

# COVID-19 spread in Houston- TX/USA

---

CARLOS ALEXANDRE NAKO

# Understanding COVID-19 relationship with venues

---

- As a new disease to which the world has never seen such reach and lethality in 100 years, no government had a clear view on how to respond to the pandemic.
- To have effective health measures, policy makers need to have a good understanding of how customer facing business, either retail, restaurants, bars and others, relates with COVID-19 spread.
- The **objective of this report** is to use available COVID cases in Houston and Foursquare business location data to understand what is the relation between COVID-19 cases across Houston and the number of businesses in each region so that policy makers can take this to create effective measures to curb COVID-19 spread.

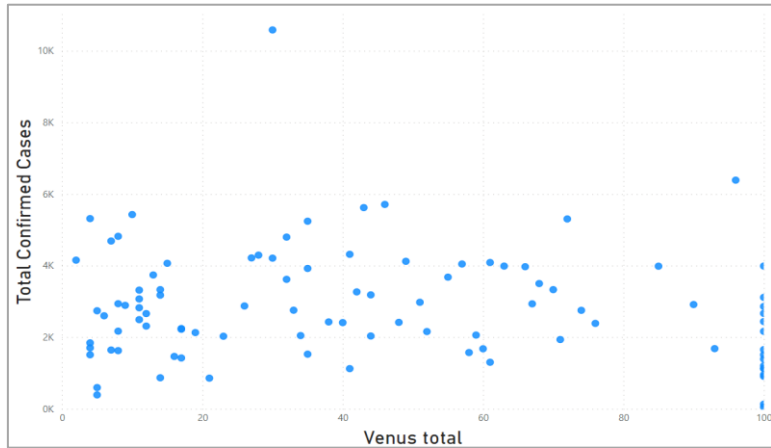
# Data Sources and Cleaning

---

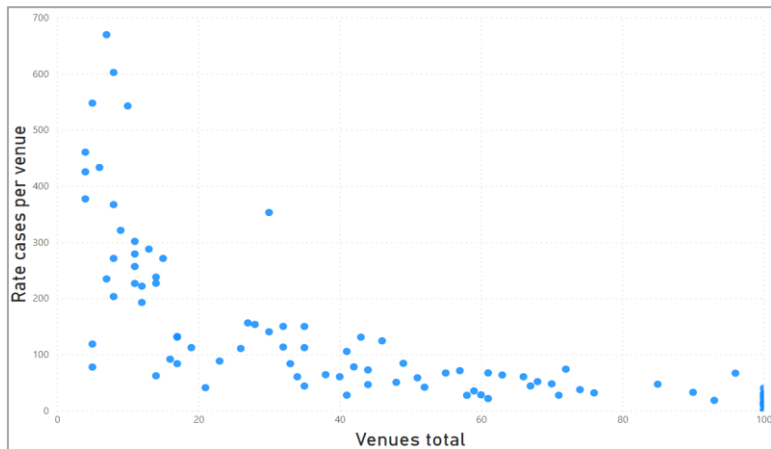
- Number of COVID-19 cases by ZIP code in Houston from 'City of Houston COVID-19 Data Hub' (Harris County Public Health, 2021). This data is gathered by Houston Health Department and is available to the general public.
- ZIP code location in Houston sourced from ZipAtlas (ZipAtlas, 2020) which is a company specialized in Zip code for US. The data from ZipAtlas (ZipAtlas, 2020) was extracted directly from the website and once it was extracted there was no extra cleaning up done before the data was ready for the analysis.
- Commerce data from Foursquare (Foursquare, 2021). To do that, ZIP location data are used to retrieve the Foursquare data. After retrieving business data for each ZIP code, they are combined into one single table.

# Number of venues and COVID-19 cases relationship

Cross-plot between  
Venues against Cases



Cross-plot between  
Venues against Rate  
of cases per venue

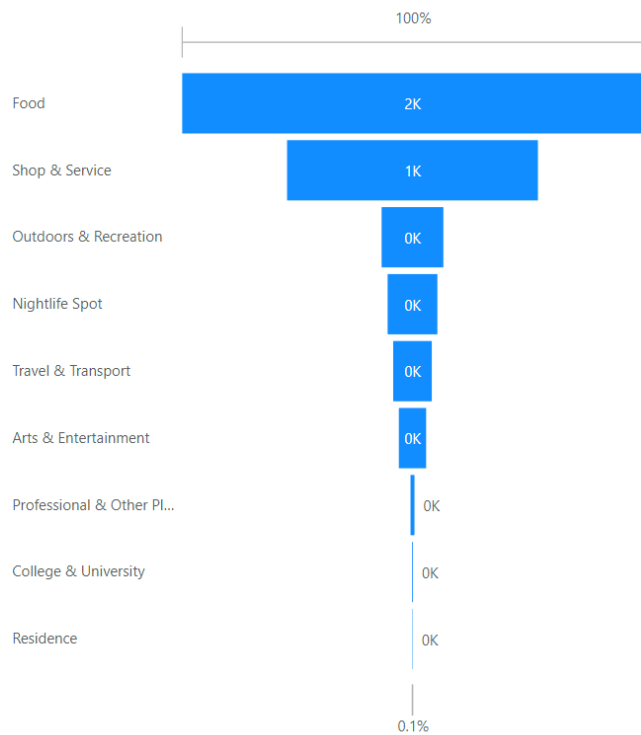


Two possible charts to verify the relationship between number of venues and cases are the number of venues against the number of cases and the number of venues against the ratio between number of cases and venues.

- The venues against cases plot doesn't clearly show a relationship between them.
- The cross-plot between venues against ratio of cases by venue clearly shows that the ratio does not remain constant when venues change as it would be in a linear relationship.

# Venue drilldown

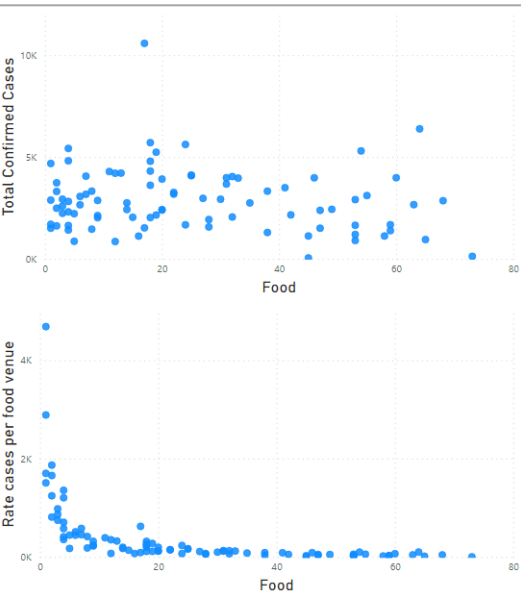
---



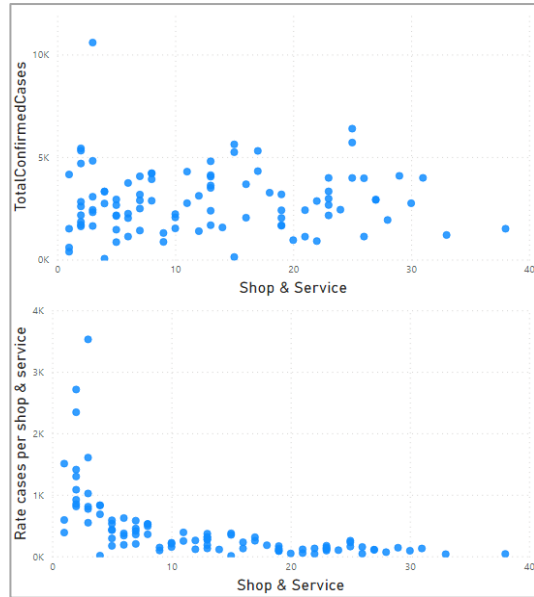
When the venues are grouped at the parent category it is possible to rank then and start the analysis with the relevant categories.

The top 3 categories, Food, Shop & Service and Outdoors & Recreation

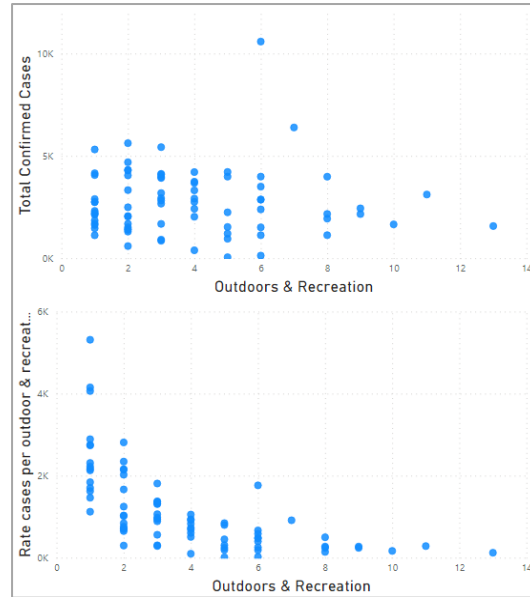
# Venue drilldown and COVID-19 cases relationship



Cross-plots with Food category



Cross-plots with Shop & Service category

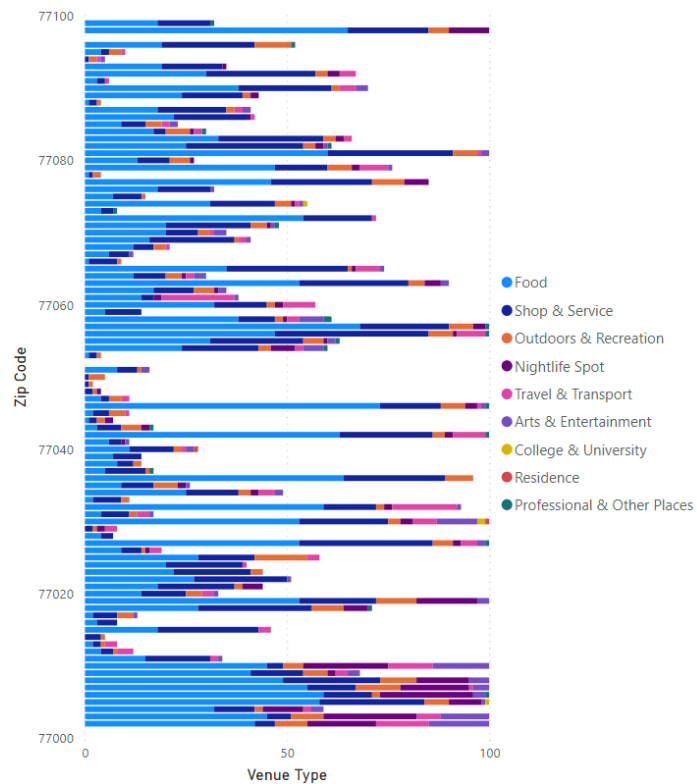


Cross-plots with Outdoors & Recreation category

- As with the total venues, the drilled down venue categories don't show a clear relationship between the number of venues and the number of COVID-19 cases.
- The cross-plot shows that it is not possible to show a clear relationship between the number of COVID-19 cases and the number of venues of a given Zip Code.

# Further drilldown

---



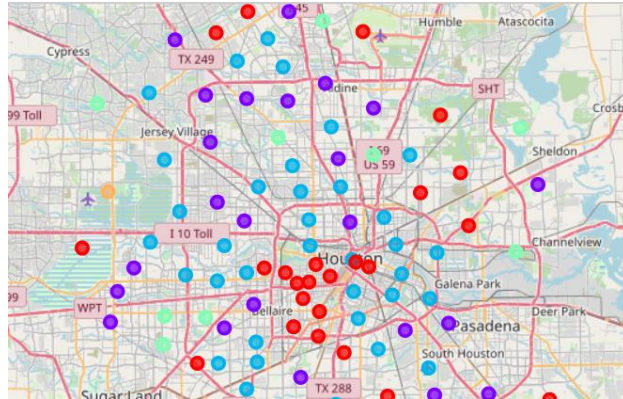
- Drill down even further reduces the number of the data points as not every Zip Code has every type of venues.
- Even at the parent category level, the number of neighborhoods with each category reduces as we move to the categories with less venues overall.



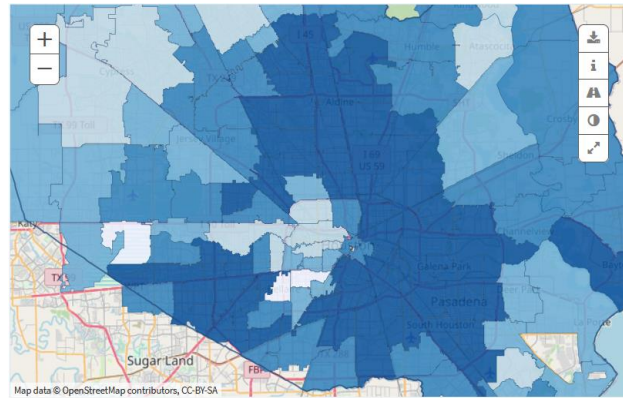
# COVID-19 clustering and other possible sources

---

Houston map  
clustered by  
number of COVID  
cases



Houston Median  
Household Income  
by Zip Code



The data exploration showed that there's no clear relationship between COVID-19 cases and number of venues by zip code.

- It is possible to cluster the neighborhoods by the number of COVID-19 cases and number of venues which allows to visualize how close together the similar clusters are.
- It is recommended to add other parameters that might not be straight away obvious or related to a virus spread such as household income. This can be found in places such as Houston state of health (Houston State of Health, 2021).



# Conclusions and Further Directions

---

- At this point it is not possible to create a final policy based on the number of COVID-19 cases and number of venues by zip code alone.
- This study showed that the COVID-19 spread is not a simple problem and that public authorities should devote resources to properly understand it to promote effective public policies.
- It is clear that using number of venues as the sole parameter to define how to define policies against COVID-19 is not a wise decision.
- In order to better understand the relationships between the disease factors contributing to its spread a larger study should be performed, which is beyond the scope of this project.

# References

---

Foursquare. (2021, 06). Places API. Retrieved from Foursquare Developers:

[https://api.foursquare.com/v2/venues/explore?client\\_id={} & client\\_secret={} & ll={},{} & v={} & radius={} & limit={}](https://api.foursquare.com/v2/venues/explore?client_id={} & client_secret={} & ll={},{} & v={} & radius={} & limit={})

Foursquare. (2021, 06 10). Venue Categories. Retrieved from Foursquare: <https://docs.foursquare.com/docs/legacy-venue-categories>

Harris County Public Health. (2021, 06 16). Download COVID Cases By Zip Codes. Retrieved from Harris County / City of Houston COVID-19 Data Hub: [https://covid-harriscounty.hub.arcgis.com/datasets/081f346d4b9b43d88fcdc79e2e6c4517\\_0/explore?location=29.833797%2C-95.435187%2C10.33](https://covid-harriscounty.hub.arcgis.com/datasets/081f346d4b9b43d88fcdc79e2e6c4517_0/explore?location=29.833797%2C-95.435187%2C10.33)

Houston State of Health. (2021). Median Household Income. Retrieved from

<http://www.houstonstateofhealth.com/indicators/index/view?indicatorId=315&localeTypeId=3&periodId=4523&>

IBM. (2021). IBM Watson Studio. Retrieved from IBM Data Platform: <https://dataplatform.cloud.ibm.com>

ZipAtlas. (2020). Population Density in Houston, TX by Zip Code. Retrieved from ZipAtlas:

<http://zipatlas.com/us/tx/houston/zip-code-comparison/population-density.htm>