

COVID-19 Community Risk Assessment

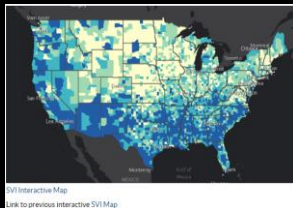


By Jacky Chan, Li (Betty) Gai

Data Sources

- 3 Main Public Data Source
- 12,000+ Variables Reviewed
- 300+ Variables Pulled for Analysis
- 15+ high correlation variables used for prediction

CDC Social Vulnerability Index



CDC 500 Cities



Census Bureau



Motivation

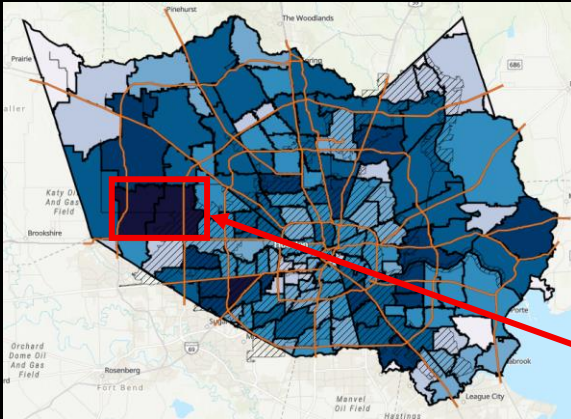
The COVID-19 pandemic has been impacting the United States since the beginning of 2020, the number of confirmed cases is still growing every day due to the public's lack of awareness of the severity of this disease.

Although confirmed cases are tracked via different zip codes, the more granular information is lacking in terms of which community has the most COVID cases and which part of the area could be considered a “**high risk area.**” Residents often need the information on the census tract level to guide their behaviors and further flatten the curve.

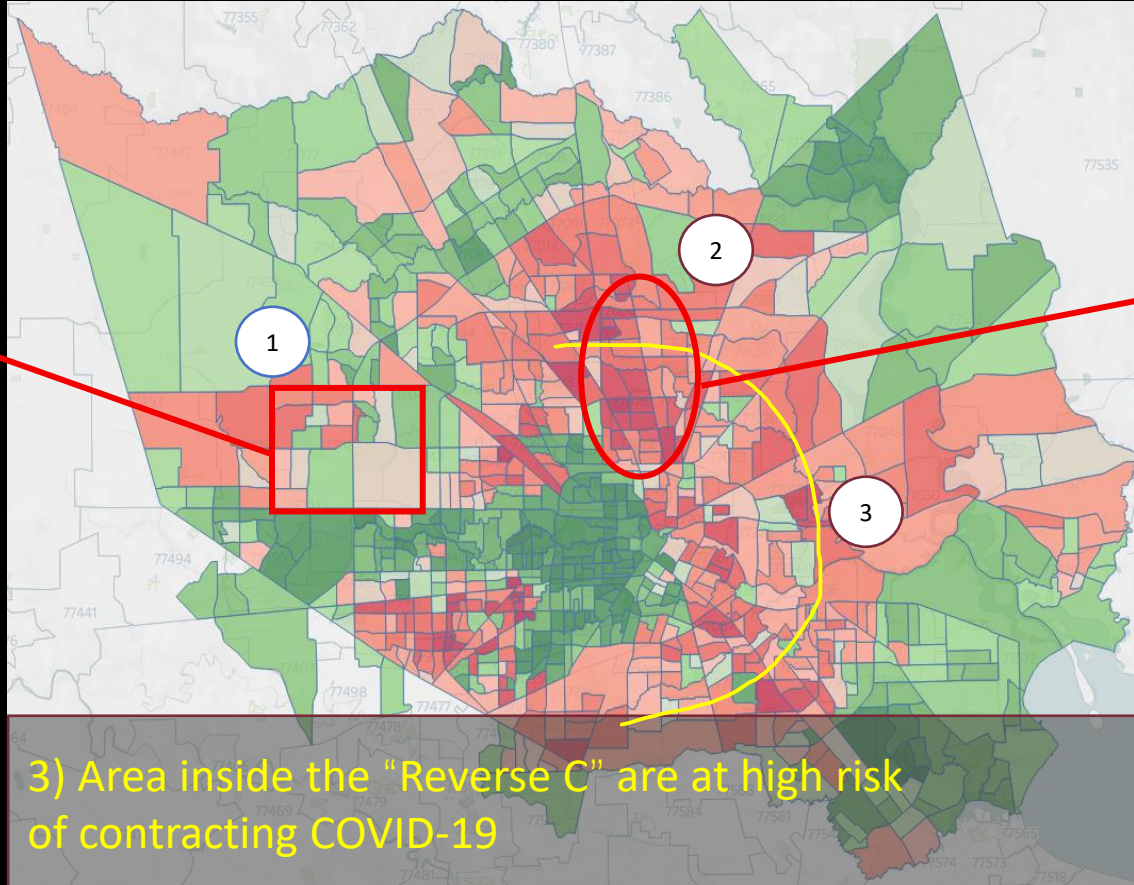
Methodology - OSEMN

1. **Obtain** census tract level data from 3 main public data sources + zip code level confirmed cases for Harris county
2. **Scrub** and Integrate the data using Alteryx and python
3. **Explore** the high correlation variables and visualization using python
4. **Modeling**: County percentile to calculate the risk score (Tract area with number lower than the county percentile would be in low risk, and Tract are with number higher than the county percentile would be in high risk) and predict against derived COVID-19 infection rate.
5. **Interpret** the community risk map developed using Tableau

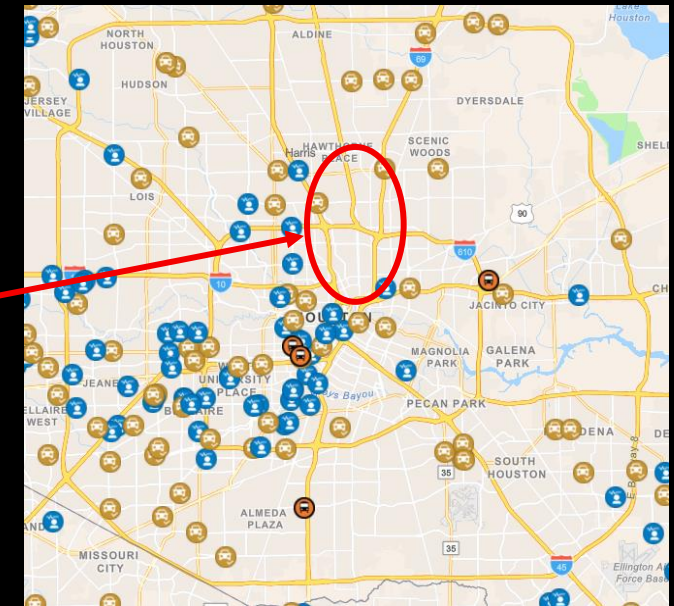
INSIGHTS AND FINDINGS



1) Confirmed case count for a zip code may not be the best metrics to show the "risk" of getting COVID-19
<https://publichealth.harriscountytexas.gov/resources/2019-novel-coronavirus>



3) Area inside the "Reverse C" are at high risk of contracting COVID-19



2) Lack of testing center in high risk communities
<https://tdem.maps.arcgis.com/apps/webappviewer/index.html?id=1e91fb79fa44417898738e5bff31a3d8/>

HIGH IMPACT VARIABLES

Abstract confirmed cases from Harris county COVID-19 dashboard

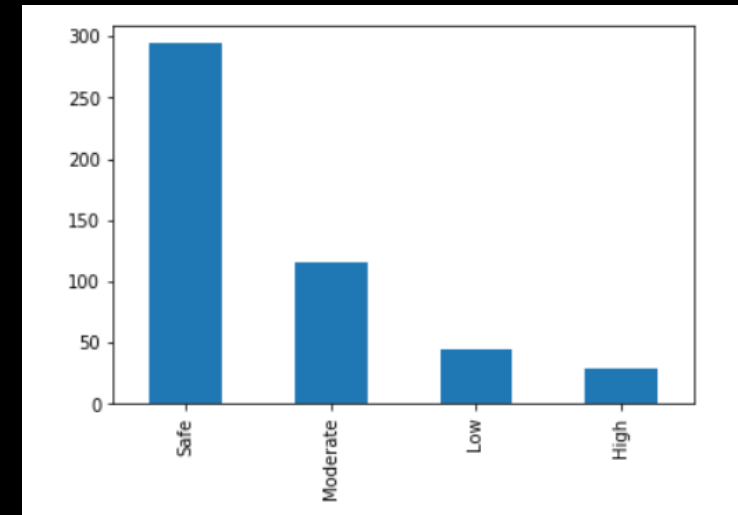
Variables	Type
Current smoking	health
No leisure time physical activity	health
Mental health not good for more than 14 days	health
Persons below poverty	socioeconomic
Percentage of single parent households with children under 18	household composition
Percentage of persons aged 17 and younger	household composition
Minority all persons except white non-Hispanic	minority status
Persons age 25 with no high school diploma	socioeconomic
Current lack of health insurance	socioeconomic
At household level occupied housing units more people than rooms	household composition
Uninsured in the total civilian noninstitutionalized population	household composition
Persons age 5 who speak English less than well	minority status
Mammography 500	health
Papanicolaou smear use 500	health
Cholesterol screening 500	health

```

IF [cases per 10K population by zipcode]<=10
THEN "Safe"
ELSEIF [cases per 10K population by zipcode] >10 AND [cases per 10K population by zipcode]<=200
THEN "Low"
ELSEIF [cases per 10K population by zipcode] >200 AND [cases per 10K population by zipcode]<=400
THEN "Moderate"
ELSEIF isnull([cases per 10K population by zipcode])
THEN "Low"
ELSE
"High"
ENDIF
    
```



	by 50 P..	County 90 P..	Risk Score
	2,399.5		-4
	258		-2
	7,502		-3
	47.4		-3
	24.3		-2
	18.09		-3
	42		-3
	18		-3
	33.5		-5
	21.4		-5
	1,583.5		-3
	1,473.5		-3



Census tract count by risk category

	Persons below poverty estimate, 2014-2018 ACS (SVI)	629	156.5	798.5	1,937	-1
48201210100	Adjunct variable - Uninsured in the total civilian noninstitutionalized population estimate, 20...	0	191.5	988	2,399.5	-5
	At household level (occupied housing units), more people than rooms estimate, 2014-2018 AC...	5	10	86	258	-5
	Minority (all persons except white, non- Hispanic) estimate, 2014-2018 ACS (SVI)	5,698	1,179.5	3,594.5	7,502	3
	Model-based estimate for crude prevalence of current lack of health insurance among adults ..	33	13.01	30.05	47.4	1
	Model-based estimate for crude prevalence of current smoking among adults aged >=18 year..	22.1	9.9	17.9	24.3	3
	Model-based estimate for crude prevalence of mental health not good for >=14 days among a...	13.3	8.9	14.1	18.09	-1
	Model-based estimate for crude prevalence of no leisure-time physical activity among adults ..	25.9	18.6	31.95	42	-2
	Model-based estimate for crude prevalence of physical health not good for >=14 days among ..	8.5	7.6	12.75	18	-4
	Percentage of persons aged 17 and younger estimate, 2014-2018 ACS (SVI)	0.6	17.2	26.7	33.5	-5
	Percentage of single parent households with children under 18 estimate, 2014-2018 ACS (SVI)	9.6	3.05	10.8	21.4	-1
	Persons (age 5+) who speak English "less than well" estimate, 2014-2018 ACS (SVI)	341	38	411	1,583.5	-1
	Persons (age 25+) with no high school diploma estimate, 2014-2018 ACS (SVI)	1,851	81	580	1,473.5	5
	Persons below poverty estimate, 2014-2018 ACS (SVI)	0	156.5	798.5	1,937	-5
48201210400	Adjunct variable - Uninsured in the total civilian noninstitutionalized population estimate, 20...	1,079	191.5	988	2,399.5	1
	At household level (occupied housing units), more people than rooms estimate, 2014-2018 AC...	157	10	86	258	3
	Minority (all persons except white, non- Hispanic) estimate, 2014-2018 ACS (SVI)	4,118	1,179.5	3,594.5	7,502	1
	Model-based estimate for crude prevalence of current lack of health insurance among adults ..	46	13.01	30.05	47.4	4
	Model-based estimate for crude prevalence of current smoking among adults aged >=18 year..	22.5	9.9	17.9	24.3	3
	Model-based estimate for crude prevalence of mental health not good for >=14 days among a...	17.5	8.9	14.1	18.09	4
	Model-based estimate for crude prevalence of no leisure-time physical activity among adults ..	42.4	18.6	31.95	42	5
	Model-based estimate for crude prevalence of physical health not good for >=14 days among ..	18.6	7.6	12.75	18	5
	Percentage of persons aged 17 and younger estimate, 2014-2018 ACS (SVI)	29.6	17.2	26.7	33.5	2
	Percentage of single parent households with children under 18 estimate, 2014-2018 ACS (SVI)	17.3	3.05	10.8	21.4	4
	Persons (age 5+) who speak English "less than well" estimate, 2014-2018 ACS (SVI)	809	38	411	1,583.5	2

Next Steps

- Incorporate data from more SDoH domains (e.g. workforce data, eviction data, criminal justice)
- Overlay risk map with community resource data (e.g. food assistance, housing, financial support, employment, health care access, transportation) and come up with COVID specific suggestions
- Refine the list of high impact variables selected and deep dive into the underlying data for high risk area in the community
- Refine predictors for risk classification model
- Expand this analysis to TX and national level
- Get feedback from and collaborate with the domain specific experts on how to improve on this work and share the knowledge and benefit the entire community!

Classification Methods

Initial Model Output

	SVM	DT	RDF	NN
K1	0.640000	0.640000	0.680000	0.640000
K2	0.600000	0.400000	0.580000	0.660000
K3	0.600000	0.680000	0.720000	0.780000
K4	0.600000	0.620000	0.660000	0.620000
K5	0.604167	0.666667	0.708333	0.687500
K6	0.604167	0.520833	0.604167	0.666667
K7	0.617021	0.531915	0.574468	0.489362
K8	0.617021	0.489362	0.553191	0.638298
K9	0.617021	0.574468	0.617021	0.638298
K10	0.630435	0.565217	0.608696	0.630435
mean	0.612983	0.568846	0.630588	0.645056