

Analyze the dataset linked below using Python and present your insights. Please do not do any manual data cleaning (only formula-based sanitization) , and do not use excel for this exercise.

This is a sample dataset of publicly available info on ecommerce sellers in the Garden category in the Amazon marketplace.

**Your task is to sanitize and analyze the data to profile the sellers present, and develop selection criteria to identify the best or most promising sellers in this dataset, that the Acquisitions team at Razor should reach out to, and acquire.**

Before you develop your selection criteria, you will need to sanitize the data , and parse out usable data from the unstructured text. Here are some tips:

- The column **sellerproductcount** gives you the count of products in the form '1-16 of over 100,000 results' , and you can parse out the product count 100,000 .
- **sellerratings** - this columns gives you the % and count of positive ratings (e.g. 88% positive in the last 12 months (118 ratings) ) if parsed correctly
- **sellerdetails** - you can use this text to parse out phone numbers, and email IDs of merchants, where available, so our team can reach out to them.
- **businessaddress** - this will give you the business locations of the sellers. You can parse them to identify if a seller is registered in the US , Germany (DE), or China (CN). Note that Razor does not acquire Chinese sellers at this point, so you can use this data to exclude sellers in China from your analysis.
- **Hero Product 1 #ratings and Hero Product 2 #ratings** - these 2 columns give you the number of ratings of the 2 'hero products' or best selling products of this seller.

Please present your insights in the form of a Jupyter Notebook. We will assess you on your data parsing process, code quality as well as the insights and selection criteria you develop to analyze the dataset and present the best sellers. **You may download the dataset from this [link](#).**