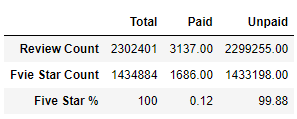
**Summary**

### Overview of the analysis:

US outdoors review of Amazon review datasets was picked and we performed the ETL to extract the dataset, transform the data, connect to NY AWS RDS instance, and load the transformed data into pgAdmin. Next, I performed analysis using Jupyternotebook to determine the proportion of vine reviews per favorable and unfavorable reviews. Finally, I did a descriptive analysis based on overall 5 star reviews whether paid or unpaid.

### Results:

[](https://github.com/jilldvn/Amazon_Vine_Analysis/blob/main/image/Rview_Five_Star_Vine.png)

#### 1. How many Vine reviews and non-Vine reviews were there?

* paid\_reviews\_count (vine): 3137,
* unpaid\_reviews\_count (non-vine): 2299255

#### 2. How many Vine reviews were 5 stars? How many non-Vine reviews were 5 stars?

* paid\_five\_star\_count (vine): 1686,
* unpaid\_five\_star\_count (non-vine): 1433198

#### 3. What percentage of Vine reviews were 5 stars? What percentage of non-Vine reviews were 5 stars?

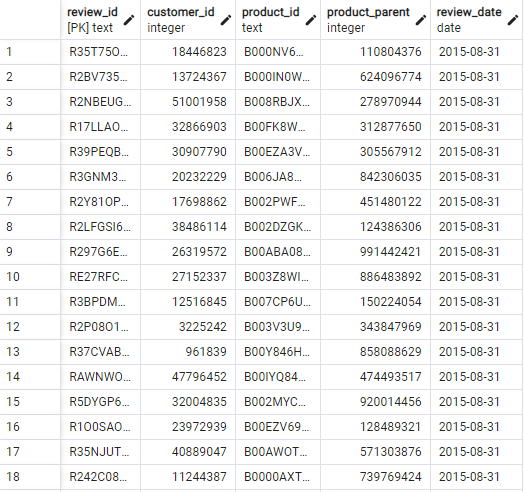
* paid\_five\_star (vine): 0.12%,
* unpaid\_five\_star (non-vine): 99.88%

### Findings:

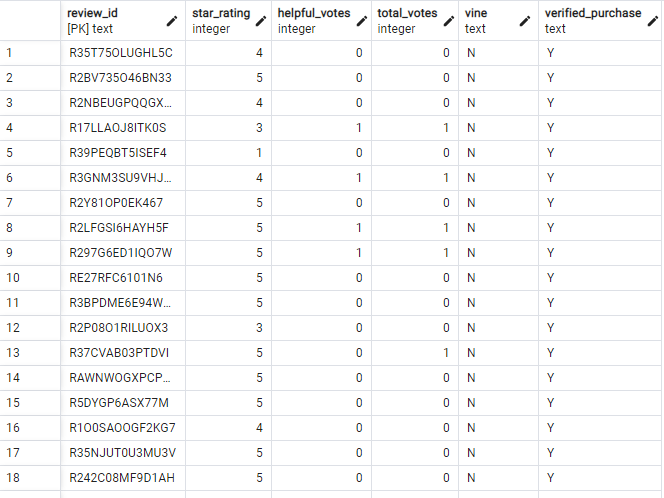
A bias could be observed due to a five-star calculation based on the overall vine and non-vine reviews from overall reviews without filtering the dataset if other customers feel the star rating is helpful or not. As a resukt, the total 5 star reviews count could be subjective. Besides, the star rating count per vine or non-vine is not grouped by product parent. With all being said, for further analysis I recommend:

1. Vine and non vine 5 star analysis should be done after filtering the data with a higher portion of helpful votes.
2. Join the tables of Review\_ID\_Table and Vine\_Table (as shown below) using review\_ID as primary key, then regroup the data based on product parent, in this case, a comparison of star rating count with vine or non-vine under the same category of product might be more meaningful.

* review\_id\_table

[](https://github.com/jilldvn/Amazon_Vine_Analysis/blob/main/image/review_id_table.png)

* vine\_table

[](https://github.com/jilldvn/Amazon_Vine_Analysis/blob/main/image/vine_table.png)