AV Experiment Analysis

Greg Welliver

Notebook

Context

We conducted an analysis of the efficacy of two email campaigns, as well as a study of user characteristics. This analysis focused on answering two key questions:

- Which of the two email campaigns should we scale for a larger campaign?
- Which characteristics of users are most indicative that they will convert to paying subscribers?

Findings & Recommendation

- Campaign B had a conversion rate of 10.4%, while Campaign A had a conversion rate of 9.1%. **Based on this finding, we recommend scaling Campaign B.**
- Variables **x4**, **x5**, **x6,** and **x8** are the most predictive of user conversion, based on multiple methods of deriving feature importance.

Detailed Insights

- The experiment consisted of 1,000 emails sent to users, with two different treatments, A and B. Treatment A was sent to 41% of users, while Treatment B was sent to 59% of users.
- Overall, 99 users converted (9.9%) to paying subscribers.
- Campaign B had a conversion rate of 10.4%, while Campaign A had a conversion rate of 9.1%.
- Generally the data was very clean. There were no duplicate IDs nor missing values. The explanatory variables x1 - x10 were reasonably normally distributed.
- Variables x5 and x8 showed some correlation with each other across their entire range of values. x8 also showed some correlation with x5 and x6 at certain levels, but not across the entire range.

Potential areas for improvement

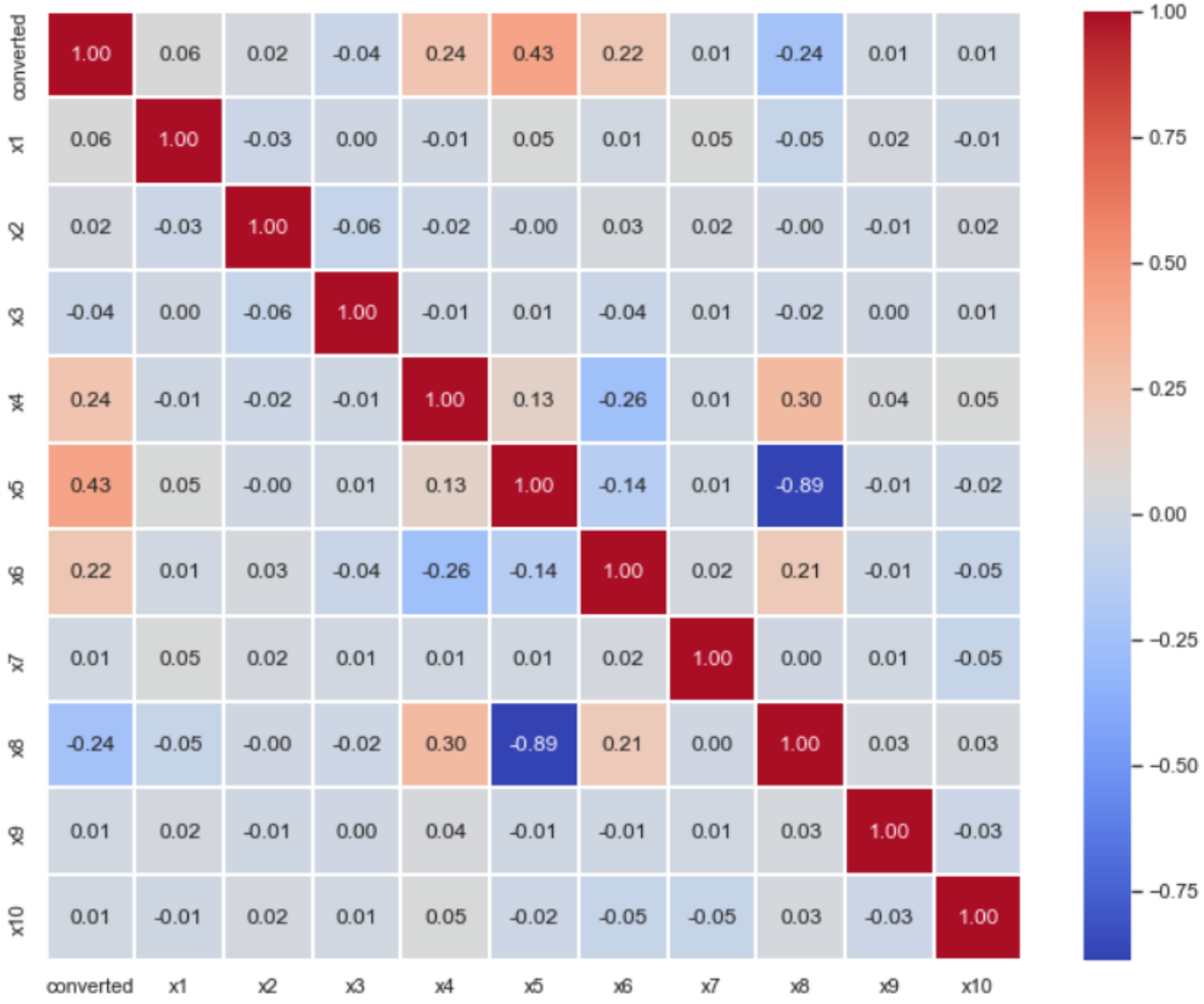
- The number of observations in each of the email campaigns was not balanced (A was sent to 41% of users, while B was sent to 59% of users). For an A/B test with two variants, it would be practical to have a 50/50 split.
- The prompt didn't go into detail on the experiment setup, but we would want to ensure that the users were randomly bucketed into groups A and B. If they weren't properly randomized, there could be bias, which could potentially contribute to the improved performance of B versus A.

- 1,000 observations is a decent sample size, but certainly we could have a much larger sample that would bolster statistical significance.
- Variables x5 and x8 showed correlation with each other across their entire range of values.
- Variable x4 showed some correlation with x6 at certain values.
- Variable x5 showed some correlation with x6 at certain values.
- Variable x6 showed some correlation with x4, x5, and x8 at certain values.
- Given that these variables were the ones that we found to be the most influential, we may want to explore this further and potentially remediate (e.g. perhaps remove variables).

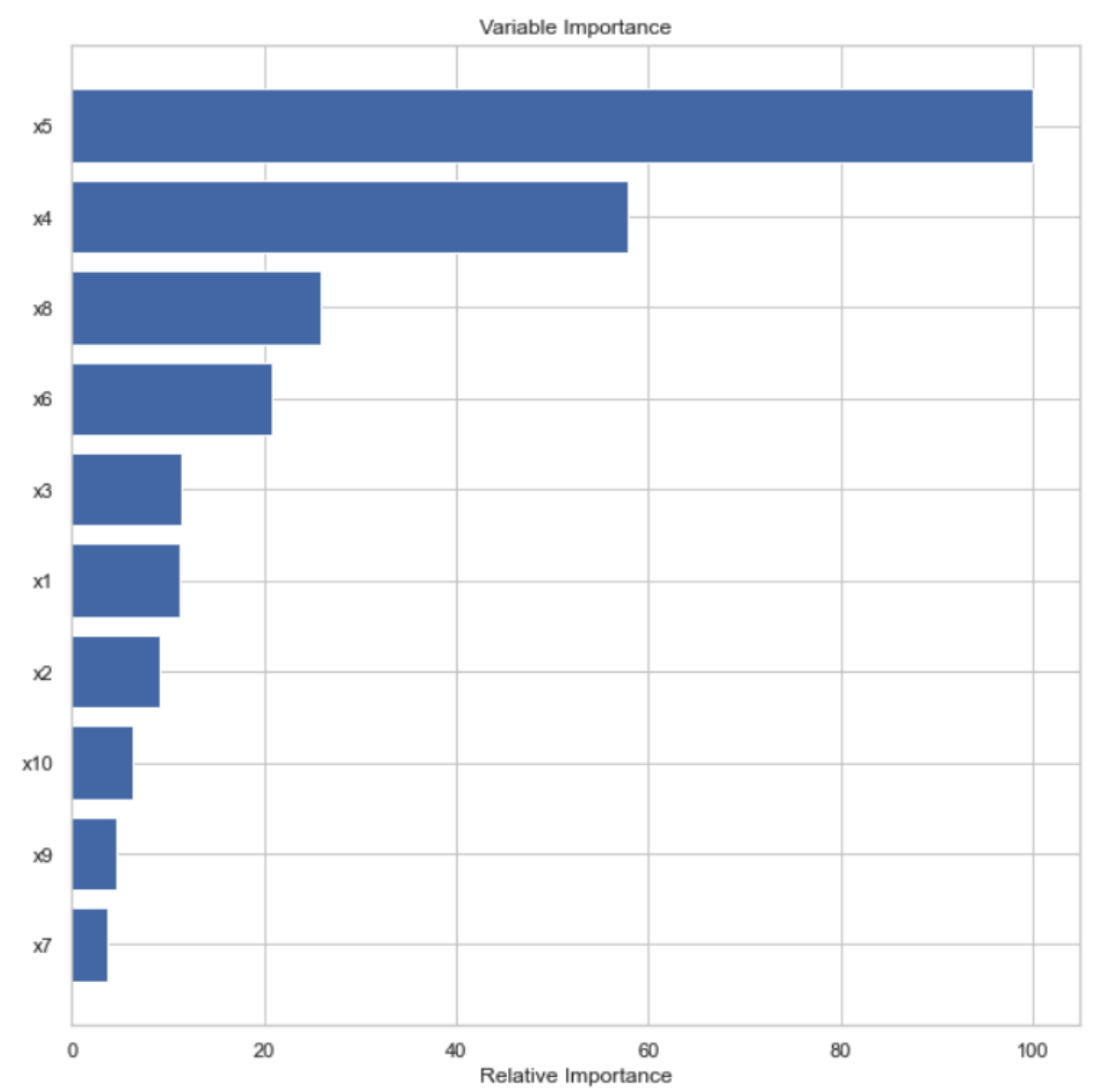
Feature importance

I utilized a few different methods to analyze feature importance. Each method confirmed the importance of x4, x5, x6, and x8. (see below for visuals)

Method 1 - Correlation Matrix



Method 2 - Random Forest Variable Performance



Method 3 - Logistic Regression Output (notice the high coefficients on x4, x5, x6, and lower coefficient for x8)

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
x1	-0.0071	0.1144	-0.0620	0.9505	-0.2312	0.2170
x2	0.1709	0.1052	1.6246	0.1043	-0.0353	0.3771
x3	-0.0542	0.1101	-0.4921	0.6227	-0.2700	0.1616
x4	1.6536	nan	nan	nan	nan	nan
x5	1.2612	nan	nan	nan	nan	nan
x6	1.2826	nan	nan	nan	nan	nan
x7	0.0408	0.1064	0.3835	0.7013	-0.1677	0.2493
x8	-0.0715	nan	nan	nan	nan	nan
x9	0.0473	0.1066	0.4441	0.6570	-0.1615	0.2562
x10	0.0465	0.1074	0.4333	0.6648	-0.1639	0.2570