Combined campaign types

- Category combinations only occur in the file for the second time period (Jan-Jun 2020), not for the first (Jul-Dec 2019) or the third (Jul-Sep 2020) time periods.
- Only 2 category combinations appear: Brand – Collection (592,769 emails) and InnovationSpotlight – Tops (344,301 emails)
- There are only 2 Brand – Collection campaigns: 'Introducing: the Dockers® Retro Collection' and 'Introducing: the Dockers® Pride 2020 Collection']
- Collection never appears without Brand in the file for the second time period, and it never appears at all in the file for the first time period.
- In the file for the third time period, there are 1,349 emails with campaign type 'Collection' (not in combination with any other type). All of these emails are from the same campaign: 'Back in stock: Dockers® x Karla'. This is a very small number of emails (by comparison, the second-smallest category is 'Replen' with 274,717 emails – over 200 times more than for 'Collection').

Suggestion 1: 'Collection' should not be considered a separate campaign type. There are only three campaigns of that type in the entire dataset; two of these campaigns (in the file for the second time period) are also of type 'Brand' and the third (in the file for the third time period) is very small (perhaps, sent to a subset of customers who previously purchased Dockers® x Karla?)

I suggest categorizing 'Brand'/'Collection' campaigns simply as 'Brand' campaigns and including the remaining 'Collection' campaign in the 'Other' category.

- In the second time period, there are 3 Tops only campaigns ('Drumroll, please: NEW, INNOVATIVE SHIRTS', 'Printed shirts for your virtual get-togethers', 'Grab these shirts by the collar'), 3 Innovation Spotlight campaigns ('What's Smart 360 Flex™?', 'Our most advanced fabric, period', 'Smart 360 Flex™: a definition'), and 1 campaign classified as Tops/Innovation Spotlight ('Supreme Flex™ from top to bottom').
- It is possible that the combined-category campaign was miscategorized (the subject lines of other Tops campaigns all focus on shirts, but this one mentions 'bottom', as well as 'top'). In any case, at first glance the subject of this campaign seems to be more similar to Innovation Spotlight than to Tops campaigns, with a primary focus on Flex™.

Suggestion 2: *The only Tops/Innovation Spotlight campaign can probably be considered simply an Innovation Spotlight campaign.*

Customer features vs. (fixed and variable) campaign features

In terms of deciding (a) for which features to explore potential correlations and (b) how different features might be combined and where they might go in the model, it might be useful to think of data columns as describing either the customer or the campaign.

Customer:
- riid, retention_score, frequency_score, sends_since_last_open,

Campaign (fixed):
- campaign_type, subject_words, message_size, discount(?), promo, sale, sent_week (?), <mark>sent_month (?) – remove, correlated with sent_week,</mark>

Campaign variable:
- sent_hr, sent_dayofweek,

Reward:
- opened, unsub, rev_3dv2, reward, optimal action