**Amazon Vine Analysis (Video Game 5 Star Reviews)**

In this analysis, Pyspark was utilized using Google Colabs to extract Amazon Video Game data, transformed into digestible/workable states, and loaded to PgAdmin servers for use. In this analysis approach, Pysparks was utilized for analysis and deliverables to capture correlation of 5 star videogame reviews for accounts that were Vine subscribers vs those that were not.

Before data was analyzed, steps were taken to clean the data where ‘total\_votes’ counts were equal or greater than 20 to prevent errors with zero divisions from populating. Dates were also converted from ‘string’ to ‘date-time’ to allow for data fluidity upon loading of data to servers.

The cleaned data was then further filtered to capture percentages where ‘helpful\_votes’/’total\_votes’ equaled or were greater than 50% to serve as the launch point of whether having a Vine account made a difference in the total of 5-Star rating reviews. Upon analyzing the data, 5 star reviews had a higher correlation with paid Vine accounts with 51.06% compared to non-subscribed Vine accounts with 38.71%. Although there is a correlation leading to paid accounts having higher 5-Star reviews, the sample size between the two types had a delta of 40,377 (Y = 94, N= 40,471). In order to find true correlation, recommendation is to compare sample sizes that are similar in magnitude as the sample difference was too great.

In conclusion, recommendation for 5-Star reviews would help to have a paid Vine subscription, but would need additional sample size to provide a more concrete result. Furthermore, it is recommended to run Machine Learning algorithms (supervised or unsupervised) to Train, Test, and fit these modals to obtain correlation, sensitivity, and utilize for predictive and prescriptive analysis.