What if Co-op Brands Sold Snow Pants?

An analysis of REI website data

Jake Weinreb (4/9/2019)

Executive Summary

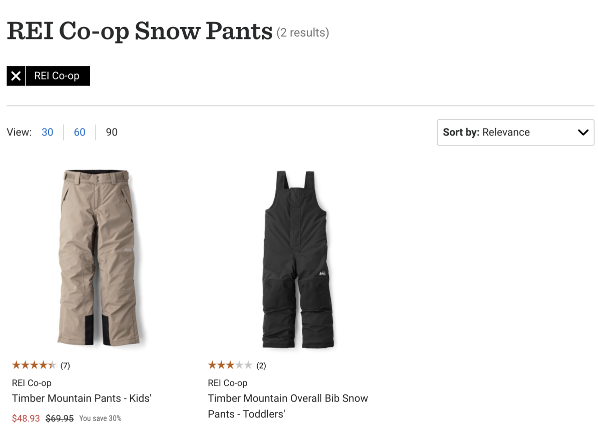
A recent post[1] on the Co-op Journal dubbed 2019 the “Year of the Bibs”. Krista Hildebrand, REI’s Senior Category Merchandising Manager for action sports apparel, mentions in the article that Feburary 2019 bib sales were up 106% YOY. That got me thinking–why doesn’t Co-op Brands sell snow pants? And if they were to start, what are some of the key considerations? This analysis of REI website data begins to answer that question.

After scraping the limited available data on REI.com and cleaning it for statistical significance, I find that snow bibs and pants are equally well-liked on average. In isolation, this result may call into question the need to jump into the snow pants arena with a bib offering, as opposed to traditional snow pants. (Full disclosure: I have been wearing a bib for the past three seasons and love it!) While bibs and pants share a similar average rating, the distributions of the ratings are different, opening up the possibility for further analysis with additional data. Perhaps most surprisingly, I find that women prefer bibs and men prefer traditional pants. Men also engage on the site at roughly twice the rate of women, adding interesting color to the implications of the bibs vs. pants question. I then provide a starting point for product inspiration by analyzing which brands have the best products in the space. Finally, I find little (if any) relationship between price and rating, and highlight potential areas for further research.

These are preliminary insights given the limited data.

Why Snow Pants?

This analysis begins with a big assumption: Co-op Brands should start selling snow pants. The reason is simple. Co-op Brands has established itself as a preeminent outdoor gear and apparel brand, building goodwill with the outdoors community in verticals from camping tents to leggings. The brand has not neglected the skiers and snowboarders either. On the contrary, almost 8% of of the SKUs on the Co-op Brands section of the REI website are categorized under skiing. When we look for ski pants, though, this is what we see:



Co-op Brands snow pants are just for kids!

Given their current lineup of kids ski pants and adult ski gear, Co-op Brands already has an existing supply chain, knowledge of the design specs, brand awareness with the likely consumer, and offers complementary products within the snow apparel vertical. While much more analysis should be done before recommending a move to a new product line, an initial review shows that snow pants are a potential area of opportunity for Co-op Brands.

Getting Set Up

I begin the analysis by loading the relevant R libraries and creating lists of URLs for scraping. The URLs are landing pages for different categories of items I'm interested in for this analysis, including: women's ski clothes, men's ski clothes, women's ski pants, men's ski pants, women's bibs, and men's bibs.

From there I scrape relevant data from the REI website. The function I wrote takes in a list of URLs and returns the relevant page data in a tidy format for analysis in R. I functionalized the scraping so that the analysis can be ported over to any product on the REI site.

With a working scraping and cleaning function, I can quickly and easily create the desired datasets.

Transforming the Data

The most pernicious issue with this dataset is the lack of reviews. With most products garnering fewer than 20 reviews, and many with fewer than 5, the data is too small to analyze directly. That's why I have chosen a Bayesian approach, first coming up with a thoughtful prior distribution of ratings (generated externally to this data), and then shrinking our data to this prior.

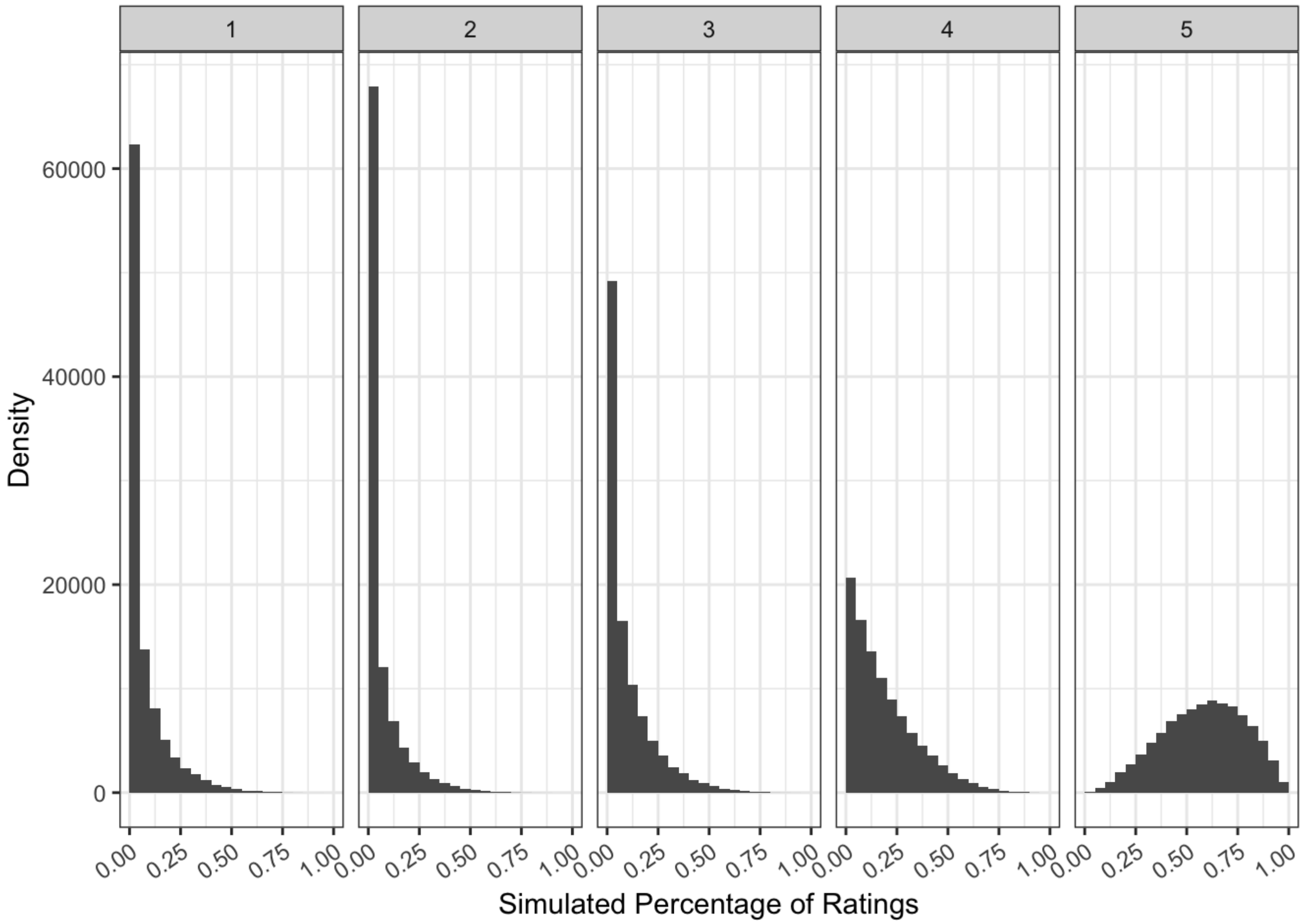
> *The data required crucial cleaning steps to ensure statistical significance. If you want to jump straight into the analysis, feel free to skip the remainder of this section.*

Ratings data are distributed multinomially, meaning every individual rating will be placed in a single category (in our case, a star rating of 1-5) with a given probability. We can also assume that each one of those probabilities come from a Dirichlet distribution, the conjugate prior of the multinomial distribution. This basically means that when we combine the Dirichlet prior probabilities with the actual data, the resulting distribution (the posterior) is also a Dirichlet distribution, just with new parameters. Conveniently, those new parameters are simply the sum of the prior parameters and the data.

When we choose our prior parameters, we generally want to do so in a way that is not already biased by the data we have, particularly if our data isn't very large. For this analysis, I chose the parameters of the prior based on a research paper[2] on the nature of online reviews. The paper lists the star ratings (1-5), number of products, and total number of reviews for various product verticals on Amazon, of which I selected 'clothing'. With 5.7 million clothing reviews on 1.1 million products, the averge number of reviews per product is approximately 5. I multiplied the probabilities of each star rating by 5 total reviews to get the prior distribution shown below.

one_star	two_star	three_star	four_star	five_star
0.35	0.3	0.5	0.95	2.9

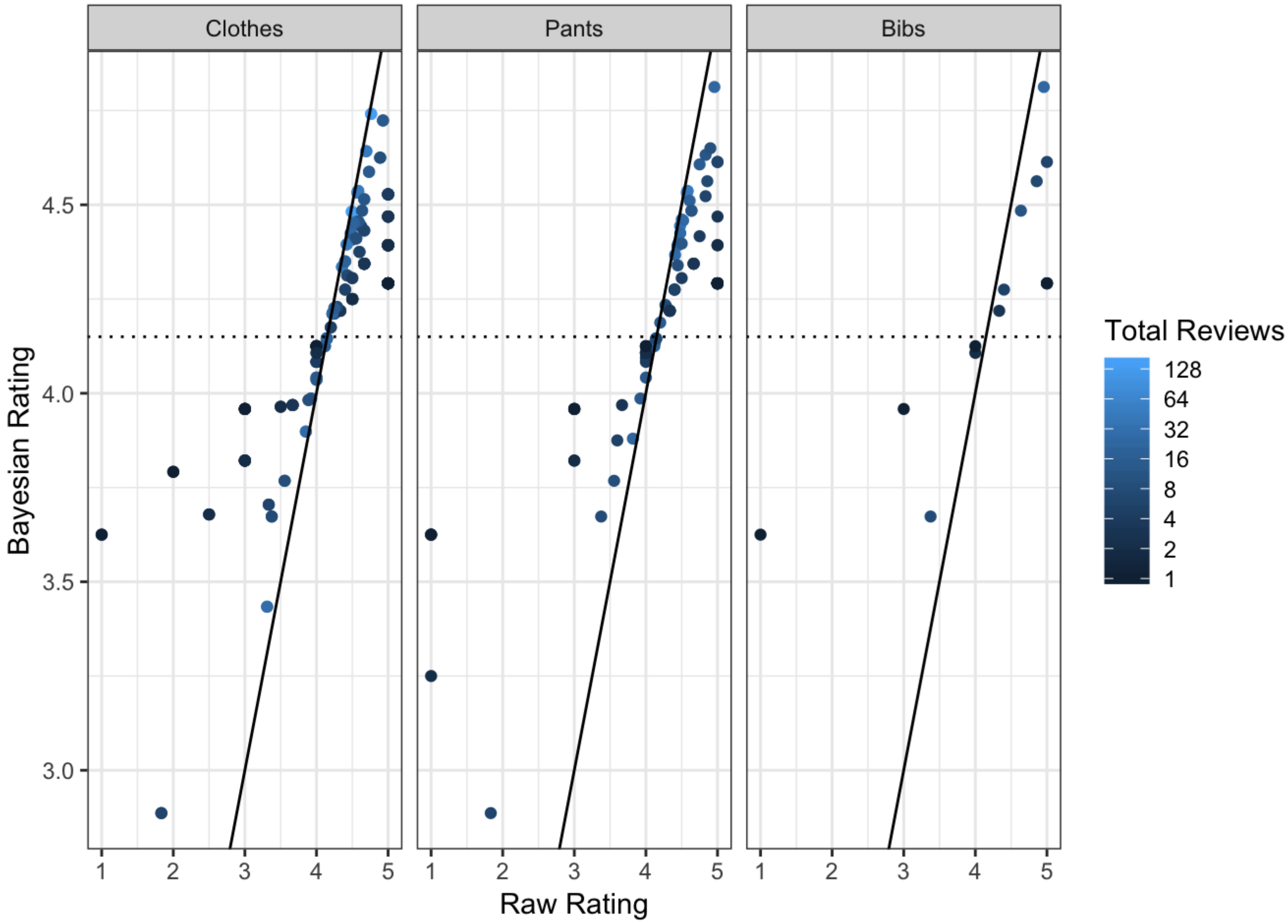
This prior is simply an estimate of what we might expect the data to look like without actually looking at our data. Since it is an estimate, there is considerable uncertainty. To visualize this uncertainty, I simulated 10,000 random ratings distributions using the parameters from our prior, and plotted them in the histogram below. This helps build the intuition that, for example, the probability of a product receiving zero 1- or 2- star ratings is quite high.



From the table below we can see that the current ratings on the REI site for general snow apparel are similar to (though slightly higher than) the Amazon ratings. We're using the Amazon data because of its volume, but this directionally corroborates our prior.

data	one_star	two_star	three_star	four_star	five_star
REI	5.40%	5.074%	9.126%	19.294%	61.29%
Amazon	7.00%	6.000%	10.000%	19.000%	58.00%

Now that we have our prior, we can use it to effectively shrink our data toward that prior, which alleviates concerns from small sample size. Products with many reviews will shrink less, while products with fewer reviews will shrink more. This makes intuitive sense, as the more data we have on a product, the more confident we can be that the reviews represent the "true" population's view on that product. We can plot the data transformation graphically as below. We can clearly see the shrinkage–observations below and to the left of the crossing lines are being inflated toward the mean, while observations above and to the right are being deflated. We can also see that the more ratings a product has, the less the estimate is shrunk (i.e. closer to the 45 degree line).



Raw product ratings vs. cleaned ratings. The horizontal dotted line is the prior distribution average, and the diagonal is the 45 degree line.

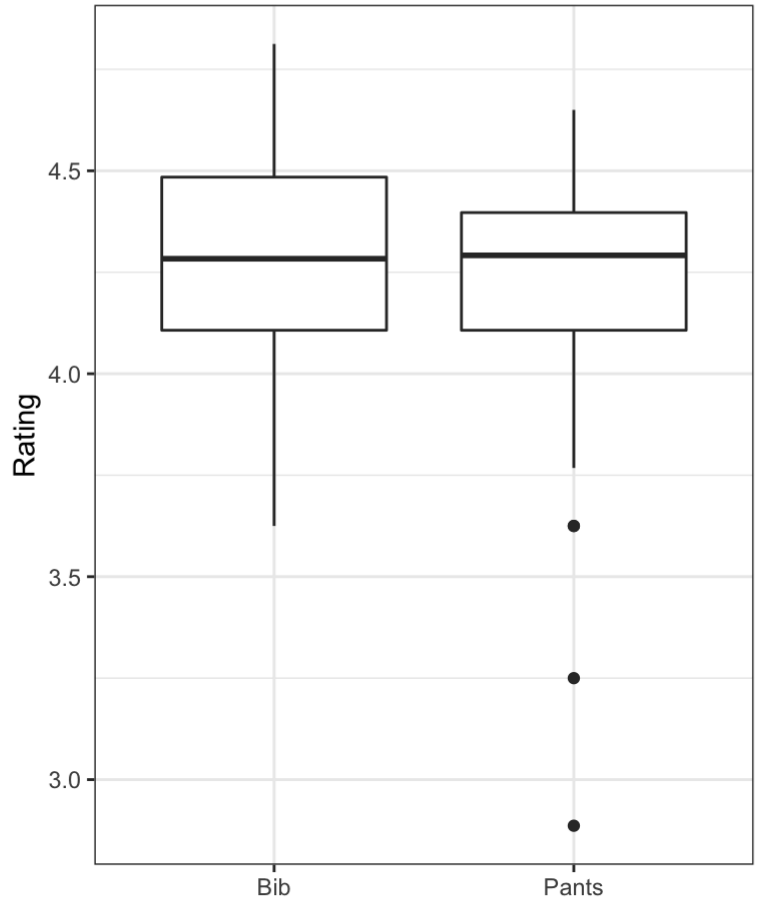
Considerations for Snow Pants Strategy

Now that we have clean, statistically significant data, we can finally get to the question at hand: what are some interesting considerations to inform our snow pants strategy?

For starters, we can look at the relative ratings of all snow pants vs. non-pants ski clothes. We see below that they are virtually the same, indicating that the opportunity set (in terms of in-place product satisfaction) in the snow pants market is similar to the other ski clothing markets in which REI currently has product.

clothing_type	rating
Other Clothes	4.222314
Pants/Bibs	4.227887

Digging into pants, we can see that the median rating for bibs and non-bibs is approximately equal, though the distributions are somewhat different. In order to interpret this data in a meaningful way, we would want to dive deeper into the relationship between ratings and sales. Another possibility for further research is to look into each one of the outliers more qualitatively to understand what led to their failure. It's interesting to note that while the median bib is rated slightly lower than the median pant, the negative outliers are entirely pants.

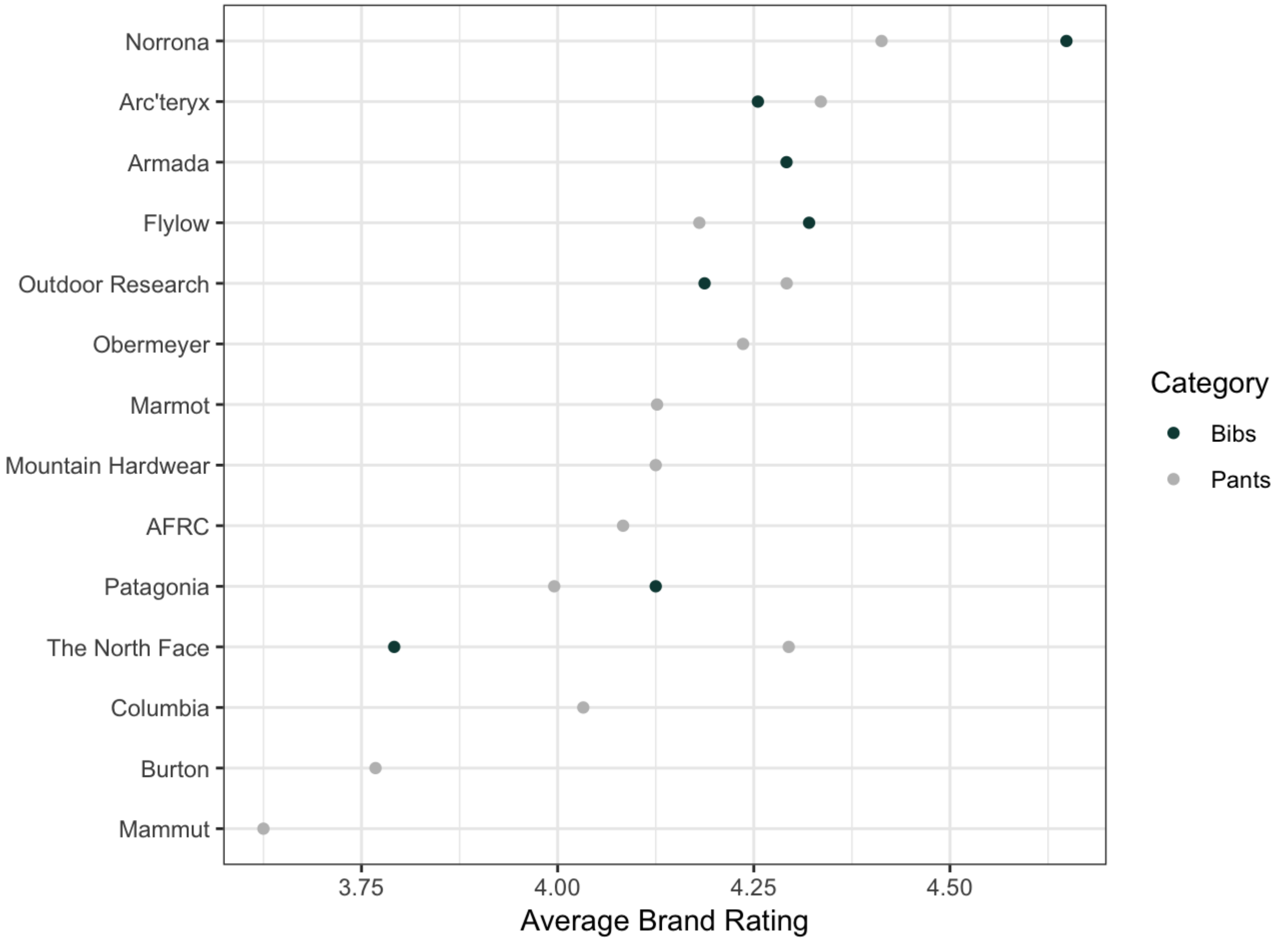


Perhaps the most surprising result of this analyis is that women enjoy their bibs more than men do, both on an absolute basis and relative to traditional pants. A common anecdote I have heard is that women find bibs less convenient, as they are more difficult to take on and off than pants. This data refutes that claim, and gives a preliminary indication that a line of bibs for women would be well-received. It also reveals an opportunity to satisfy a potentially underserved need for men. An interesting wrinkle is the engagement—on average men leave almost twice the number of ratings on the REI website across all ski products. If this correlates to engagement with the brand overall, the bibs vs. pants question may come down to doubling down on an existing core customer base versus expanding with a less engaged customer.

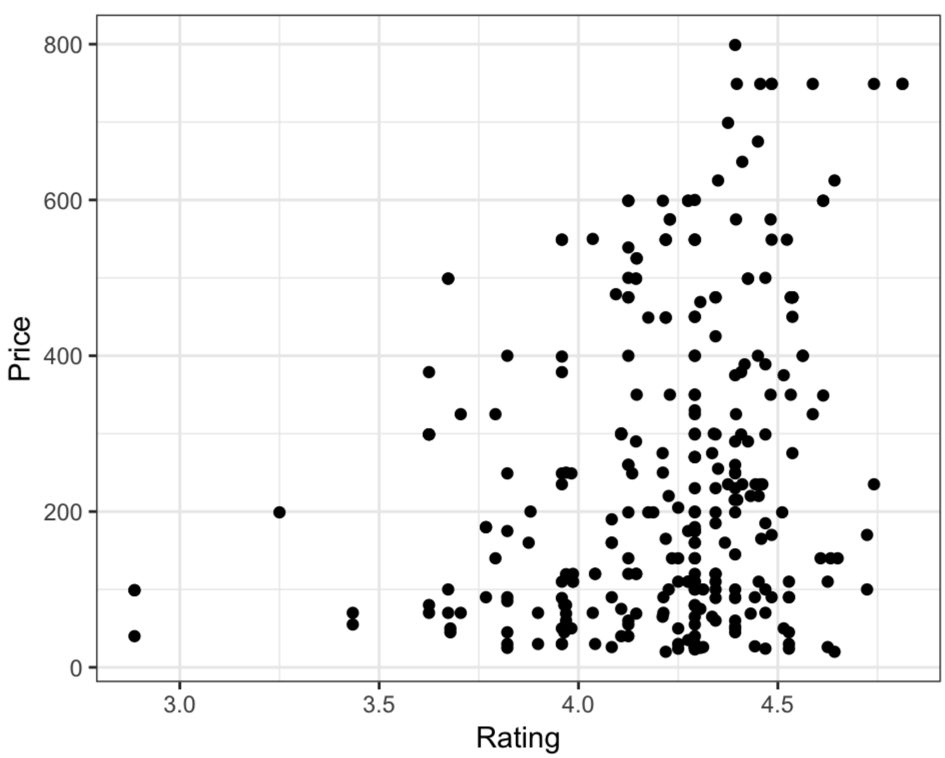
category	rating
men_bib	4.186218
men_pants	4.249179
women_bib	4.306762
women_pants	4.163071

gender	engagement
men	13.165644
women	7.065217

We can also look at overall brand ratings for bibs and non-bibs. This might give us an idea of which brands are currently executing best (or worst), so we can dive deeper into why that's the case. For example, Norrona is impressive in both bibs and pants, while The North Face's bib offering isn't as successful as their pants.



Finally, we can start to tease apart factors that influence the rating, such as price, color, specs, etc. As a starting point, I have plotted price against rating and found the Spearman[3] correlation between the two to be .13, reflecting a weak relationship. One insight to draw from this chart is that REI could theoretically come to market with the relative lack of competition at the higher end of the market as compared to the cheaper options.



Future Steps

This report is merely the tip of the iceberg for moving into a new product category. We find that the decision for Co-op Brands to make its first foray into adult snow pants with bibs versus the more traditional (but equally well-rated) pants is not so cut-and-dry. We also find that women tend to prefer the current bib lineup, while men tend to prefer pants. The strategic implications may vary based on target customer focus, and could potentially signal an opportunity to either satisfy an unmet need in men's bibs, or come to market with more confidence in women's bibs. These data points can't tell the whole story on their own, but are potentially valuable as part of a larger analytics mosaic.

With the data we scraped, we can continue to probe colors and tech specs to see if there are any patterns that correlate with better ratings. With additional data, we can study the relationship between sales and customer ratings, as well as an outlier analysis, which could reveal important strategic insights. Finally, we can tailor the product offering to our target customer and build a robust financial model to solidify our projections.

Appendix

Below you will find all the code for this analysis, from scraping web data to Bayesian prior simulation to ratings boxplots. I have included the code as an appendix so as to reduce clutter in the report.

```
#load libraries and set a global plot theme
library(tidyverse)
library(rvest)
library(rebus)
library(jsonlite)
theme_set(theme_bw())

#list of cateogry page urls
url_women_pants <- c(
  url_women_pants_1 ='https://www.rei.com/c/womens-downhill-ski-pants?pagesize=90&
ir=category%3Adownhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhill-ski-cl
othing%7Cdownhill-ski-pants%7Cwomens-downhill-ski-pants&sort=rating%7Cnum-reviews&
origin=web'
)

url_men_pants <- c(
  url_men_pants_1 = 'https://www.rei.com/c/mens-downhill-ski-pants?pagesize=90&ir=
category%3Adownhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhill-ski-cloth
ing%7Cdownhill-ski-pants%7Cmens-downhill-ski-pants&sort=rating%7Cnum-reviews&origi
n=web'
)

url_women_bib <- c(
  url_women_bib_1 = 'https://www.rei.com/c/womens-downhill-ski-pants/f/f-bibs?page
size=90&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdo
wnhill-ski-clothing%7Cdownhill-ski-pants%7Cwomens-downhill-ski-pants%3Bf&origin=we
b'
)

url_men_bib <- c(
  url_men_bib = 'https://www.rei.com/c/downhill-ski-pants/f/f-bibs?pagesize=90&ir=
category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhill-ski-
clothing%7Cdownhill-ski-pants%3Bf%3Bgender%3AMen%27s&origin=web'
)

url_women_ski_clothes <- c(
  url_women_ski_clothes_1 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=
90&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhil
l-ski-clothing%3Bgender%3AWomen%27s&sort=title&origin=web',
  url_women_ski_clothes_2 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=
90&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhil
l-ski-clothing%3Bgender%3AWomen%27s&sort=title&origin=web&page=2',
  url_women_ski_clothes_3 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=
90&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhil
l-ski-clothing%3Bgender%3AWomen%27s&sort=title&origin=web&page=2',
  url_women_ski_clothes_4 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=
90&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhil
l-ski-clothing%3Bgender%3AWomen%27s&sort=title&origin=web&page=2',
  url_women_ski_clothes_5 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=
```

```
90&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhil
l-ski-clothing%3Bgender%3AWomen%27s&sort=title&origin=web&page=2',
  url_women_ski_clothes_6 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=
90&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhil
l-ski-clothing%3Bgender%3AWomen%27s&sort=title&origin=web&page=2'
)

url_men_ski_clothes <- c(
  url_men_ski_clothes_1 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=90
&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhill-
ski-clothing%3Bgender%3AMen%27s&sort=title',
  url_men_ski_clothes_2 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=90
&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhill-
ski-clothing%3Bgender%3AMen%27s&sort=title&page=2',
  url_men_ski_clothes_3 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=90
&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhill-
ski-clothing%3Bgender%3AMen%27s&sort=title&page=2',
  url_men_ski_clothes_4 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=90
&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhill-
ski-clothing%3Bgender%3AMen%27s&sort=title&page=2',
  url_men_ski_clothes_5 = 'https://www.rei.com/c/downhill-ski-clothing?pagesize=90
&ir=category%3Amens-downhill-ski-clothing&r=category%3Adownhill-skiing%7Cdownhill-
ski-clothing%3Bgender%3AMen%27s&sort=title&page=2'
)

#define a function to take in a specific product page url and read the html/json i
nto a usable format
get_individual_data <- function(url) {
  page <- read_html(url)
  json <- page %>% html_nodes(css = 'script') %>% html_text()
  fromJSON(json[[which(str_detect(json, 'specs'))]])
}
#this is the main scraping function that takes in a list of category page urls (re
call some product categories will have multiple pages of urls), loops through each
of the products on the page to create product-level urls, gathers the html/json fo
r those urls (using previous function), pulls out the relevant data from each pag
e, and performs a few cleaning steps to make the data easier to work with
get_web_data <- function(url_list) {
  json <- map(url_list, read_html) %>%
    map(html_nodes, '#page-data') %>%
    map(html_text)
  output <- map(json, fromJSON) %>%
    map(~.$data$results) %>%
    map(select, index, prodId, cleanTitle, brand, regularPrice, availableColors) %
>%
    bind_rows(.id = 'category') %>%
    mutate(webTitle = tolower(str_remove_all(cleanTitle, pattern = PUNCT) %>%
                              str_replace_all(pattern = one_or_more(" "), replac
ement = "-")))
```

```
  url_list_full <- paste('https://www.rei.com/product', output$prodId, output$webT
itle, sep = "/")
  page_data_full <- map(url_list_full, get_individual_data)

  df <- vector("list", length = length(page_data_full))
  for(i in seq_along(page_data_full)) {
    if (length(page_data_full[[i]]) < 13) {
      df[[i]]$total_reviews <- page_data_full[[i]][[1]]$reviewsSummary$total
      df[[i]]$avg_rating <- page_data_full[[i]][[1]]$reviewsSummary$overall
      df[[i]]$rating_1 <- page_data_full[[i]][[1]]$reviewsSummary$ratingHistogram
$`1`
      df[[i]]$rating_2 <- page_data_full[[i]][[1]]$reviewsSummary$ratingHistogram
$`2`
      df[[i]]$rating_3 <- page_data_full[[i]][[1]]$reviewsSummary$ratingHistogram
$`3`
      df[[i]]$rating_4 <- page_data_full[[i]][[1]]$reviewsSummary$ratingHistogram
$`4`
      df[[i]]$rating_5 <- page_data_full[[i]][[1]]$reviewsSummary$ratingHistogram
$`5`
      df[[i]]$would_recommend <- page_data_full[[i]][[1]]$reviewsSummary$wouldReco
mmendCount
      df[[i]]$would_not_recommend <- page_data_full[[i]][[1]]$reviewsSummary$would
NotRecommendCount
    } else {
      df[[i]]$total_reviews <- page_data_full[[i]]$reviewsSummary$total
      df[[i]]$avg_rating <- page_data_full[[i]]$reviewsSummary$overall
      df[[i]]$rating_1 <- page_data_full[[i]]$reviewsSummary$ratingHistogram$`1`
      df[[i]]$rating_2 <- page_data_full[[i]]$reviewsSummary$ratingHistogram$`2`
      df[[i]]$rating_3 <- page_data_full[[i]]$reviewsSummary$ratingHistogram$`3`
      df[[i]]$rating_4 <- page_data_full[[i]]$reviewsSummary$ratingHistogram$`4`
      df[[i]]$rating_5 <- page_data_full[[i]]$reviewsSummary$ratingHistogram$`5`
      df[[i]]$would_recommend <- page_data_full[[i]]$reviewsSummary$wouldRecommend
Count
      df[[i]]$would_not_recommend <- page_data_full[[i]]$reviewsSummary$wouldNotRe
commendCount
    }
  }
  df <- bind_rows(df, .id = "index")

  df_full <- left_join(output, df, by = "index") %>%
    gather(key = stars, value = ratings_count, rating_1:rating_5) %>%
    select(-webTitle) %>%
    janitor::clean_names(case = 'snake') %>%
    mutate(category = str_remove_all(category, pattern = START %R% "url_") %>% str
_remove_all("_" %R% DGT %R% END))

  df_full$ratings_count[is.na(df_full$ratings_count)] <- 0
```

```
  df_full
}

#run the function for each product category
women_pants <- get_web_data(url_women_pants)
men_pants <- get_web_data(url_men_pants)
women_bib <- get_web_data(url_women_bib)
men_bib <- get_web_data(url_men_bib)
men_clothes <- get_web_data(url_men_ski_clothes)
women_clothes <- get_web_data(url_women_ski_clothes)

#combine category dataframes into metacategories, and combine those into one compl
ete dataframe
df_clothes <- bind_rows(men_clothes, women_clothes) %>% select(-available_colors)
 %>% distinct()
df_pants <- bind_rows(men_pants, women_pants) %>% select(-available_colors) %>% di
stinct()
df_bib <- bind_rows(men_bib, women_bib) %>% select(-available_colors) %>% distinct
()

df_all <- bind_rows(df_clothes, df_pants, df_bib, .id = 'metacategory') %>%
  mutate(metacategory = case_when(
    metacategory == 1 ~ "Clothes",
    metacategory == 2 ~ "Pants",
    metacategory == 3 ~ "Bibs"
  ),
  metacategory = factor(metacategory, levels = c("Clothes", "Pants", "Bibs")))

#set up the prior distribution
test_reviews_total <- 5
amazon_weights <- c(.07, .06, .1, .19, .58)
test_reviews_weighted <- test_reviews_total * amazon_weights
stars_fct = factor(c('one_star', 'two_star', 'three_star', 'four_star', 'five_sta
r'), levels = c('one_star', 'two_star', 'three_star', 'four_star', 'five_star'))

df_prior <- tibble(test_reviews_weighted = test_reviews_weighted,
              stars = stars_fct) %>%
  spread(stars, test_reviews_weighted)

df_prior %>% kable() %>% kable_styling(full_width = FALSE)

#simulation data for prior distribution
sim <- data_frame(parameters = list(test_reviews_weighted),
                  name = map_chr(parameters, paste, collapse = ", ")) %>%
  mutate(simulation = map(parameters, ~ VGAM::rdiric(1e5, .))) %>%
  unnest(map(simulation, reshape2::melt, varnames = c("rep", "stars")))

#histogram of simulated prior values
ggplot(sim, aes(value)) +
```

```
  geom_histogram(binwidth = .05, boundary = 0) +
  facet_grid(. ~ stars) +
  labs(x = "Simulated Percentage of Ratings", y = "Density") +
  theme(axis.text.x = element_text(angle = 35, hjust = 1))

#average percentage of 1-5 star ratings for REI website snow apparel vs. Amazon pr
ior data
current_ratings <- df_all %>%
  filter(total_reviews > 0) %>%
  mutate(avg_count = ratings_count / total_reviews) %>%
  group_by(stars) %>%
  summarize(percent_of_ratings = mean(avg_count)) %>%
  mutate(percent_of_ratings = percent_of_ratings,
         stars = stars_fct) %>%
  spread(stars, percent_of_ratings) %>%
  bind_rows(tibble('one_star' = .07, 'two_star' = .06, 'three_star' = .1, 'four_st
ar' = .19, 'five_star' = .58)) %>%
  mutate_if(is.numeric, scales::percent) %>%
  mutate(data = c('REI', 'Amazon')) %>%
  select(data, everything())

current_ratings %>% kable() %>% kable_styling(full_width = FALSE)

#create primary dataframe for analysis by adding cleaned ratings
df_bayes <- df_all %>%
  filter(total_reviews > 0) %>%
  spread(stars, ratings_count) %>%
  mutate(avg_rating = (rating_1 + 2 * rating_2 + 3 * rating_3 + 4 * rating_4 + 5 *
rating_5) / total_reviews) %>%
  mutate(rating_bayes = ((rating_1 + df_prior$one_star) +
                         (rating_2 + df_prior$two_star) * 2 +
                         (rating_3 + df_prior$three_star) * 3 +
                         (rating_4 + df_prior$four_star) * 4 +
                         (rating_5 + df_prior$five_star) * 5) /
           (total_reviews + sum(test_reviews_weighted)))

#compute the mean of the prior distribution
prior_mean <- sum(c(1:5) * test_reviews_weighted) / sum(test_reviews_weighted)

#facetted plot of raw ratings vs. cleaned ratings
df_bayes %>% ggplot(aes(avg_rating, rating_bayes, color = total_reviews)) +
  geom_point() +
  geom_abline() +
  geom_hline(yintercept = prior_mean, linetype = 'dotted') +
  scale_color_continuous(trans = 'log2', breaks = c(1, 2, 4, 8, 16, 32, 64, 128),
 name = 'Total Reviews') +
  facet_grid(. ~ metacategory) +
  labs(x = 'Raw Rating', y = 'Bayesian Rating')
```

```r
#compare average rating of ski clothes to pants/bibs
df_bayes %>%
  select(prod_id, clean_title, rating_bayes, category) %>%
  mutate(clothing_type = ifelse(prod_id %in% union_all(df_bib$prod_id, df_pants$pr
od_id), 'Pants/Bibs', 'Other Clothes')) %>%
  filter(!((category == 'men_clothes' | category == 'women_clothes') & clothing_ty
pe == 'pants')) %>%
  distinct() %>%
  group_by(clothing_type) %>%
  summarize(rating = mean(rating_bayes)) %>%
  arrange(rating) %>%
  kable() %>%
  kable_styling(full_width = FALSE)

#boxplot of ratings by category
df_bayes %>%
  filter(metacategory != "Clothes") %>%
  mutate(bib_lgl = ifelse(prod_id %in% df_bib$prod_id, 'Bib', 'Pants')) %>%
  ggplot(aes(bib_lgl, rating_bayes)) +
  geom_boxplot() +
  labs(x = '', y = 'Rating')

#bibs vs. pants by gender
df_bayes %>%
  filter(metacategory != "Clothes") %>%
  select(prod_id, clean_title, rating_bayes, category) %>%
  mutate(bib_lgl = ifelse(prod_id %in% df_bib$prod_id, 'bib', 'no bib')) %>%
  filter(!((category == 'men_pants' | category == 'women_pants') & bib_lgl == 'bi
b')) %>%
  distinct() %>%
  group_by(category) %>%
  summarize(rating = mean(rating_bayes)) %>%
  kable() %>%
  kable_styling(full_width = FALSE)

#average engagement across all products by gender
df_bayes %>%
  select(prod_id, clean_title, rating_bayes, category, total_reviews) %>%
  mutate(gender = ifelse(str_detect(category, 'women'), 'women', 'men')) %>%
  distinct() %>%
  group_by(gender) %>%
  summarize(engagement = mean(total_reviews)) %>%
  kable() %>%
  kable_styling(full_width = FALSE)

#summary dataframe for ratings by brand and item
df_brand <- df_bayes %>%
  filter(metacategory != "Clothes") %>%
  gather(stars, star_count, rating_1:rating_5) %>%
```

```
  mutate(bib_lgl = ifelse(prod_id %in% df_bib$prod_id, 'bib', 'no bib')) %>%
  filter(!((category == 'men_pants' | category == 'women_pants') & bib_lgl == 'bi
b')) %>%
  group_by(brand, bib_lgl) %>%
  summarize(avg_brand_rating = mean(rating_bayes))

#plot brand ratings
df_brand %>%
  ggplot(aes(reorder(brand, avg_brand_rating), avg_brand_rating, color = bib_lgl))
+
  geom_point() +
  coord_flip() +
  labs(y = 'Average Brand Rating', x = '') +
  scale_color_manual(name = 'Category', labels = c('Bibs', 'Pants'), values = c("#
0C4742", "gray"))

#scatterplot of price vs. rating
df_bayes %>%
  ggplot(aes(rating_bayes, as.numeric(regular_price))) +
  geom_point() +
  labs(x = 'Rating', y = 'Price')
```

---

1. https://www.rei.com/blog/snowsports/the-year-of-the-bibs?cm_mmc=sm_tw_76514-_-content-_-orgcon-_-bib_year (https://www.rei.com/blog/snowsports/the-year-of-the-bibs?cm_mmc=sm_tw_76514-_-content-_-orgcon-_-bib_year)↩
2. http://www.columbia.edu/~on2110/Papers/Schoenmueller_netzer_stahl_2018.pdf (http://www.columbia.edu/~on2110/Papers/Schoenmueller_netzer_stahl_2018.pdf)↩
3. Spearman correlation is a more robust measure of correlation that considers non-linear, monotonic relationships.↩