select vine, count(\*) as reviews\_count, avg(star\_rating) as avg\_star\_rating,

avg(helpful\_votes) as avg\_helpful\_votes, avg(total\_votes) as avg\_total\_votes,

(avg(helpful\_votes) / avg(total\_votes)) as percentage\_helpful\_votes

from vine\_table

where vine = 'Y' or vine = 'N'

group by vine

;

--filter based on # of votes. chose 7 being the average # of votes received by vine reviewers

select vine, count(\*) as reviews\_count, avg(star\_rating) as avg\_star\_rating,

avg(helpful\_votes) as avg\_helpful\_votes, avg(total\_votes) as avg\_total\_votes,

(avg(helpful\_votes) / avg(total\_votes)) as percentage\_helpful\_votes

from vine\_table

where vine = 'Y' and total\_votes > 7 or vine = 'N' and total\_votes > 7

group by vine

;

--checking for duplicates...some customers reviewed products in both datasets and are showing up twice in customer table

select customer\_id, count(customer\_id) from customers

group by customer\_id

having count (customer\_id) > 1

order by count(customer\_count) desc ;

select \* from customers

where customer\_id = 5179904 or customer\_id = 5179869

select review\_id, count(review\_id) from review\_id\_table

group by review\_id

having count(review\_id) > 1

select customers.customer\_id, count(customers.customer\_count) as total\_ratings, avg(vine\_table.star\_rating) as avg\_star\_rating

from review\_id\_table

join vine\_table on vine\_table.review\_id = review\_id\_table.review\_id

join customers on review\_id\_table.customer\_id = customers.customer\_id

where customers.customer\_count > 100

group by customers.customer\_id

order by total\_ratings desc