# Battle of Neighborhoods Houston

**Author: Arnav Mangal**
**Creation Date: 2/3/2021**

Introduction:

With the recent COVID-19 crisis, big cities have been impacted the most with their numbers skyrocketing recently (due to population density). Harris County's Public Health department has revealed that there is an average increase of over 1,000 new cases daily. Specifically in Houston, this upwards trend has negative implications for the future of current (active) non-chain restaurants; many will go out of business.

However, this also means that following this pandemic, this empty void can be refilled with new restaurants. Given how big the city of Houston is, how can these restaurant venturers determine where to open up? The goal of this analysis project is to determine the best prospective neighborhoods to open up a restaurant in Houston. One of the most important factors to consider when determining where to place a restaurant, is the type of restaurants that are popular in a certain area, which we will find by using K Means Machine Learning to cluster venues from the Foursquare API.

Data:

We will be obtaining the neighborhood name data from this page: https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods. It shows that there are a total of 87 Neighborhoods in Houston. We will be taking the relative location column's data as well, but ignoring the approximate boundaries column for each neighborhood, as this information is irrelevant to our research. Using the Nominatim package, we will be finding the longitude and latitude coordinates of these neighborhoods to then access the Foursquare API, which contains location data. Specifically, we will access data about restaurants in each neighborhood, to determine the most popular types of restaurants and form clusters based on this.

First five entries of Houston's neighborhood dataset (from wikipedia):

| # | Name | Location relative to Downtown Houston | Approximate boundaries |
|---|---|---|---|
| 1 | Willowbrook | Northwest | Along Texas State Highway 249 northwest of Beltway 8 |
| 2 | Greater Greenspoint | North | Around the junction of Beltway 8 and Interstate 45 North |
| 3 | Carverdale | Northwest | South of the junction of Beltway 8 and U.S. Route 290 |
| 4 | Fairbanks / Northwest Crossing | Northwest | Along U.S. Route 290 between Interstate 610 and Beltway 8 |
| 5 | Greater Inwood | Northwest | North of Fairbanks / Northwest Crossing and east of Acres Home |

As you can see, the only information that concerns us is the Neighborhood names (column 1) and Relative Location to Downtown (Column 2).

Methodology:

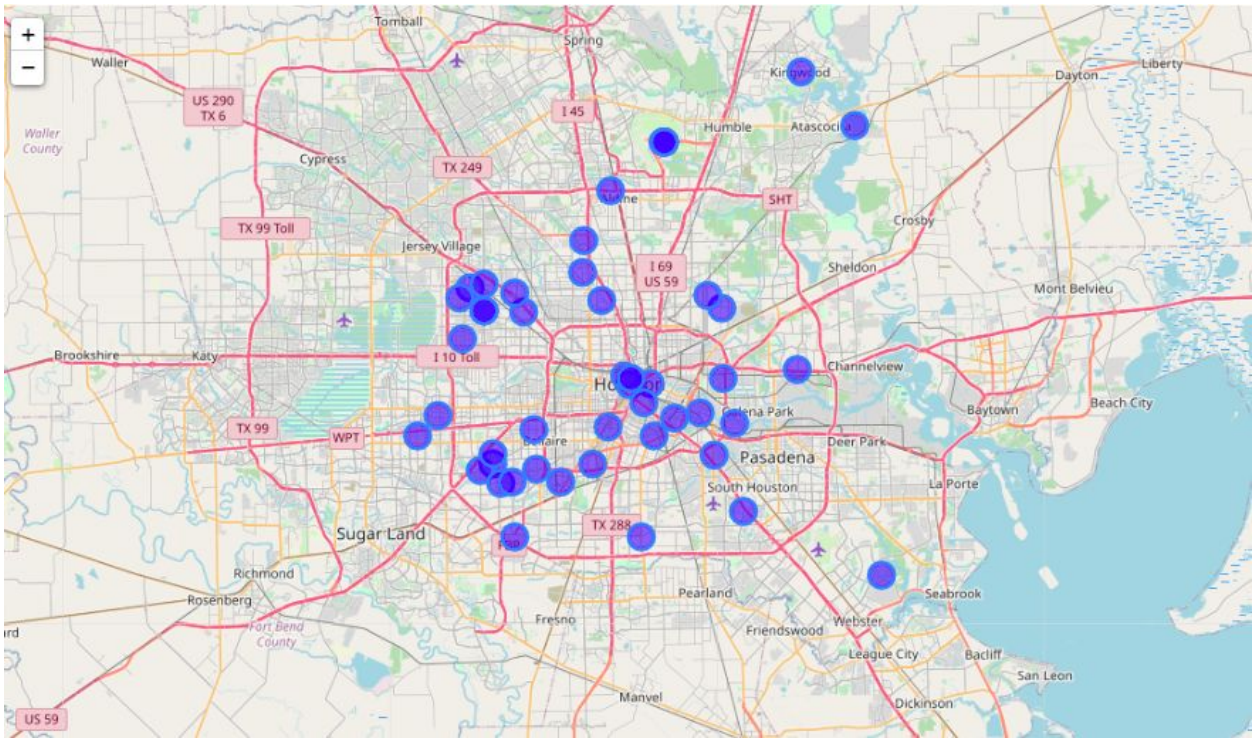
To start analyzing the Houston data, we first need to scrap the data from the Wikipedia page and add it into a pandas dataframe. We can keep the first two columns, which inform each Neighborhood's name and relative location within Houston.

Using the Nominatum package, we can find the latitude and longitude coordinates of each neighborhood. The dataframe should then look like this:

| | Neighborhood | Relative_Location | Latitude | Longitude |
|---|---|---|---|---|
| 1 | Greater Greenspoint | North, Houston | 29.938997 | -95.389077 |
| 2 | Carverdale | Northwest, Houston | 29.848687 | -95.539450 |
| 3 | Fairbanks / Northwest Crossing | Northwest, Houston | 29.852726 | -95.524386 |
| 4 | Greater Inwood | Northwest, Houston | NaN | NaN |
| 5 | Acres Home | Northwest, Houston | 32.634285 | -83.689674 |

As you can see, there are some Neighborhoods that have missing Lat/Long coordinates. We will drop these Neighborhoods from the dataframe.

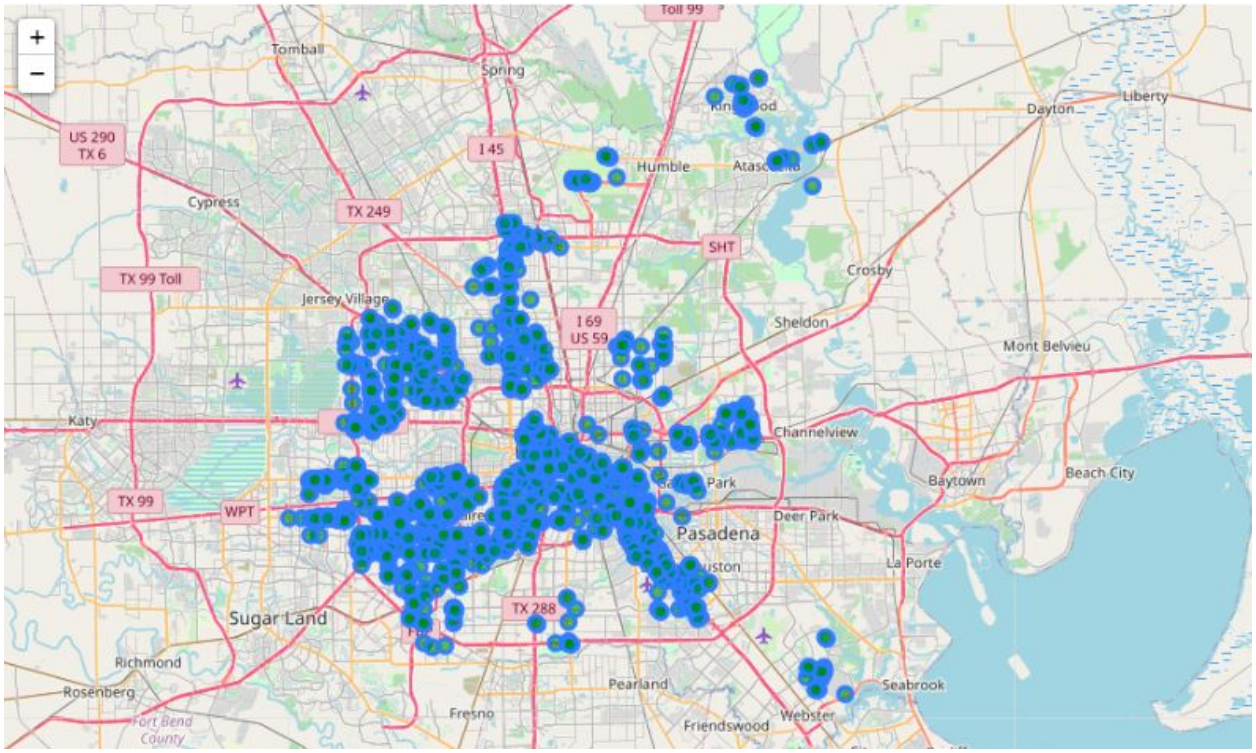
We can use foliam to generate a map of houston with the neighborhoods:




Now that we have the location data, we will use the Foursquare API to get venues within 3000 meters of each neighborhood, and also their venue category. We can add this to the current data table we have:

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Greater Greenspoint | 29.938997 | -95.389077 | Super Chicken | 29.937545 | -95.396312 | Fried Chicken Joint |
| 1 | Greater Greenspoint | 29.938997 | -95.389077 | Jimmy G's Cajun Seafood Restaurant | 29.941073 | -95.400880 | Cajun / Creole Restaurant |
| 2 | Greater Greenspoint | 29.938997 | -95.389077 | Pappas Seafood House | 29.929609 | -95.412938 | Seafood Restaurant |
| 3 | Greater Greenspoint | 29.938997 | -95.389077 | Popeyes Louisiana Kitchen | 29.938028 | -95.398008 | Fried Chicken Joint |
| 4 | Greater Greenspoint | 29.938997 | -95.389077 | Schlotzsky's | 29.933500 | -95.377600 | Restaurant |

And Graph this into Folium:




This folium map displays the geographic location for each venue in houston.

We next will make a data table with each neighborhood vs the number of restaurants they contain that fall under each restaurant type.

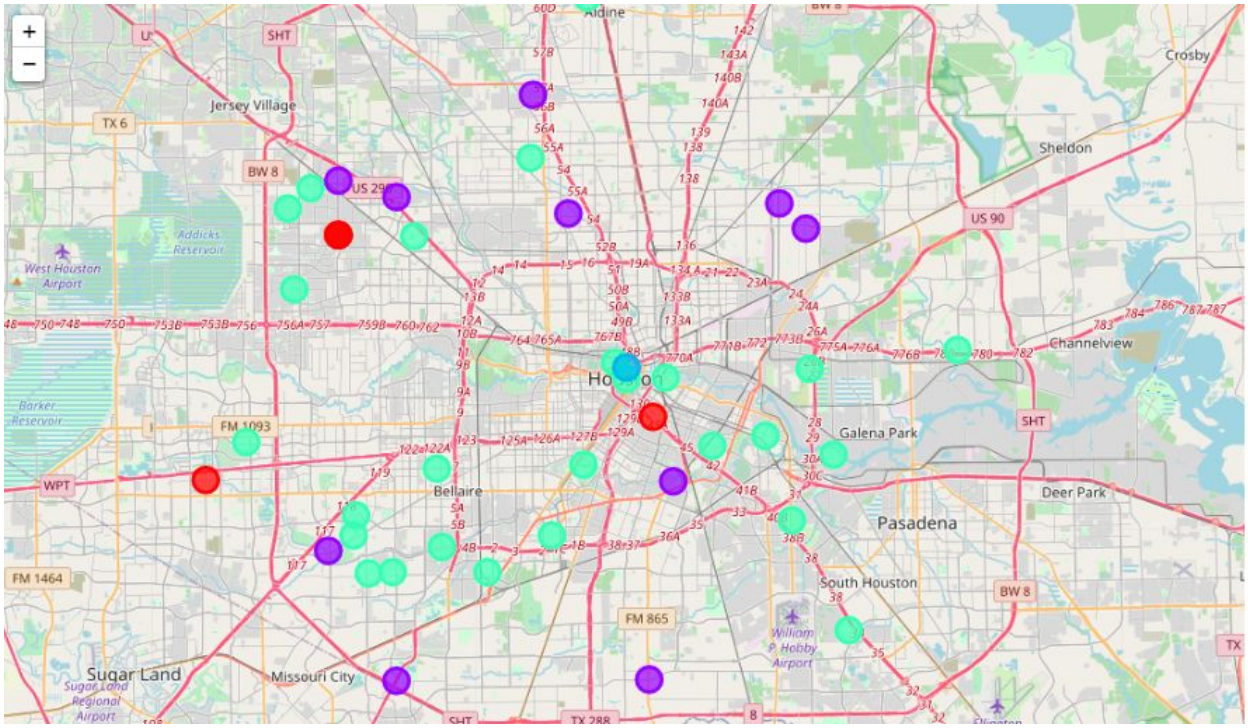
Using one-hot encoding, it is possible for us to see which restaurant type is most common in each neighborhood. This can help us create the following data set:

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th M Comm Ve |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Acres Home | Fast Food Restaurant | Pizza Place | American Restaurant | Chinese Restaurant | Mexican Restaurant | Sandwich Place | Japanese Restaurant | Wings Joint | Steakhouse | Breakfa Spot |
| 1 | Alief | Vietnamese Restaurant | Sandwich Place | Pizza Place | Cajun / Creole Restaurant | Fast Food Restaurant | Café | Breakfast Spot | Noodle House | Restaurant | Wings Joint |
| 2 | Astrodome Area | Pizza Place | Sandwich Place | Burger Joint | Italian Restaurant | BBQ Joint | Restaurant | Fast Food Restaurant | Mexican Restaurant | Chinese Restaurant | Breakfa Spot |
| 3 | Braeburn | Fast Food Restaurant | Fried Chicken Joint | Mexican Restaurant | Pizza Place | Donut Shop | Food | Sandwich Place | BBQ Joint | Food Truck | Filipino Restau |
| 4 | Braeswood | Mexican Restaurant | Food | Donut Shop | Fried Chicken Joint | Chinese Restaurant | Cajun / Creole Restaurant | Fast Food Restaurant | Breakfast Spot | Café | Burger Joint |
| 5 | Carverdale | Café | Sandwich Place | BBQ Joint | Donut Shop | Chinese Restaurant | Fast Food Restaurant | Mexican Restaurant | Asian Restaurant | Taco Place | Burger Joint |

(this data only represents the first six neighborhoods)

Finally, it is time to use K Means Machine Learning to create five clusters. Each cluster will share similar common restaurant types.

The neighborhoods our now each assigned to a cluster, each cluster represented in a different color when graphed in folium.




When looking at the results next, we can be clued in as to why cluster data may be important to prospective restaurant owners.

<u>Results:</u>

**Cluster 1**

| | index | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 24 | Alief | Vietnamese Restaurant | Sandwich Place | Pizza Place | Cajun / Creole Restaurant | Fast Food Restaurant | Café | Breakfast Spot | Noodle House | Restaurant | Wings Joint |
| 1 | 61 | Midtown | Vietnamese Restaurant | Thai Restaurant | Fried Chicken Joint | Café | Empanada Restaurant | Deli / Bodega | Dim Sum Restaurant | Diner | Doner Restaurant | Donut Shop |

**Cluster 2**

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | Fairbanks / Northwest Crossing | Fast Food Restaurant | Food | Pizza Place | American Restaurant | Deli / Bodega | Seafood Restaurant | Food Truck | Vietnamese Restaurant | Mexican Restaurant | Burger Joint |
| 6 | Hidden Valley | Fast Food Restaurant | Fried Chicken Joint | Mexican Restaurant | Pizza Place | Sandwich Place | Breakfast Spot | Burger Joint | Taco Place | Donut Shop | Bakery |
| 12 | Independence Heights | Fast Food Restaurant | Mexican Restaurant | Fried Chicken Joint | Pizza Place | BBQ Joint | Wings Joint | Sandwich Place | Chinese Restaurant | Food Truck | Donut Shop |
| 29 | Braeburn | Fast Food Restaurant | Fried Chicken Joint | Mexican Restaurant | Pizza Place | Donut Shop | Food | Sandwich Place | BBQ Joint | Food Truck | Filipino Restaurant |
| 40 | Fort Bend / Houston | Fast Food Restaurant | Sandwich Place | Fried Chicken Joint | Food | Food Truck | Chinese Restaurant | Seafood Restaurant | Mexican Restaurant | Burger Joint | Bakery |
| 43 | Lake Houston | Seafood Restaurant | Fast Food Restaurant | Breakfast Spot | Fried Chicken Joint | Vietnamese Restaurant | Food Truck | Food | Bakery | Sushi Restaurant | Italian Restaurant |
| 48 | East Houston | Fast Food Restaurant | Fried Chicken Joint | Food | Sandwich Place | Seafood Restaurant | Pizza Place | Mexican Restaurant | American Restaurant | BBQ Joint | Donut Shop |
| 49 | Settegast | Fried Chicken Joint | Fast Food Restaurant | Donut Shop | American Restaurant | Pizza Place | BBQ Joint | Bakery | Restaurant | Sandwich Place | Seafood Restaurant |
| 69 | Pecan Park | Fast Food Restaurant | Fried Chicken Joint | Pizza Place | Food Truck | BBQ Joint | Chinese Restaurant | Cajun / Creole Restaurant | Latin American Restaurant | Sandwich Place | Seafood Restaurant |
| 76 | Minnetex | Fast Food Restaurant | Chinese Restaurant | Fried Chicken Joint | BBQ Joint | Restaurant | Food | Sandwich Place | Seafood Restaurant | Burger Joint | Donut Shop |
| | | | | Fried | | | | | | | |

## Cluster 3

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 60 | Downtown | Sandwich Place | Wings Joint | Creperie | Deli / Bodega | Dim Sum Restaurant | Diner | Doner Restaurant | Donut Shop | Dumpling Restaurant | Eastern European Restaurant |

## Cluster 4

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Greater Greenspoint | Fast Food Restaurant | Sandwich Place | American Restaurant | Restaurant | Mexican Restaurant | Breakfast Spot | Fried Chicken Joint | Seafood Restaurant | Pizza Place | Chines Restau |
| 2 | Carverdale | Café | Sandwich Place | BBQ Joint | Donut Shop | Chinese Restaurant | Fast Food Restaurant | Mexican Restaurant | Asian Restaurant | Taco Place | Burger Joint |
| 5 | Acres Home | Fast Food Restaurant | Pizza Place | American Restaurant | Chinese Restaurant | Mexican Restaurant | Sandwich Place | Japanese Restaurant | Wings Joint | Steakhouse | Breakfa Spot |
| 7 | Westbranch | Mexican Restaurant | Fried Chicken Joint | Food | Burger Joint | Pizza Place | Vietnamese Restaurant | Taco Place | Thai Restaurant | Fast Food Restaurant | Noodle House |
| 9 | Spring Branch West | Mexican Restaurant | Korean Restaurant | Steakhouse | Seafood Restaurant | Mediterranean Restaurant | Fast Food Restaurant | Burger Joint | Pizza Place | Bakery | New Americ Restau |
| 10 | Langwood | Mexican Restaurant | Sandwich Place | Fast Food Restaurant | Pizza Place | Seafood Restaurant | Chinese Restaurant | Fried Chicken Joint | Taco Place | Donut Shop | Food Tr |
| 14 | Greater Heights | Mexican Restaurant | Chinese Restaurant | Fried Chicken Joint | Sandwich Place | Seafood Restaurant | Burger Joint | Donut Shop | Cajun / Creole Restaurant | Deli / Bodega | Bakery |
| 15 | Memorial | American Restaurant | Italian Restaurant | Mexican Restaurant | Pizza Place | Burger Joint | Mediterranean Restaurant | Food Truck | Café | New American Restaurant | Southe Soul Fo Restau |
| 18 | Westchase | Vietnamese Restaurant | Asian Restaurant | Mediterranean Restaurant | Mexican Restaurant | Sandwich Place | Seafood Restaurant | Burger Joint | Pizza Place | Breakfast Spot | Bakery |
| 25 | Sharpstown | Sandwich Place | Pizza Place | Fast Food Restaurant | Mexican Restaurant | Chinese Restaurant | Latin American Restaurant | Donut Shop | Fried Chicken Joint | Food | Vietnan Restau |
| 26 | Gulfton | Burger Joint | Mexican Restaurant | American Restaurant | Pizza Place | Seafood Restaurant | Fast Food Restaurant | Mediterranean Restaurant | Donut Shop | Thai Restaurant | Cajun / Creole Restau |
| 27 | University Place | Mexican Restaurant | Fast Food Restaurant | Pizza Place | American Restaurant | Asian Restaurant | Sandwich Place | Burger Joint | Fried Chicken Joint | BBQ Joint | Restau |
| 28 | Westwood | Mexican Restaurant | Fast Food Restaurant | Fried Chicken Joint | Sandwich Place | Burger Joint | Chinese Restaurant | Donut Shop | Breakfast Spot | American Restaurant | Food Tr |

(this goes on for 10+ more rows)

**Cluster 5**

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 41 | IAH Airport | Food | Wings Joint | Ethiopian Restaurant | Deli / Bodega | Dim Sum Restaurant | Diner | Doner Restaurant | Donut Shop | Dumpling Restaurant | Eastern European Restaurant |

Discussion

Looking at the cluster data, we can observe the neighborhoods that share many similarities with other neighborhoods, and those that are unique.

In particular, Clusters 1, 3, and 5, hold significance to prospective restaurant owners, as they can clue in the type of restaurants needed in particular neighborhoods.

For example, Cluster 1's neighborhoods Alief and Midtown may be great places for African or Afghani restaurants. They would not be as successful for new Vietnamese restaurants, which may face a lot of competition.

Cluster 3 has one sole neighborhood, Downtown Houston, whose restaurants seem to be either predominantly European or Chinese. Any extraneous restaurant cuisine would succeed without much competition.

Cluster 5, only containing the George Bush International Airport (IAH), has a pretty diverse range of restaurants, but even so, there lacks some major cuisines, such as Indian.

As demonstrated above, restaurant owners can infer patterns from this cluster data and decide where in Houston they may want to open a particular restaurant cuisine.

## Conclusion

With data from a wiki page that only gave us the neighborhood names in Houston, we were able to eventually obtain clusters that grouped restaurants in Houston neighborhoods that share similar cuisines. This research proves the effectiveness of Data Science and Machine Learning algorithms such as K Means. These two can be utilized in domains that extend beyond geographical data, such as the healthcare, economical, and technological industries. But for now, hopefully aspiring Houston restaurant owners can benefit from this research, and bounce back stronger after COVID times.
Thank you for reading!