Page 1



**Overview**

Best Buy is the largest specialty retailer in the United States [consumer electronics](https://en.wikipedia.org/wiki/Consumer_electronics) retail industry.  It was originally founded by [Richard M. Schulze](https://en.wikipedia.org/wiki/Richard_M._Schulze) and James Wheeler in 1966as an audio specialty store called Sound of Music. In 1983, it was re-branded under its current name with an emphasis placed on consumer electronics.

Page 2

**My research on Best Buy Refrigerators**

Best Buy is one of the largest kitchen appliance retailers throughout the US. I am curious about how brands, number of reviews, ratings and styles affect the refrigerators’ prices move. I used scrape to scrape all the information I need for every single refrigerator.

**Questions**

1. What’s the most important factor influence the variation of the price?

* Prices vs brands?
* Prices vs number of reviews?
* Prices bs ratings?
* Prices vs style?

1. Do higher highest number reviews cause the lower ratings?

Page 3

**Method**

**Web Scraping Technique:**

To get the data on each product, I created a spider in scrapy, the spider will go through each result page which shows 24 products per page. Spider will also go into individual product page to extract the data I need. I extracted 1000+ products after the spider went through the loop.

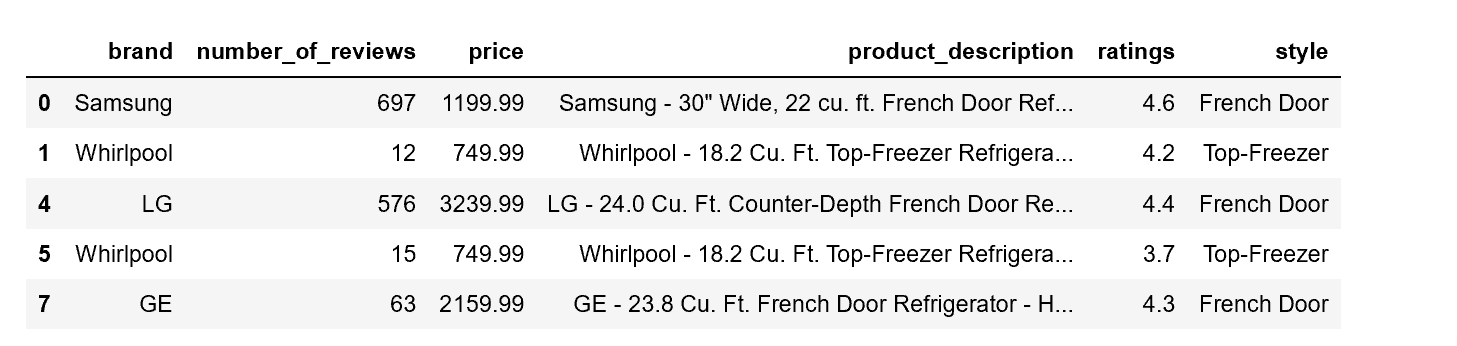
For each refrigerator I found the following factors for my analysis:

* Brand
* Number of reviews
* Ratings
* Product description
* Style

Page 4

**Data Cleaning**

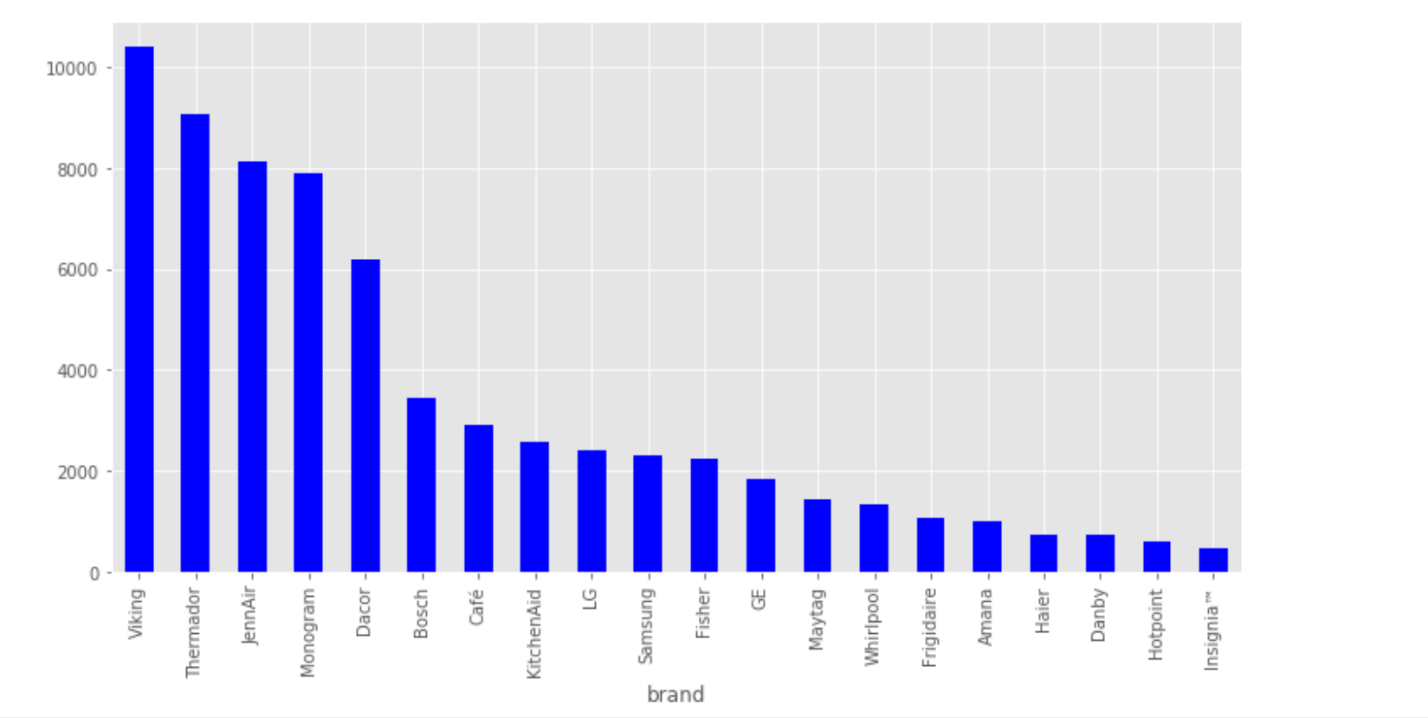
It took me a couple of hours to scrape all the data I need. The majority of the raw data I grabbed from the websites is not in a workable form. Since I have a tremendous difficulty to grab the style category. I decided to extract the “style” column from the “product description”. See the screenshot below. Product\_description column contains many useful information, such as the style of each refrigerator. I created a new style column by using regular expression by filter through all the words. That probably is the biggest challenge through the entire cleaning process. After get the new style column, there were still many NaN value in different columns. Dropna is the method to filter through all the data. Eventually I retrieved a completed workable form data. However, the number of rows reduced to 662 rows



**Page 5**

**Conclusions:**

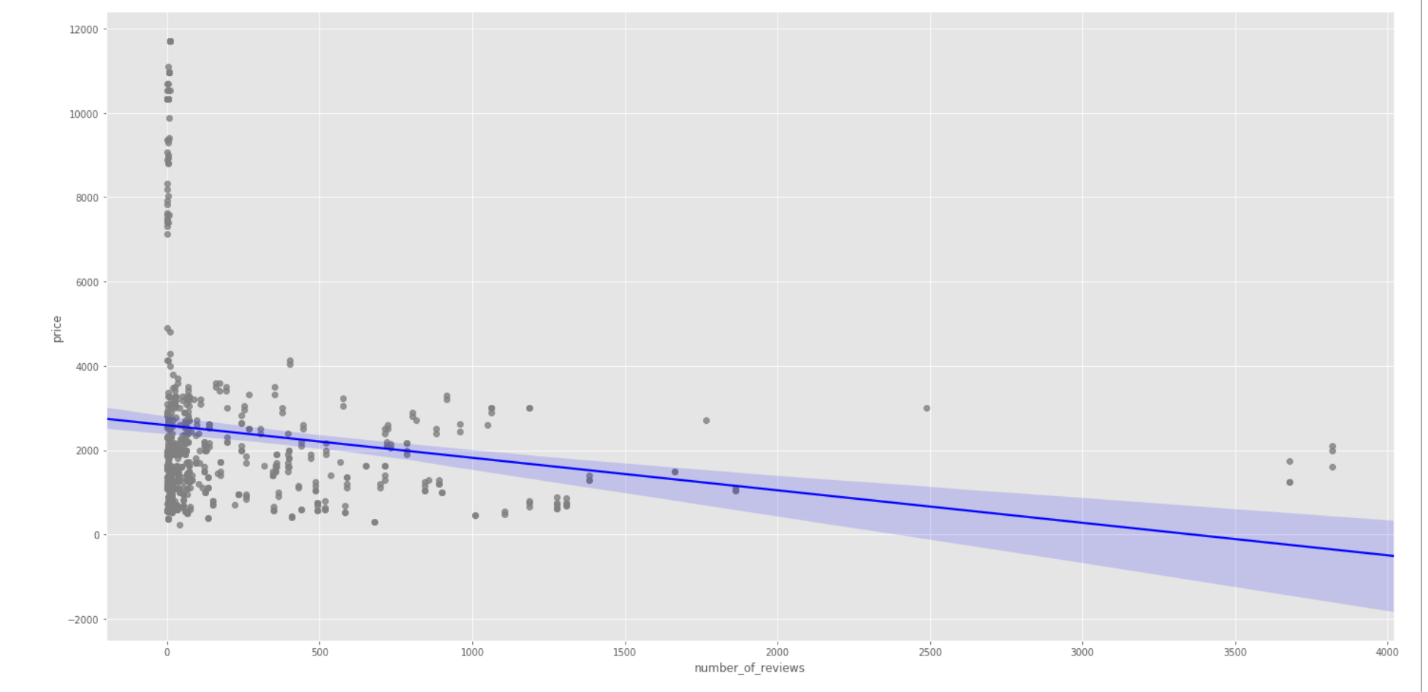
1. Check the relationship between brand and price by visualizing those 2 variables.



Based on the chart, we do see a down trend of mean price by each brand. After I researched some brands, I discovered that brand has a significant impact on the price variation. For example. “Viking” only manufactures customized high-end refrigerators. Most of price for “Viking” are over $10,000. Therefore, brand is key factor influence the prices.

Page 6

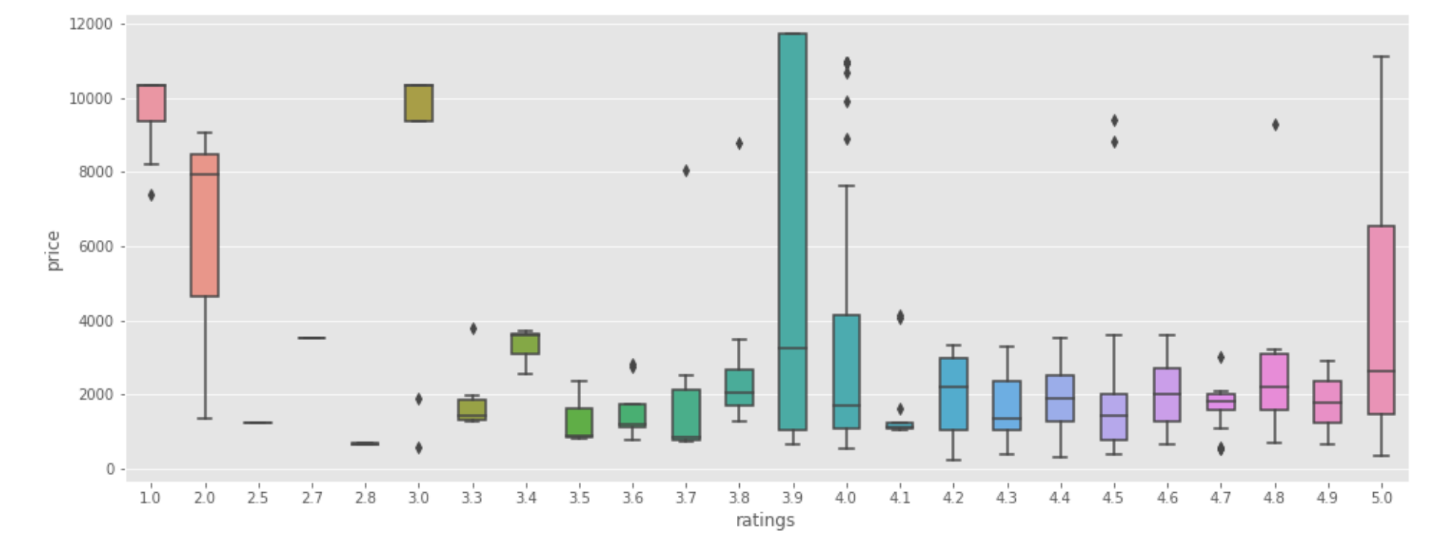
1. Relationship between number of reviews and prices.



I used linear plot method to generate the above chart. Obviously, the price and number of reviews has an inverse relationship. Less reviews on a product have higher price of the product. This is not necessarily true that the variation of the prices is based on the number of reviews.

Page 7

1. Relationship between ratings and prices



Above graph shows price range for each rating point. As showed in graph, we can see the widest price range is on rating 3.9. (from $300 to $12,000). Some of the products are only rated as 1.0, but the prices are more or less than $10,000. Therefore, there is no relationship between ratings and prices. Ratings will not affect the price range for each product.

Page 8

1. Relationship between style and price

