

Tania Papandrea
Homework 6: Data Collection for Final Project

Etsy is an online marketplace for artists and artisans to sell their goods. I've always been interested in opening an Etsy shop, but since it is such a large database it will be presumably difficult to make your own shop stand out from the rest. It becomes highly advantageous to appear early in the search results, as most customers will simply search for items they're looking for, only considering the top few results instead of fully exploring individual shops. This is frustrating for new shop owners because as in the real world, it is difficult to get business without being an established shop. I will define a 'successful' Etsy shop as one that has a high rank. I intend to investigate the factors that contribute most to becoming successful, and find other useful information that can help people who are looking to start an Etsy shop.

I did a search on Etsy for "knitted hats", and for each product on the list, I went to the storefront for the respective seller and web scraped for data. My resulting DataFrame shows several different characteristics for each shop. The shops are listed by their rank, or their position in the search results. -1 means that the specified information is not available. Here is an example of the information collected on this sampling of successful Etsy shops:

	Admirers	Age	Gift_card	Location	Name	Num_items	Num_reviews	Num_sections	Rating	Sales
1	1055	977	False	Gdynia, Poland	MyKnitCroch	233	229	10	4.8696	1646
2	1967	1685	False	Florence, Italy	viafashion	69	493	4	3.5143	2308
3	3647	1539	False	Europe, Turkey	Puik	18	441	2	4.9185	2541
4	1800	1635	True	Suncoast of FL	SandyCoastalDesigns	369	751	10	4.9091	3563
5	8209	2368	True	Los Angeles, CA, United States	MaxMelody	260	1220	9	4.8411	5013
6	1119	2653	True	Fort Collins, CO, United States	knittingtreasures	48	255	4	4.9268	711
7	331	741	False	-1	SunshineBabyProps	48	7	6	5.0000	263
8	1282	1147	True	Liverpool, England, United Kingdom	accessoriesbyrita	439	284	9	4.9036	1703
9	4712	2206	False	Europa	Ebruk	324	436	10	4.9085	2948
10	23719	2913	True	Zephyrhills, FL, United States	pixiebell	1661	6483	10	4.8909	25165

The items are organized with their rank (position in the search results) as their index. The information collected so far includes number of admirers, age (number of days since they joined Etsy), whether or not they accept Etsy gift cards as payment, location, total number of items in the shop, number of reviews, number of sections (categories for different types of products), average rating, and number of sales.

Out of the 225 shops on the first 5 pages of search results, the shop with the most items is pixiebell, who has the 10th rank. The shop with the most admirers is ileaiye, who has the 13th rank. The shop with the most reviews is lajumpinbean, who has the 77th rank.

Performing a linear regression on the data will allow me to isolate which attributes of a shop contribute most to its position in the search results. It seems logical that admirers, rating, number of items, and number of sales should all indicate a high rank, so there should be positive correlations between several of these factors and rank.