# \*\*Amazon\_vine\*\*

## Overview of Project

This project’s objective is to analyse Amazon reviews written by members of the paid Amazon Vine program.

The deliverables for this project are:

\* Perform ETL on Amazon Product.

\* Determine Bias of Vine.

## Software

PySpark, AWS, Google Colab and SQL.

## Results

### Perform ETL on Amazon Product.

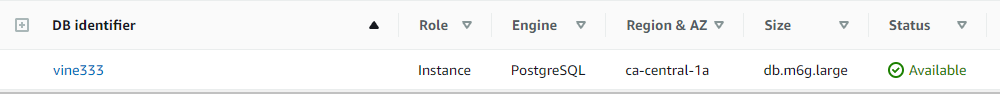
Using Google Colab and PySpark, a dataset from Amazon Review dataset was extracted as dataframe.

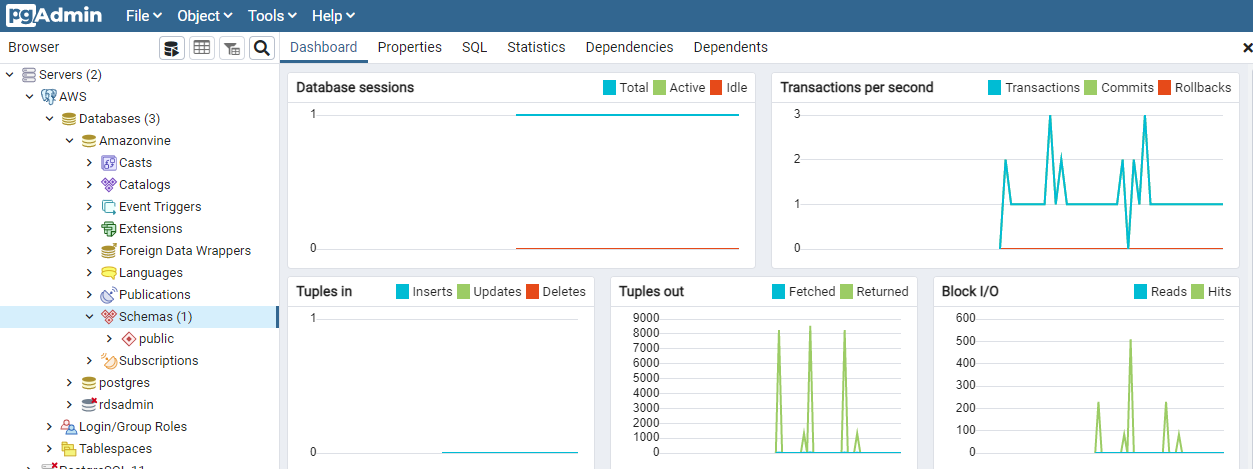


The dataframe was divided in four datasets and uploaded in AWS RDS database. The new dataframes are customer, product, review and vine.



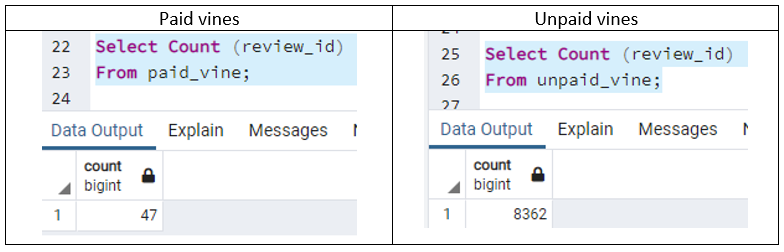
The new dataframes were transferred from AWS RDS to pgAdmin.





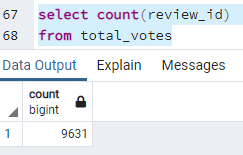
### Determine Bias of Vine.

* How many Vine reviews and non-Vine reviews were there?

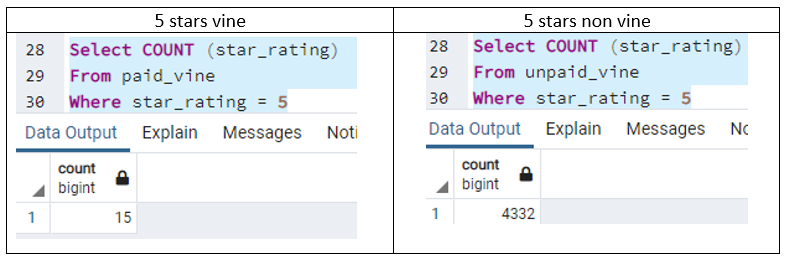


|  |  |
| --- | --- |
| Paid vines | Unpaid vines |
|  |  |

* How many Vine reviews and non-Vine reviews were there?



* How many Vine reviews were 5 stars? How many non-Vine reviews were 5 stars?



|  |  |
| --- | --- |
| 5 stars vine | 5 stars non vine |
|  |  |

* What percentage of Vine reviews were 5 stars? What percentage of non-Vine reviews were 5 stars?



###Summary

In this analysis, any vine with less than 20 reviews were excluded from the total count. The analysis shows that there is no bias in the Amazon vine reviews as the numbers of paid reviews is significantly smaller than the unpaid reviews.