

ClientComm Data Analysis 3

Gwen Rino and Eric Giannella

7/20/2018

OVERVIEW AND KEY FINDINGS

This is our final report (for now!) on our investigation of supervision outcomes in the Baltimore ClientComm data. We can now pull data from AWS at any time with confidence that the time zone is correct, we can filter the data to more accurately find messages from clients, and we can articulate a robust and meaningful model that predicts supervision success.

Key Findings

- 1) Relationships in which a client participated modestly (by sending one or two messages) had a significantly higher chance of a positive supervision outcome than relationships in which a client sent no messages. Relationships in which a client participated more than that (by sending three or more messages) had even better chances of a positive outcome.
- 2) Client participation is the only variable that is significantly associated with positive outcomes, regardless of the number of observations or other variables included in the model.
- 3) Relationship-level variables that aggregate the counts of PO-sent messages by time of day or day of week reveal no significant association with supervision outcomes.
- 4) Data munging is improved so that modeling results are more meaningful; see details below. All future modeling should process data accordingly.

DATA IMPROVEMENTS

Time Stamps

Data is now pulled and processed in such a way that time stamps reliably reflect the correct time zone.

Inbound Messages

Inbound messages are now filtered to exclude system messages regarding updated phone numbers or client reassignments.

ANALYTICAL STORY

Client participation in a texting relationship matters. A client who never texts has a 9.9% chance of supervision failure; a client who texts once or twice has a 6.5% chance of failure; a client who texts more than twice has a 5.4% chance of failure. Client participation is associated with as much as a 45% reduction in the chance of supervision failure.

Other variables that reflect EBPs (`future_appointment_date`, `max_scheduled_diff`) do seem to be meaningfully associated with a slight improvement in a client's chance of supervision success, but neither of these is statistically significant.

Other variables (time of day, day of week, and length of PO message) show no relationship to supervision outcome. This is probably because the possible explanatory variable is far "upstream" from the response variable, and many other unmeasured events may be much more influential. For example, a PO sending a message on a Thursday morning likely has an immeasurably weak influence compared to the impact of a child's illness on the outcome of the supervision relationship.

The good news is that we can investigate client participation as a much more proximate response to qualities of PO texts. This will be the **next phase** of our investigation: What PO message qualities are associated with client participation?