Analysis of Invitations on Etsy

Ramin Ziaei

April 28, 2021

Introduction

After using all the filters on the "Invitation & Paper" page, each of them was saved as a separate csv file, and then, they were joined to make a new csv file that contains all of the options of a category. So, finally, I ended up having 8 different categories including: Orientation, Price, Color, Type, Theme, Season, Fold and Occasion. For each data set, there are 7 attributes, namely, Price, Discount, Ratings, Reviews, Best sellers, Free shipping and the category itself.

Exploratory Data Analysis

Orientation

There are 2 possible orientations for each card namely Horizontal and Vertical. Almost half of the cards are horizontal and the rest vertical. If we assume the number of written reviews is a representation of the number of people who actually bought the item, there are 40% more reviews for the Vertical cards compared to the Horizontal ones. Also, the discount for vertical cards is almost twice that of Horizontal ones. Finally, the sale price (after discount) is a little cheaper for the Vertical cards. If best seller items are of interest, 68% of the best sellers come from the Vertical cards while only 32% come from the Horizontal ones. Therefore, all the elements, suggest that making cards in a vertical format can be more appealing to the customers.

Color

There are many different options for the color on this page. However, to keep things easy to compare, I only chose the top 8 colors that contain almost 98% of all the products. These colors include, in order, Blue(17%), Black(17%), White(17%), Green(16%), Red(12%), Purple(9%), Yellow(6%) and Rose-gold(4%). The top three colors that received most reviews

on average are Red, Green and Blue. Ratings are really close for all of them (all around 4.9 out of 5), so this element may not be used to compare different colors. Green color is the one with highest discount (11% off) followed by White and Black in order. Finally, Yellow, Green and Red are cheapest colors among others in terms of their sale price. When it comes to best seller items, since there are different number of each color in the data set, it does not make sense to just compare the number of their best sellers. Instead, the number of best sellers for every 1000 items for each color will be used. The top 5 best seller colors in every 1000 products are Rose-gold with 4, White and Green each with 2.1, Black with 1.3 and Blue with 0.9 items. It is worth mentioning that although Rose-gold is the most expensive color in average (33% more expensive compared to the cheapest color), it is the best seller.

Type

There are two types of these cards namely Handmade and Vintage. While almost all the cards are Handmade (98%), only 2% of them are Vintage. The number of reviews for the Handmade items is only about 4% more than that of the Vintage products. While the discount for the Handmade items is almost 4 times that of the Vintage items, Their final price on average is almost equal. Vintage products have no best seller while Handmade items have 8 for each 1000 products.

Theme

There are 10 different themes to choose from including Classic(22%), Garden(21%), Rustic(15%), Minimalist(13%), Boho(8%), Beach(5%), Bling(4%), Destination(4%), Fairytale(4%) and Country(3%). Rustic has most reviews (which is 35% more than the second highest review), it is with the second highest discount, and finally, its sale price is the cheapest among all themes. Among all 10 categories, only 6 of them have best seller items; Rustic with 0.7 per 1000 items, Garden with 0.6 and Minimalist with 0.4 per 1000 products are top three of this list. It seems that the Rustic theme is the most popular theme among all the available themes on this page of the website.

Season

Another category in this page is 4 seasons where Spring, Summer, Fall and Winter contain 36%, 26%, 22% and 15% of the data, in order. Fall is the season that receives most reviews followed by Summer. Also the same pattern applies to their discounts. Finally, While Winter is cheapest card among the 4, Summer is the most expensive one considering the fact that their prices are pretty close to each other. When it comes down to best seller items, Fall

and Spring have no items in the list. Summer has 2.7 per 1000 items and Winter has 0.6.

Price

Prices are categorized into 4 categories: less than $10, between $10 and $100, between $100 and $200 and over $200. In order, these categories shape 28%, 31%, 26% and 15% of the invitation cards. As one may expect, the highest number of reviews, and consequently, the highest number of sales come from the cheapest category (under $10). The pattern is the same for other prices as well; The cheaper, the more reviews. It is interesting to notice that discount also follows the same pattern. The cheaper an item is, higher discount it offers. Discounts range from 6.5% for under $10 items to 0.23% for over $200 items. There are best sellers only from 2 of these categories, and again, as probably expected, those categories are the two cheapest ones. Under $10 items have 11.5 best sellers for each 1000 items and items between $10 and $100 have 3.5 best selling products.

Fold

These invitation cards can be either Folded or Flat. Folded cards compose a third of the cards and the rest are Flat. The number of reviews for the flat cards is 3 times that of the folded ones. Their discount is also 3 times the discount of the folded cards, 9% and 3% in order. Finally, Folded cards are about 40% more expensive compared to flat ones. Regarding the best selling items, Flat cards have 0.8 item per 1000 products, and Folded cards have 0.5.

Occasion

Invitation cards fall into 6 category of occasions namely: Bridal(37%), Wedding(34%), Engagement(14%), Bachelorette(9%), Anniversary(6%) and finally Bechelor(0.75%). Top 3 most reviews are from Bridal, Wedding and Bechelorette invitations while the least comes from Bachelor party. While Bachelor invitations seem to be the least favorite among others, they offer the highest discount (14%). The next two items with high discounts are Bridal and Bachelorette with 10% and 8% respectively. On average, Bridal cards are the cheapest($10) while Wedding invitations are the most expensive($21). For best selling items, Wedding ranks first with 1.8 per 1000 products. Then, in order, there are Anniversary, Bridal, Bachelorette and Engagement cards with 1.6, 0.3, 0.2 and 0.1 products per 1000 items.

Regression Analysis

To see the relation between different parameters of each category, a regression analysis was conducted. The dependent parameter was set to be the Price, and all other elements were chosen as independent. For all of them, Being a best seller is not statistically significant since its p-value has a big value. Number of reviews has negligible impact on the price. Rating and discount have an inverse relationship with the sales price. So, as the price goes down, discount goes up as was discussed in the Price section. Also, as the ratings go up, prices come down.

Results

Such an analysis can be really beneficial to any retailer or company since it gives insight as to what products are more favorable and appealing to customers. Based on such analyses, they can decide on what item they need to focus to increase their sales. Well, of course it is not as simple as that, but it can be a great help in the decision making process of any company. According to the previous discussions, if a designer or a company wants to come up with new cards, they better be, Vertical, Flat, Handmade, with a Rustic/Garden theme, with Fall/Summer seasons, and less than $10. Bridal and Wedding ceremonies are the more popular ones as well.

Discussion

One of the problems that I faced while scraping the Etsy website was, for some reason, some of its tags would change after a while. So, a working code will need a minor revision since the previous tag will not be usable anymore. It actually happened to me twice. One was for the "next page" tag, and the other one for the "ul" tag when parsing to get all items of a page. Also, in the beginning, I was not aware of the fact that each page has four rows of advertisement. So, from my perspective, the BeautifulSoup library could not capture all the items on a page. That is why I changed my code to make use of another package called "Selenium". That library would capture all 64 items on a page, including the advertisements. After consulting with the instructor, I realized that BeautifulSoup was working fine, so I changed my code back for the last time. Another issue arose when there were too many items in each category; the error I was getting was the error number 429 which means too many requests were sent from my end. For those items, I had to break down the data into smaller groups, scrape and join them later. Also, I wish, aside from plotly, we could cover static graphs and tables that are super useful when it comes to writing a report.