Below is a typical problem that we try to solve in the Strategy group. It's an open problem with no correct answer. If you have any questions at all, feel free to e-mail Zach Marine, our analytics team lead, at zmarine@squarespace.com.

Overview

On any given day, a certain number of people begin trials. We'll call all the people who begin a trial on the same date a trial cohort. A percentage of this cohort will eventually go on to subscribe. The cohort's conversion rate is defined as the percentage of people from that cohort who go from starting a trial to subscribing. This number is influenced by product and marketing decisions we make, and as a result conversion rate is a key source of feedback on the effectiveness of our decisions.

Many of our customers who create trials become paying subscribers the same day, but some become paying customers weeks or months later. We can't wait months for feedback, can we? We'd like to be able to estimate a trial cohort's conversion rate as soon as possible. In addition, as a trial matures, we get a better picture of how it's performing (through additional conversions), so our estimated conversion rate should become increasingly accurate over time.

For example:
500 people began a trial on 4/1/2013 and eventually 230 of them converted. So, we can say that our conversion rate for this cohort is 46%. The goal of this problem is to estimate the conversion rate, but rather than waiting for many days/months, make that estimate at the end of day 1, day 2, day 3, etc. for any given cohort.

So if you take all the trialers for a given date (let's say 2,686 on 5/1/2013), and count the number that eventually subscribed (538) we can say that the eventual conversion rate is 538/2686 or 20% for that cohort - now just find a way to predict this number from the first day and to improve its accuracy as the cohort matures (i.e. you get additional information on the cohort).

Problem

Can you create a method that will help us predict the final conversion rate for a cohort of trailers and which becomes more accurate as the cohort matures?

Please include:

1) An explanation of your methodology for making this estimation. Please provide relevant code used and explain why you choose this method instead of an alternative.
2) Any code you used for data exploration.
3) A description (and code) for how you evaluated your method's accuracy.
4) Any other interesting findings or suggested next steps.

Data

We're attaching the data for the problem. You get a list of approximately 75,000 people who began trials, and a unique ID for each person. You also get the date, hour, and day of week from which they began a trial. Next to some customers you also get the date and hour they subscribed (if they did); if they never subscribed it was left blank.

* Note: Data has been altered from its original version for the purpose of this question.

Good luck!