# Predicting Subscriber Churn

DAT7, Summer 2015

# Question(s)

- For a given subscriber, what is the likelihood of churn?

- What are the major determinants of churn?

# Why?

Subscribers

=

Eyeballs and/or Interactions

=

Advertising $$$$

=

Profit

## Problem

Recruiting New Subscribers Is Difficult and Expensive

## Solution

Retention

# Data

<u>Dependent Variable</u>

Subscriberid + Status (U, or S) + Date_Effective = Churn (binary)

<u>Independent Variables(?)</u>

Product (200+ in Universe)

Normalized Values (Position Function & Level,  Company Type)

Duration of Subscription

Engagement Metrics (Date of Last Open, Date of Last Click)

Type of Acquisition Channel

<u>Domain Knowledge</u>

Type of Churn

Time

# Modeling

Accuracy is important…

- Why spend extra money on subscribers who would have stayed anyway?

So is comprehensibility…

- Internal stakeholders need to trust that results are consistent with past experiments
- Insight into the determinants of churn (generally)
- Adoption by business users (Audience Development, Sales)

Logistic Regression (Default)
Decision Trees
Support Vector Machines
Survival Analysis (Time to Churn!)