



1/17/16

Customer Cube & Predictive Model

Tony Caballero

AGENDA

- Problem
- Data
- Hypothesis



Problem

- Objective:

- Leverage the data warehouse to generate insight into customer cohorts and serve as the foundation for a predictive model to predict Dollar Based Net Expansion

- Problem Statement:

- Dollar based net expansion is difficult to predict. Doing so (accurately) on a monthly basis for the next year will guide Optimizely's focus into certain countries and industries as well as guide how we invest our resources to service segments where we find favorability
 - Accurate Dollar Based Net Expansion requires predicting it correctly for the next 3 months within +- 3 percentage point confidence interval



Data

- So far, I've collected over 40 variables in the customer data cube
 - Some variables are time series (MRR, Traffic, # of logins)
 - others are static variables (industry, segment, country, region, AE, etc)
 - Type of data = Salesforce data, Optimizely product data, finance data
- 200,000 records generated via SQL query
- Simplified the data set by excluding customers who've never paid \$300 in MRR since they have a different behavior profile
- Cube consists of 5 years' worth of data – the integrity of the data improves over time as well as amount of data



Hypothesis

- At least one of the variables that I captured will be a significant linear predictor of dollar based net expansion
- If my preferred analysis hypothesis doesn't work out, there's so many variables in the data set, that I can investigate the relationship among the variables themselves. Even discovering colinearity among the independent variables will be valuable insight.

