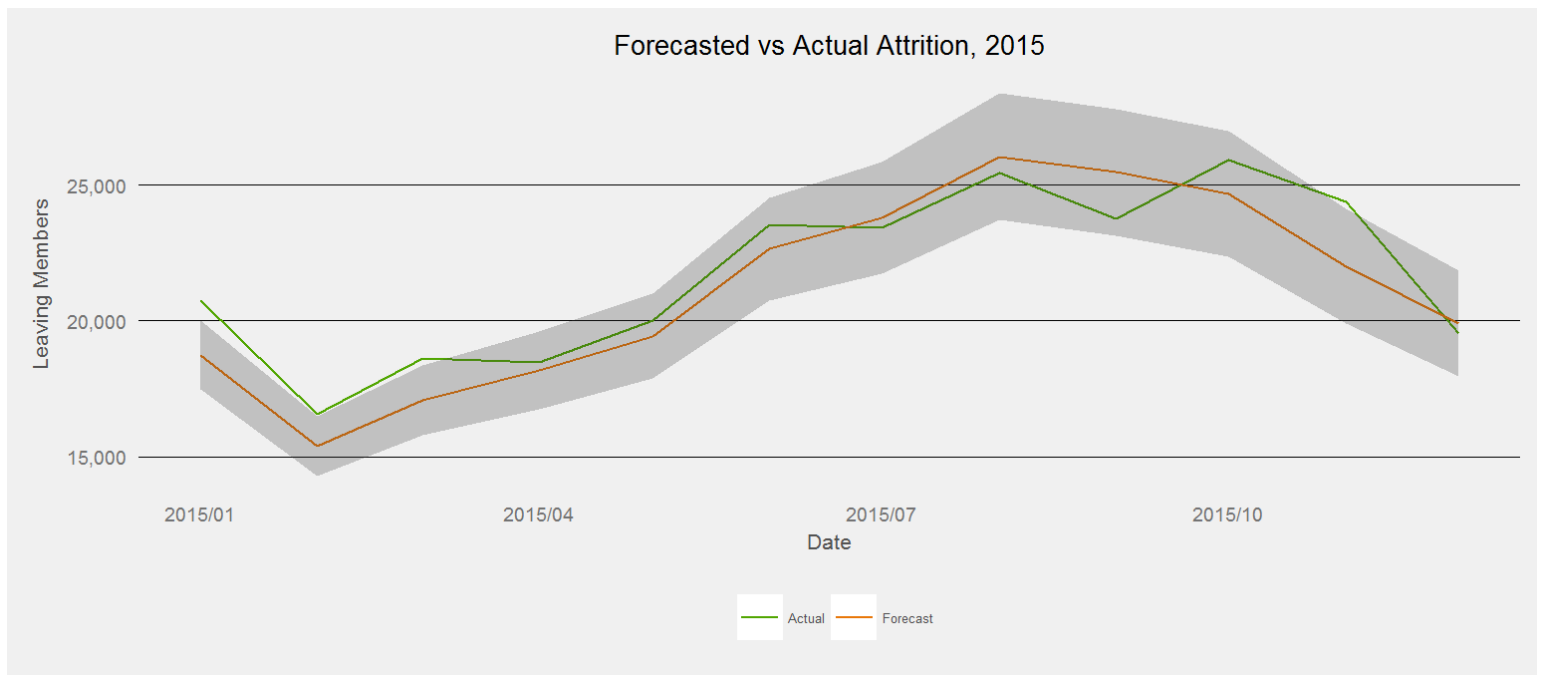


fig 6

## 4.2 Final Modeling Results

The winning model was refit and had a MAE of 707 on the training dataset. The MAE on the test dataset was 1285.6.

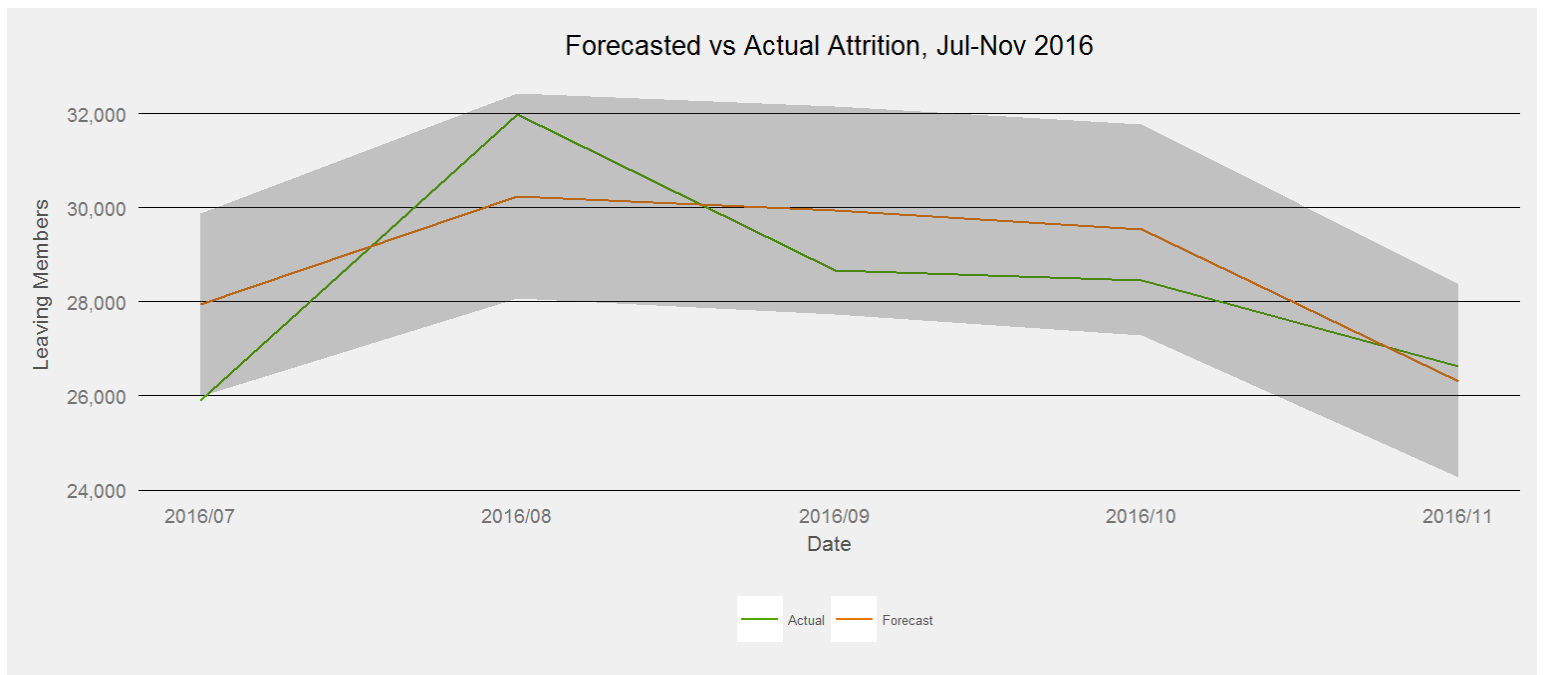


fig 7

## 5 Churn Modeling

### 5.1 Initial Modeling Results

The results from the initial modeling run can be seen in the table below (Fig 8). An additional model type was introduced into this round: a multivariate time series was used to try and draw out the effects on the change in member base composition (specifically the increase of Uni as a share of total members). An Exponential Triple Smoothing model was the best fitting

model, with a MAE of 0.0011 on the training dataset and 0.0019 on the test dataset.

fig 8

Model Type	Mean Absolute Error
MAM	0.0030662
MMM	0.0019489
stl	0.0028639
arima	0.0028129
multi	0.0077591

The fit of the winning model and associated confidence intervals for the test data set can be seen in Figure 9

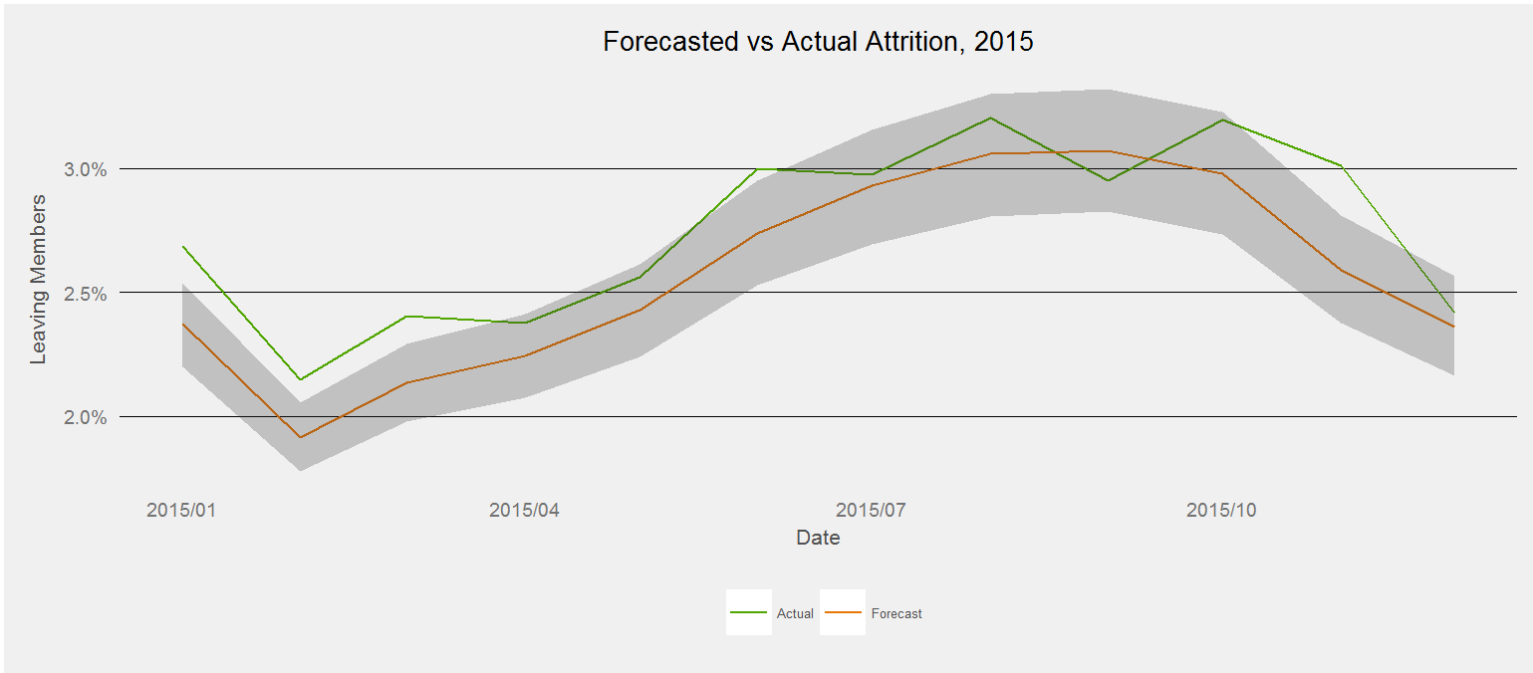


fig 9

## 5.2 Final Modeling Results

The winning model was refit and had a MAE of 0.11% on the training dataset. The MAE on the test dataset was 0.16%.

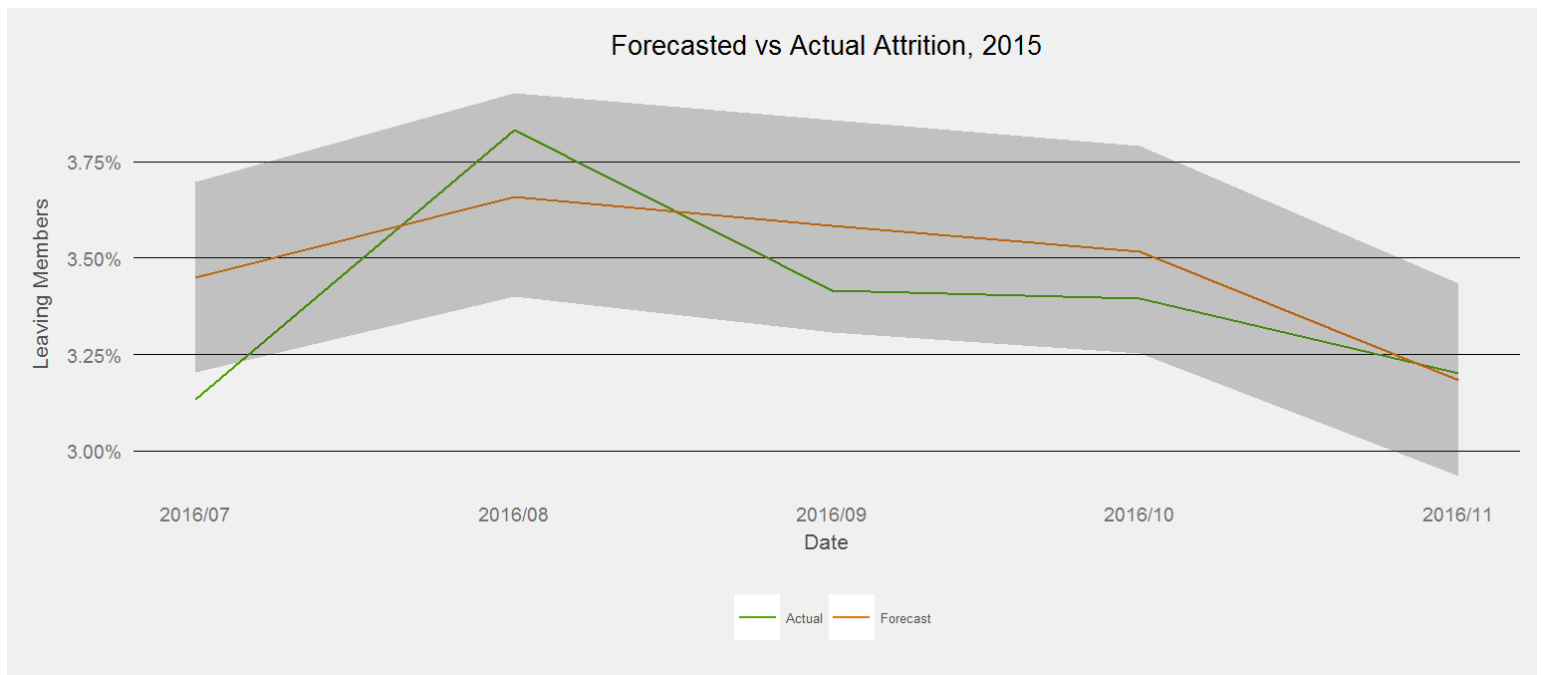


fig 11

## 6 Conclusions

Results are within the normal bounds of expectations for both Total Monthly Leavers, as well as churn rate. While August 2016 saw elevated results, the attrition numbers stabilize for the remaining observation periods. All results are within acceptable error limits. Speaking extemporaneously, you could potentially slice this data up into various 2 or 3 year views and get a variety of results. That is, 2015 appears to be somewhat of an outlier in terms of attrition performance, and the growth rate for 2011-2013 is not reflective of the current state. Additionally, a multivariate model was attempted as an alternative to the univariate time series to determine if using business unit share and size was a more accurate predictor vs. the trend and seasonal data used by the various time series algorithms (it was not).

Long term I would suggest not using a time series model as a final predictor, but rather build a traditional independent linear regression model (with a time series model as a potential input) to attempt to predict attrition. There are a multitude of potential predictors that are unaccounted for in this quick modeling exercise, and a full data discovery should be undertaken to create a more comprehensive model.

## 7 Appendix A - Sql queries

```
with Leaving_Members as ( select to_char(ada.adm_date,'YYYY-MM') as leaving_ym ,rpc.segment_class ,rpc.plan_name
,rpc.fee_freq ,zf.level_5_key as zipfleet ,count() as leaving_members from account_driver_activity ada join zipfleet_rollups zf
on ada.last_left_zipfleet_id=zf.rollup_id and zf.level_3_key='continent_north_america' JOIN analytics.rate_plan_cat rpc on
ada.prior_rate_plan_id = rpc.rate_plan_id where action='left' and account_count = 0 and adm_date > date '2011-01-01' group
by to_char(ada.adm_date,'YYYY-MM') ,rpc.segment_class ,rpc.plan_name ,rpc.fee_freq ,zf.level_5_key ) , total_members as (
select to_char(mmd.month,'YYYY-MM') as member_ym ,rpc.segment_class ,rpc.plan_name ,rpc.fee_freq ,zf.level_5_key as
zipfleet ,count() as total_members from v_member_month_details mmd join zipfleet_rollups zf on
mmd.joined_zipfleet_id=zf.rollup_id and zf.level_3_key='continent_north_america' JOIN analytics.rate_plan_cat rpc on
mmd.join_rate_plan_id = rpc.rate_plan_id where month>= date '2011-01-01' group by to_char(mmd.month,'YYYY-MM')
,rpc.segment_class ,rpc.plan_name ,rpc.fee_freq ,zf.level_5_key )
```

```
select tm.* ,lm.leaving_members from total_members tm left join leaving_members lm on tm.member_ym=lm.leaving_ym
and tm.segment_class=lm.segment_class and tm.plan_name=lm.plan_name and tm.zipfleet=lm.zipfleet
```