Monica Siegel

SI 206

Project 4 – Newman

25 March 2018 – 11 April 2018 UPDATE

Ben and Jerry’s Ice Cream

For this final project I want to crawl and scrape Ben and Jerry’s ice cream. From the crawling and scrape I want to select Ice Cream Pints from the product type and from there select the flavors of ice cream with a description of what is in each pint, along with the rating and ingredients in both dairy and non-dairy. I also want to select the top 10 flavors of pint ice cream based on their ratings (which would be another scrape). From what I have described above I think that I would give myself a challenge score of 8.

I want to create two databases. The first one will be a ‘pint’ database. I want this database to include the ice cream name, description of the ice cream, the ingredients in the ice cream, and the average rating for the ice cream. The second database will be a ‘non-dairy’ database. This database will include the flavor name, the ingredients in it, and the description. For my foreign key I will have the ‘non-dairy’ database include a column that is linked to my ‘pint’ database if the non-dairy flavor is also offered in the regular pint section. Additionally, I will incorporate a link to a website where users can look at the picture of the ice cream.

I plan to use plotly to track the top 10 Ben and Jerry’s flavors, the ratings of the flavors, a graph that shows how much of an ingredient is in a flavor, and how many non-dairy flavors are also regular pint flavors. This will require me to scrape again into the Sears reports on the Ben and Jerry’s website.

UPDATE 4/11:

* I have decided to scrape only the pint ice creams instead of also scraping the dairy-free version.
* I am now also finding reviews, if product is recommended, and account name
* My two data bases will be ‘flavors’ and ‘reviews’
  + I want the ‘flavor’ database to include the ice cream name, description of the ice cream, and the ingredients in the ice cream. I have not decided yet, but I additionally might incorporate a link to a website where users can look at the picture of the ice cream as well.
  + The ‘reviews’ database will include the id, account name, the review, the rating, and the flavor id. The flavor id will be the foreign key that is connected to the flavor id in the ‘flavor’ table.
* For plotly I am still going to track the top 10 Ben and Jerry’s flavors, but additionally now the ratings of the flavors, the distribution of ratings from all reviewers, the top ice cream ingredients used in flavors, and the 10 most frequent words used in a description.