**Amazon Vine Analysis**

Overview

In this analysis, we are analyzing Amazon reviews written by members of the paid Amazon Vine Program. Amazon Vine program is a service that allows manufacturers and publishers to receive reviews for their products.

For this challenge, we are analyzing the Amazon reviews for US Wireless phones. Using PySpark, we will Extract the dataset, Transform the data and connect to an AWS RDS instance and Load the transformed data into pgAdmin.

In the first deliverable, we extracted the dataset for US Wireless phones and transformed the dataset into four DataFrames (The Customers table, The Products Table, The Review ID Table, and the Vine Table DataFrames).

In the second deliverable, we created a DataFrame for the Vine Table, and filtered the DataFrame by total votes greater than or equal to 20 votes, percentage of helpful votes equal to or greater than 50%, created DataFrame where there is and isn’t a Vine review, filtered the DataFrame by total number of reviews, number of 5-star reviews and percentage of 5-star reviews calculated for all Vine and non-Vine reviews.

Results:

* How many Vine reviews and non-Vine reviews were there? According to our data, there were 613 paid vine reviews and 64,968 unpaid non-Vine reviews

Graphical user interface, text

Description automatically generated

Graphical user interface, text

Description automatically generated

* How many Vine reviews were 5-stars? How many non-Vine reviews were 5 stars? 222 Vine reviews were 5-stars and 30,543 non-Vine reviews were 5-stars

Text

Description automatically generated with low confidence

Graphical user interface

Description automatically generated with low confidence

* What percentage of Vine reviews were 5 stars? What percentage of non-Vine reviews were 5 stars? 36.22% of Vine reviews were 5-stars while 47.01% of the non-Vine reviews were 5-stars

Text

Description automatically generated with medium confidence

Text

Description automatically generated with medium confidence

Summary

Based on our findings, we can conclude that there isn’t any positivity bias for reviews in the Vine program as 5-star reviews for Vine members (36%) is less than the 5-star reviews for non-Vine members (47%). In fact, there were only 613 Vine members reviews while non-Vine members total reviews was 64,968 and 30,543 out of that were 5-star reviews.

Additionally, we can also calculate measures of Central Tendency by calculating the mean, median and mode of the 5-star reviews to determine how skewed the dataset will be.