How do businesses react to COVID-19

In this web page, we will analyze how business react to the COVID-19. We jointly use the data from two sources: the Yelp business dataset(https://www.yelp.com/dataset/download) and the CMU COVIDcast dataset(https://covidcast.cmu.edu/). We will use these data to explore how different factors effect the businesses after COVID-19 happened.

1. Business state overview

Let's first look at the raw dataframe from Yelp COVID19 dataset.

	business_id	highlights	delivery or takeout	Grubhub enabled	Ca
0	9kXRUIkwdDtnAPO6tVo51g	FALSE	FALSE	FALSE	
1	H6D5HOTfMjrZt7r1EObZ1g	FALSE	FALSE	FALSE	
2	FYddq7fUtzobZcw4jOJgVA	FALSE	FALSE	FALSE	
3	c75jLTjlgA9q3gImLEGT6w	FALSE	FALSE	FALSE	
4	YfzPiY50h_10Sjlg3mnNWQ	FALSE	FALSE	FALSE	

The features of the original dataset can be divided into four types:

- Bool type: **delivery or takeout**, **Grubhub enabled**, **Call To Action**, **Request a Quote Enabled**
- Json type: **highlights**
- Str type: **Covid Banner**, **Virtual Services Offered**
- Datetime type: **Temporary Closed Until**

In our exploration, we use these features as the indicators for the states of the businesses. Before further explore what factor will influence these features, lets first try to analyze the correlation of them.

1.1 Covid feature correlation

We first change non-bool type features into bool types to see whether there is some inner correlation among these features.

☐ Show new data

Now let's explore whether these features are related. You may select several interested features below and press GO to run.

Show your interested subsets

Choose an option

GO!

You may also want to see how multi features affect certain feature.

Select one affected feature.

highlights

Gooo!

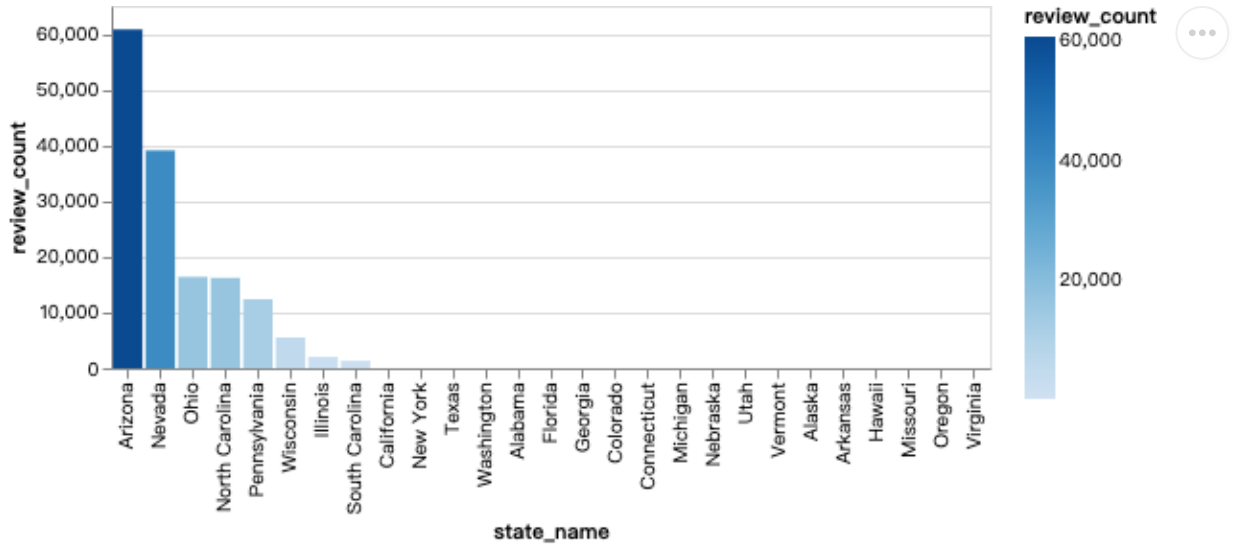
2. How Businesses' location affect their reaction?

2.1 How is the situation for the business in different county/state?

First, we try to have an overview that how many data points belongs to eachs states. This will give intuition about the missing values.

Select one target about the businesses' condition you want to explore more.

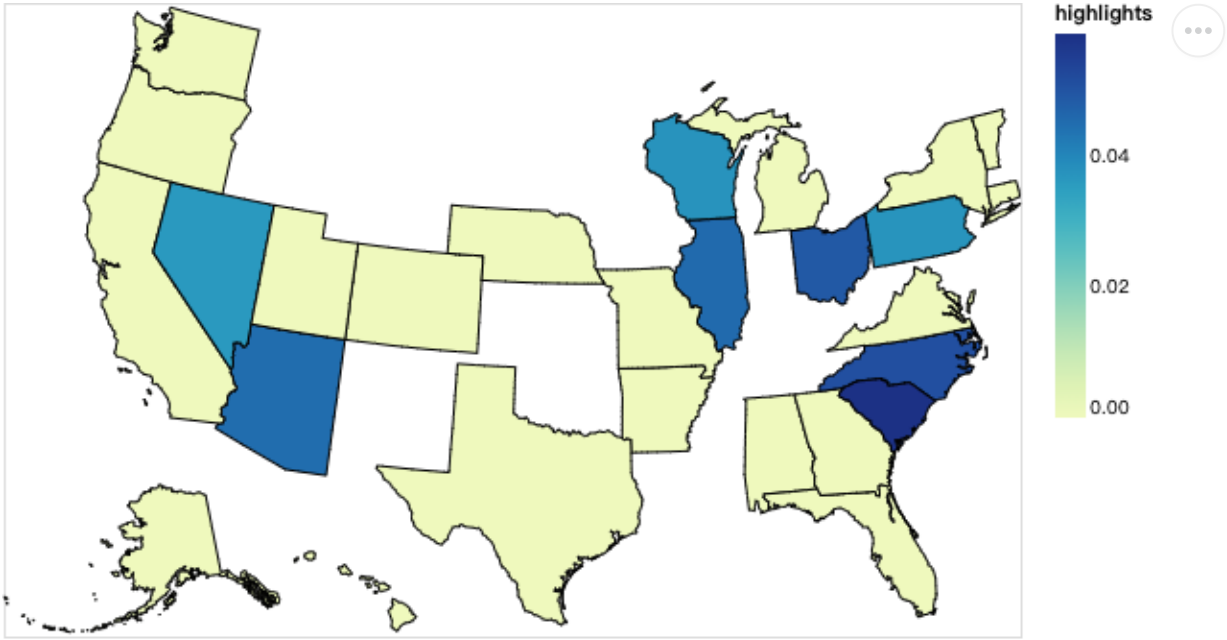
highlights



We can see that for most of the states, there are nearly no businesses' data point.

Then we will make more exploration about what happen to The following figure show that the selected feature value for businesses in each state/county, the color refers to the TRUE rate for selected bool varaible.

highlights rate in each state



> Notice that we have preprocessed the data. The original dataset contains data from countries other than U.S., e.g., data from Canada. We remove all these data entries. The white state/county here refer to that there is no business record in these places.

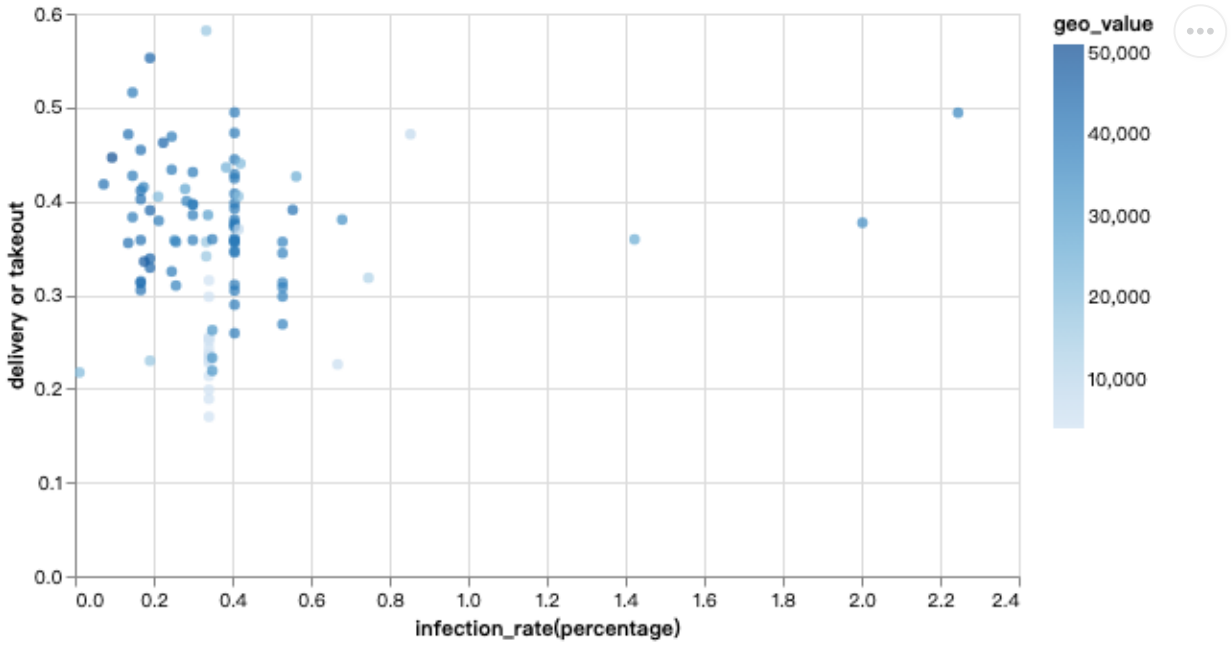
2.2 Reasons for the differences?

Why there are differences between businesses in different locations? Let's explore more about this. A possible reason is that different place have different infection rate. Let's find out how infection rate will affect the businesses situation!

You may also want to see how is the situation for different country across the country.

Select one target you want to explore more.

delivery or takeout

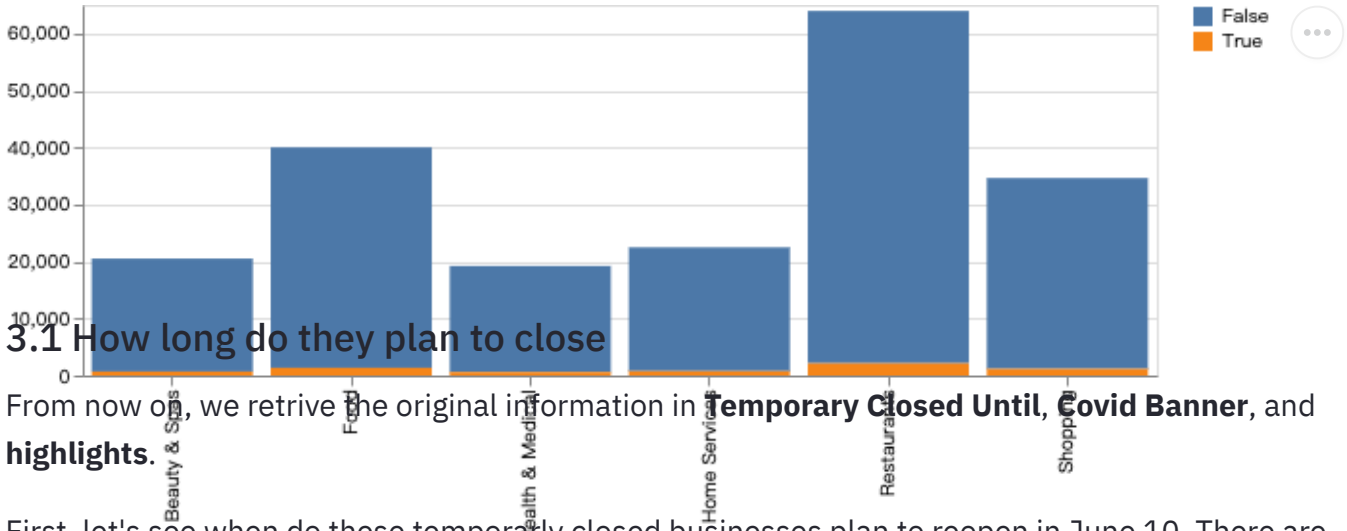


3. How Businesses' affect their reaction?

Let's now explore how businesses of different categories behave. We start by looking at whether different categories react differently with the above COVID features.

Select a Covid feature you are interested in:

highlights

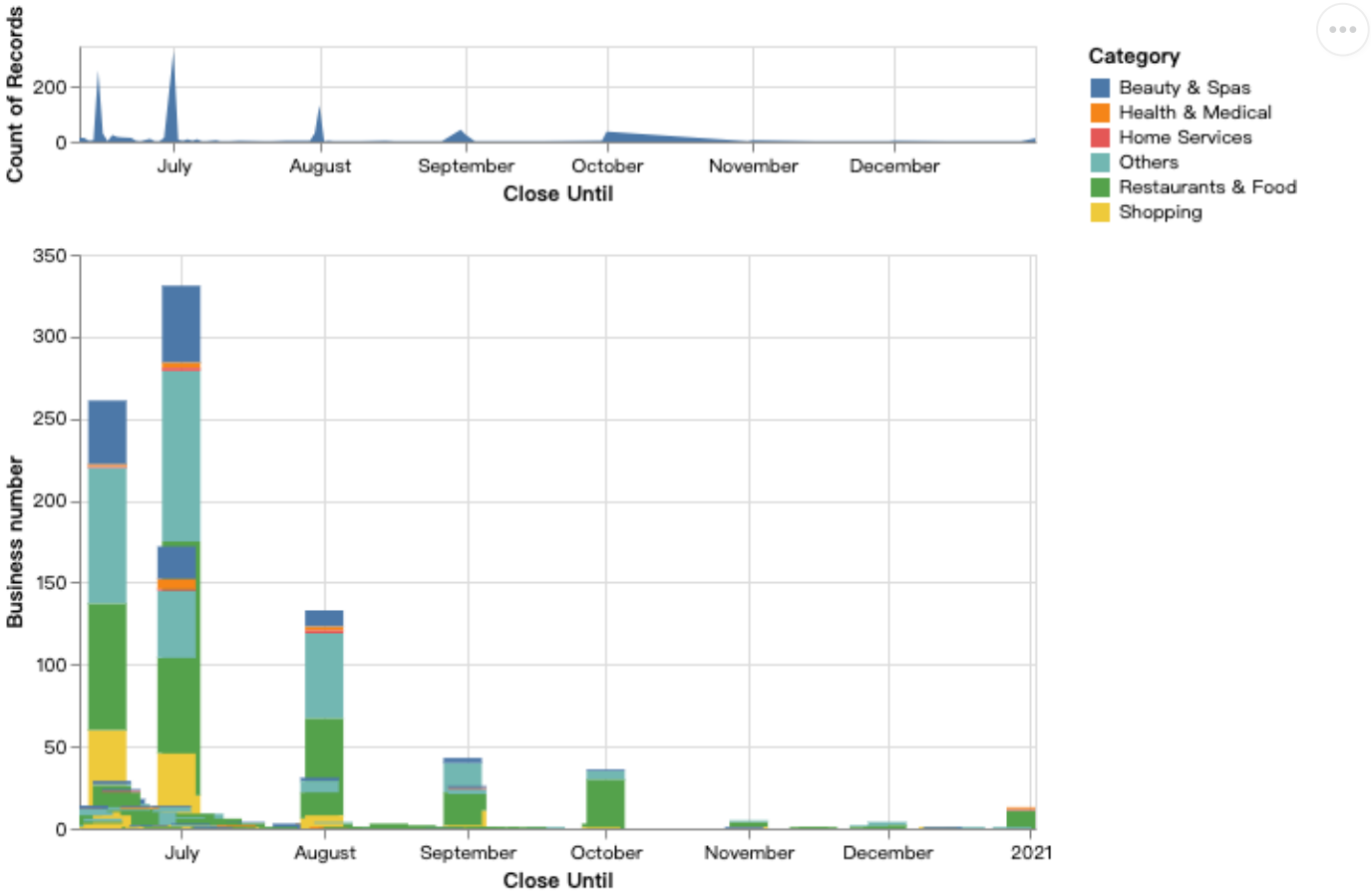


3.1 How long do they plan to close

From now on, we retrive the original information in **Temporary Closed Until**, **Covid Banner**, and **highlights**.

First, let's see when do these temporarily closed businesses plan to reopen in June 10. There are 1379 businesses uploading closure notification.

You may select certain category you are interested in from the bottom box, and select certain time sub-interval to take a closer look, which could be done through bruching on the upper figurg.



Category of Restaurants & Food

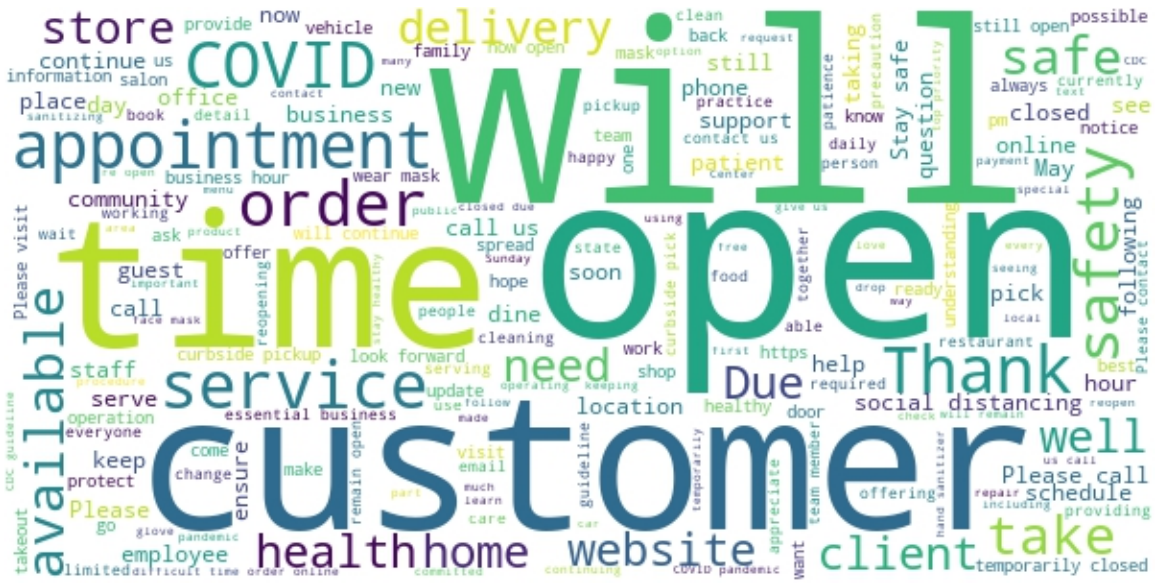
3.2 What do Covid Banner say

Choose a category you are interested in:

whole

Next, let's turn to Covid Banner. There are 13895 businesses in all uploading their Covid Banner.

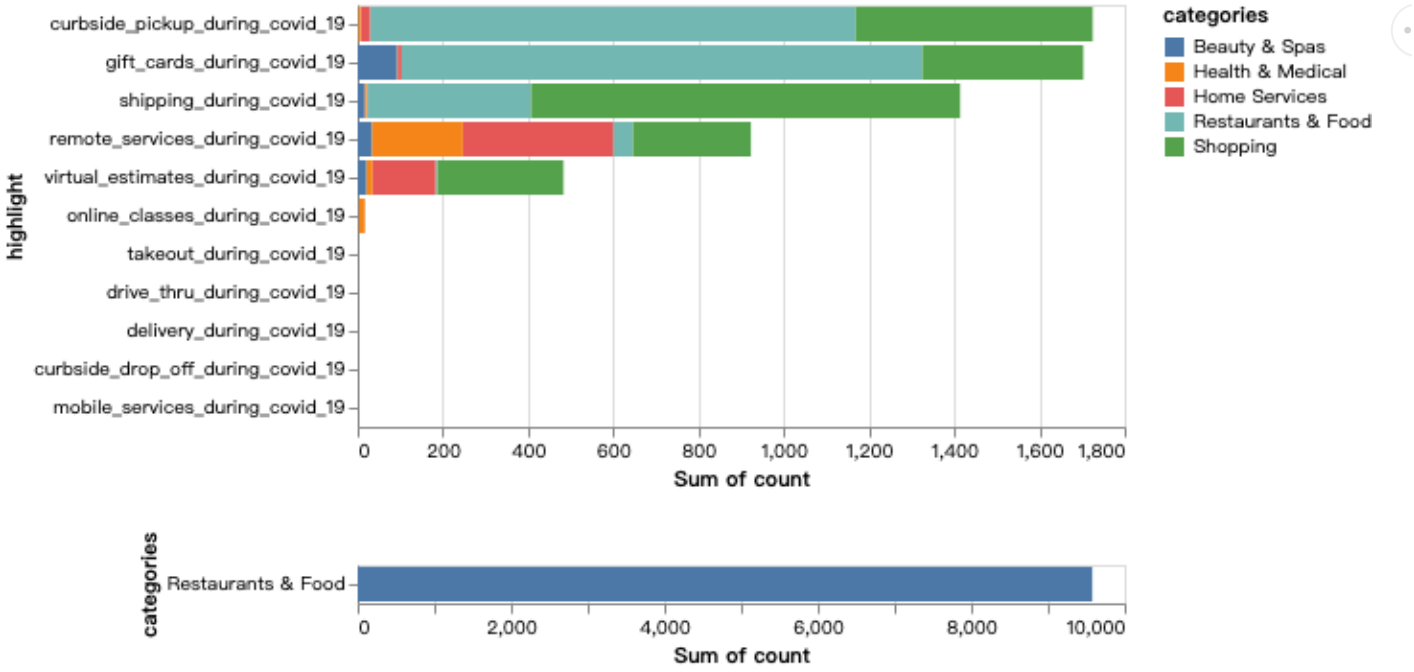
You may want see what words frequently appear in that banner, in whole, or in certain category.



3.3 What are in the highlights

Now let's take a look at hightlights, which are put in the business information page. There are 7045 businesses in all updating their highlights in response to Covid 19. We can see what highlights are about.

Below, we extract 11 covid 19 related highlights, and show their numbers decomposing in different categories. You may select one highlight and see in more detailed about its decomposition in the second figure. You may also select one particular category and compare highlight types of business belong to that category.



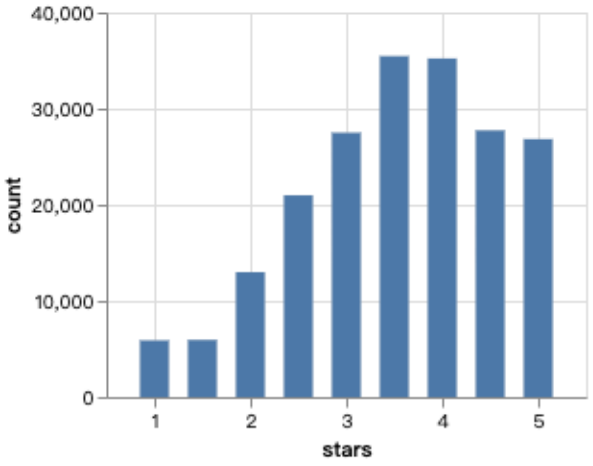
4. How businesses' quality affect their reaction?

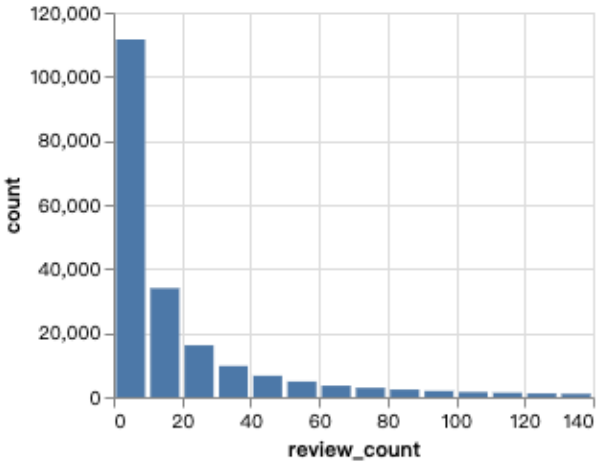
Does business quality before COVID-19 have some relationship with their state during COVID-19? We would look at their popularity, measured by review counts, and their ratings.

4.1 Quality overview

We first look at the distribution of **star ratings** and **review counts** in our dataset. Recall that the quality data is collected in March 2020, when the COVID was not that a serious concern, and thus could be used as a previous quality measure.

Brush certain intervals from both figures to explore into certain stars and/or review counts.





You may find that **stars** are relatively scattered, while for **review counts**, they are pretty concentrated, and actually, though the most popular business can have more than 10,000 reviews, the 95 percentile of review count is 145.

4.2 Stars, review counts, and COVID features

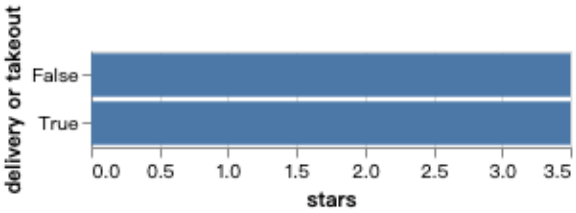
Now, let's explore into the relationship among **stars**, **review counts** and **COVID features**. It is likely that business states are affected by their previous quality.

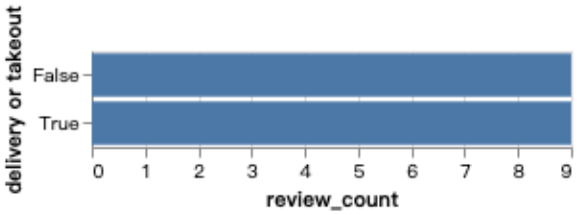
Let's start by selecting one **COVID feature** to see whether a business has it or not depends on stars and review counts.

Select the covid feature you want to explore:

delivery or takeout

Let's see how the whether the median of stars and review counts are different between businesses having delivery or takeout, and those not.





We can go deeper into individuals! However, due to the huge size of original dataset, we use sampling strategy. Let's choose a **sample size**!

Select how mamy points you want to sample for each case:

100

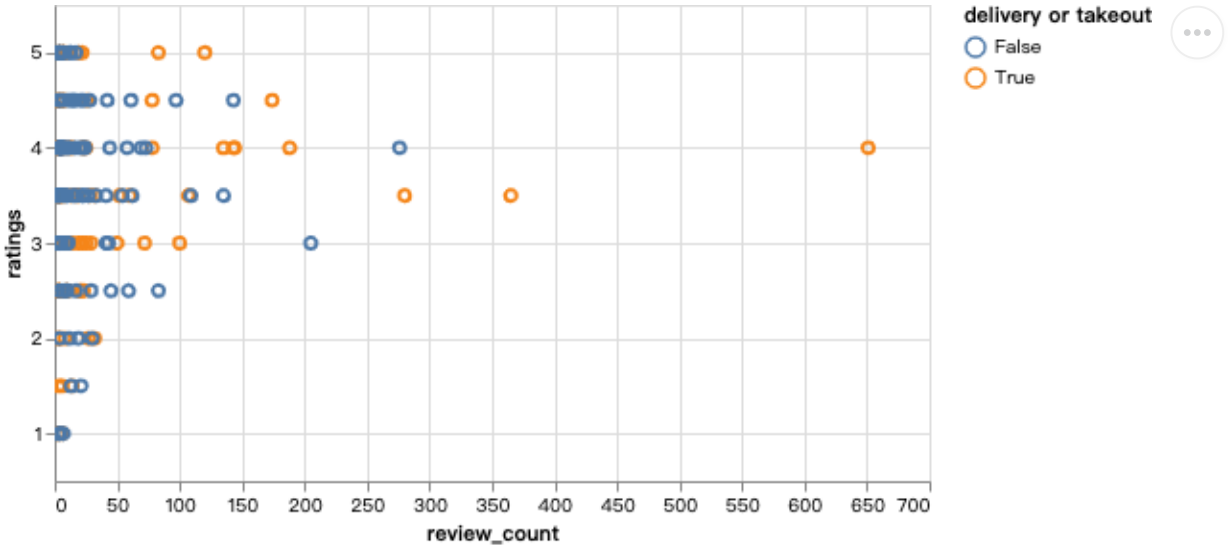
50 500

You may want filter some extremely large review counts, according to your observation in the **Quality Overview**. By default, we are showing you the whole range.

I only want to sample points from review counts under:

10130 − +

Now you have your graph! Try to double click one point or a True/False on the right to highlight a certain group.



You may see that stars do not influence much, while more popular businesses are more likely to react.

5. Using ML model to prediction?

Now we try to deploy a machine model to predict the business reaction for a selected business data point. We try to make this into a classification problem so we make all the prediction target to be True/False, even for highlights and covid banner, we use the location, category, ratings and review_count as the prediction factors.

To simplify the problem, we transfer all first two factors into one-hot-vector. Please notice that location refers to the state here. We set 80% of the data as the training set and the rest as test set.

Select one target you want to explore more.

Request a Quote Enabled ▾

The dataset after preprocessing is like this:

	stars	review_count	Restaurants	Shopping	Food	Home Services	Beauty &
0	3.5000	36	0	1	0	0	
1	5	4	0	0	0	0	
2	5	5	0	0	0	0	
3	2.5000	3	0	1	0	1	
4	4.5000	26	0	0	0	1	

Choose the ML Model

Logistic Regression

The prediction acc is: 98.82038301733607

Made with Streamlit