**Module 16 Challenge: Amazon Vine Analysis: (Colab / Pyspark)**

**Overview:**

The purpose of this analysis is to analyze Amazon reviews written by members for paid Amazon Vine Program and compare those to non-paid public review submitted. The objective is to determine if there is a bias from the Vine members as they submit reviews of products. My review used the luggage category from the Amazon data base and utilizing Google Colab and Pyspark to perform the ETL from AWS webserver.

**Analysis Results:**

### Total Qualified Luggage Reviews Vine (paid) vs. non-Vine (unpaid):

* There was a total of 6,711 luggage review after applying the following filters:
* Total votes were greater than or equal to 20
* Helpful votes were more than or equal to 50% of total votes
* 6,690 reviews from non-Vine while only 21 reviews were from Vine members
* Almost 99% of the Luggage reviews in the filtered data set came from non-Vine users

Text

Description automatically generated with medium confidence

Text

Description automatically generated

Text, Word

Description automatically generated

### Five-star reviews Vine (paid) Reviews vs Non-Vine (unpaid):

* Total five-star reviews from the filtered data were 3,458, which is 51% of all the luggage reviews.
* 3,448 of the five-star reviews were from non-Vine users, while only 10 were from Vine members.

A picture containing logo

Description automatically generated A picture containing company name

Description automatically generated Text

Description automatically generated

### Percentage of five-star reviews for Vine (paid) vs non-Vine (unpaid):

* 47.6% of the Vine reviews were five-star
* 51.3% of the non-Vine reviews were five-star

A picture containing text

Description automatically generatedText

Description automatically generated with low confidence

### **Summary:**

There is not a positive bias for Vine (paid) reviews verses non-Vine (unpaid) reviews in Amazon’s Luggage category. In fact, a great percentage of the unpaid review were listed as five-star reviews. This dataset had an insignificant number of paid reviews with over 99% of the luggage reviews coming from unpaid users.