Progress 2

Goal:

- [x] ~~Decide on data source~~
- [x] ~~Collect Product Information~~
- [x] ~~Import and read in MATLAB~~

Work Done:

- After I had my list of potential user input ideas I could decide on what information I needed to collect from a chosen retailer.
- I chose MECCA as they do not currently have a recommendation system and finding non-bias specific recommendations can be hard as an Australian as we have limited brands available here. Additionally, their site is very ease to scrape and has no blockers.
- This phase mostly relied on learning I have been doing outside this course with building simple web scraper in python.
- Once I had the CSV I read it into MATLAB and applied basic commands to check that it was being read properly.

Test Case:

- Showing first 5 rows of dataset in MATLAB

First 5 rows of the dataset:

ProductNumber	Brand	ProductName	Price	Category	Subcategory	Weight_g_	Weight_ml_	StockStatus	ReviewCount	Rating
2	{'MECCA COSMETICA' }	{'Gel Cream 50ml' }	{'$55.00' }	{'Skincare'}	{'Moisturisers' }	{'N/A'}	50	{'In Stock'}	58	4.6
3	{'Summer Fridays' }	{'Lip Butter Balm' }	{'$39.00' }	{'Skincare'}	{'Lip Care' }	{'N/A'}	NaN	{'In Stock'}	239	4.3
4	{'Dr. Dennis Gross' }	{'DermInfusions™ Fill + Repair Eye Cream 15ml'}	{'$118.00' }	{'Skincare'}	{'Eye Care' }	{'N/A'}	15	{'In Stock'}	NaN	0
5	{'Youth To The People'}	{'Superfood Cleanser' }	{'$24.00–$67.00'}	{'Skincare'}	{'Cleanser & Toner'}	{'N/A'}	NaN	{'In Stock'}	3919	4.8
6	{'Tower 28' }	{'SOS Daily Balancing Gel Cleanser 150ml' }	{'$33.00' }	{'Skincare'}	{'Cleanser & Toner'}	{'N/A'}	150	{'In Stock'}	284	4.7

Reflection:

- It quickly became noticeable that the file would need typecasting as many of the variables were not being read as simple strings
- From here I was able to see what could and couldn't be read directly from the file so I could start taking note of what would need cleaning.