The purpose of the Challenge was to analyze Amazon Vine reviews written by members of the paid Amazon Vine program and non-Vine members. Amazon Vine is a service that allows manufacturers and publishers to receive reviews of their products and determine if there are any biases among the Vine members and non-Vine members reviews. Companies will sometimes provide free products to Vine members who are expected to publish a review. I used the furniture reviews for my analysis:

https://s3.amazonaws.com/amazon-reviews pds/tsv/amazon\_reviews\_us\_Furniture\_v1\_00.tsv.gz

To determine if there is bias from Vine members vs non-Vine members the data was filtered to only use rows where the “total\_votes” count is equal to or greater than 20 in order to pick reviews that are more likely to be helpful and to avoid having division by zero errors later on in the analysis. The data was further filtered to retrieve all the rows where the number of “helpful\_votes” divided by “total\_votes” is equal to or greater than 50%. The data below was extracted from the data frame, **paid** indicates a verified purchase.

- Total Number of Helpful Vine and Non-Vine Reviews Paid and Not Paid: 18,155

- Total Number of Helpful Vine Reviews Paid: 3

- Total Number of Helpful Vine Reviews Not Paid: 133

- Total Number of Helpful Non-Vine Reviews Paid: 15,080

- Total Number of Helpful Non-Vine Reviews Not Paid: 2,939

Five Star

- Total Number of Helpful Five Star Reviews: 8,556

- Total Number of Helpful Five Star Vine Reviews Paid: 2

- Total Number of Helpful Five Star Vine Reviews Not Paid: 72

- Total Number of Helpful Five Star Non-Vine Reviews Paid: 7,323

- Total Number of Helpful Five Star Non-Vine Reviews Not Paid: 1,159

Five Star Percentage

- Percent of Five Star Vine Reviews Paid: 67

- Percent of Five Star Vine Reviews Not Paid: 54

- Percent of Five Star Vine Reviews Paid: 49

- Percent of Five Star Vine Reviews Not Paid: 39

Review Totals

- Total Number of Reviews: 792,113

- Total Number of 'Helpful' Reviews: 18,155

- Total Number of Five Star Reviews: 447,716

The counts of Vine and non-Vine reviews were placed in a new data frame using Pandas and plotted on a bar graph using Matplotlib.

The percentage of Vine and non-Vine reviews were placed in a new data frame using Pandas and plotted on a bar graph using Matplotlib.

The following were used to complete the analysis:

* Spark
* PySpark
* Python
* Pandas
* Matplotlib
* Google Colaboratory

Of the 50 datasets, I chose to analyze reviews that were made by users in the **"** **Furniture"** category.

Dataset used for this analysis can be found [here](https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_us_Electronics_v1_00.tsv.gz)