

# Utilizing Yelp Data to Estimate the Number of Businesses in Houston

Joe Darby  
Alex Harmon  
Sydney Herndon  
Zoe Nawar



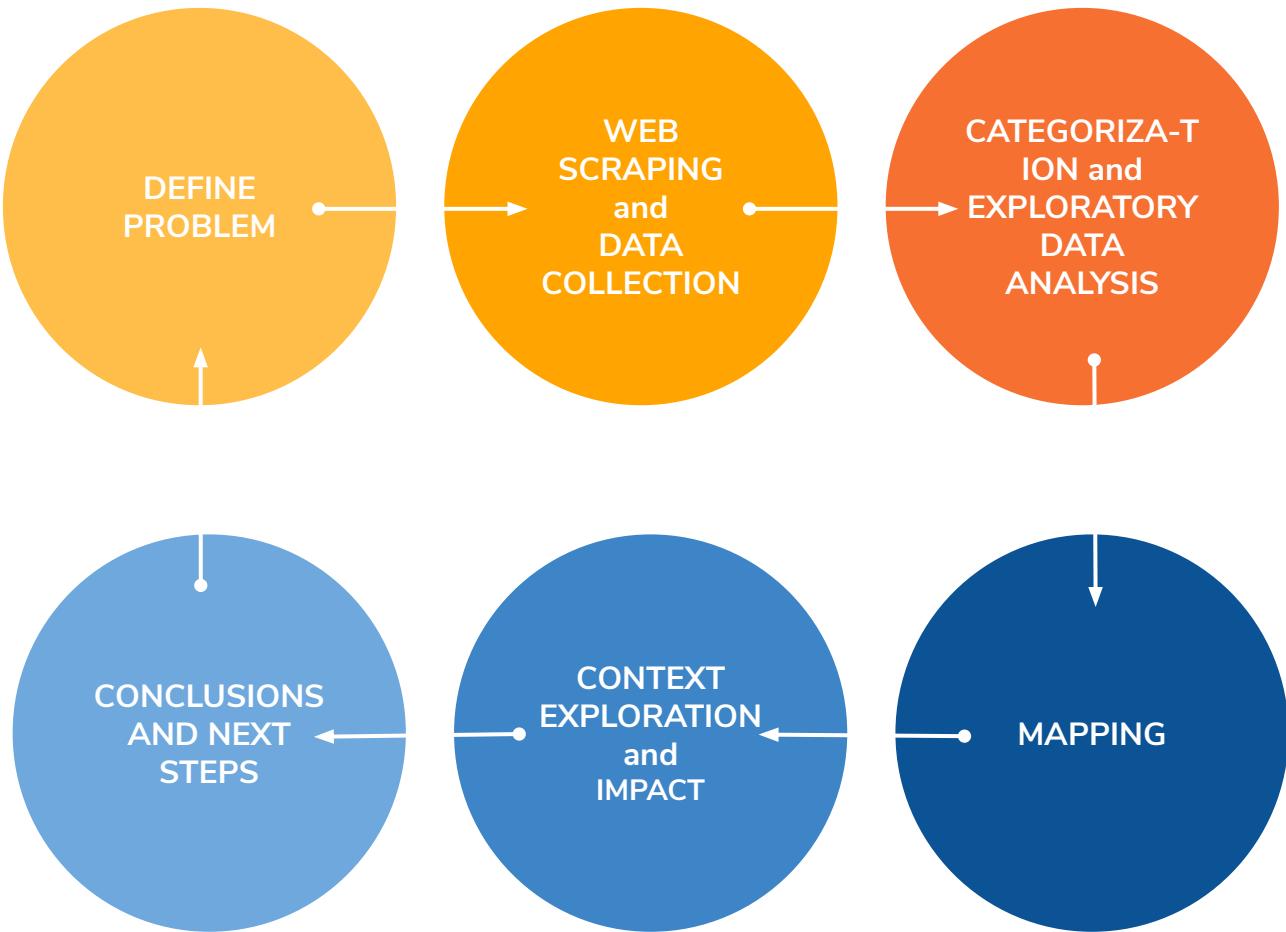
# Defining the Problem and Stakeholders:

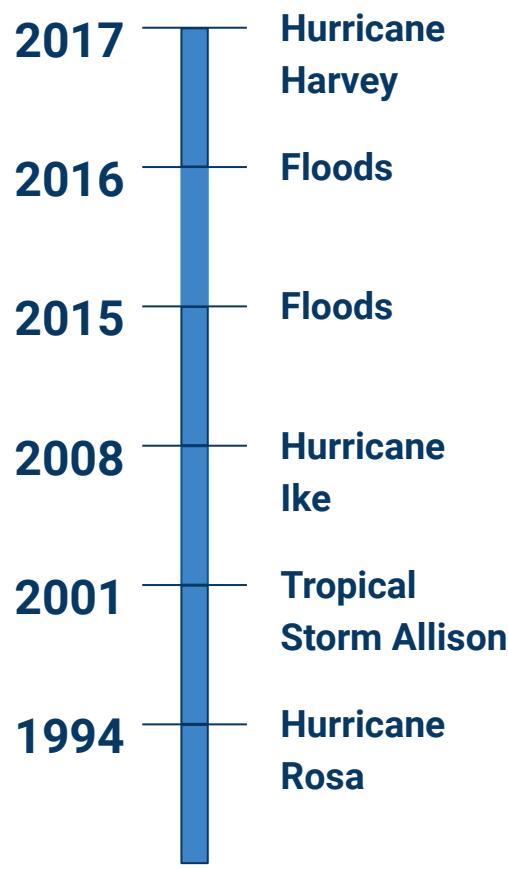
**Problem Statement:** Prior to and during a disaster, it is important to understand the projected and actual effects of the event on the community, including its economic effects on critical services. Our clients need a tool to estimate the potential effects of an event on each of FEMA's seven lifelines.



FEMA

# OUR PROCESSES:





# Timeline of Major Natural Disasters in Houston





Find grocery stores

Near Houston



🍴 Restaurants ▾

🏡 Home Services ▾

🚗 Auto Services ▾

More ▾

✍ Write a Review

🏪 For Businesses

# Scraping Yelp Data

## DATA POINTS:

- Business Name
- Business Category
- Address
- Zip Code
- Latitude
- Longitude
- Phone Number
- Price
- Rating
- Is\_closed

## LIMITATIONS:

- Only 1000 businesses can be scraped per category parameter
- Limited data (number of employees, business revenue, etc.)
- Majority of price values missing making it impossible to predict potential economic impact
- Lack of alignment between Yelp categories and FEMA Lifelines

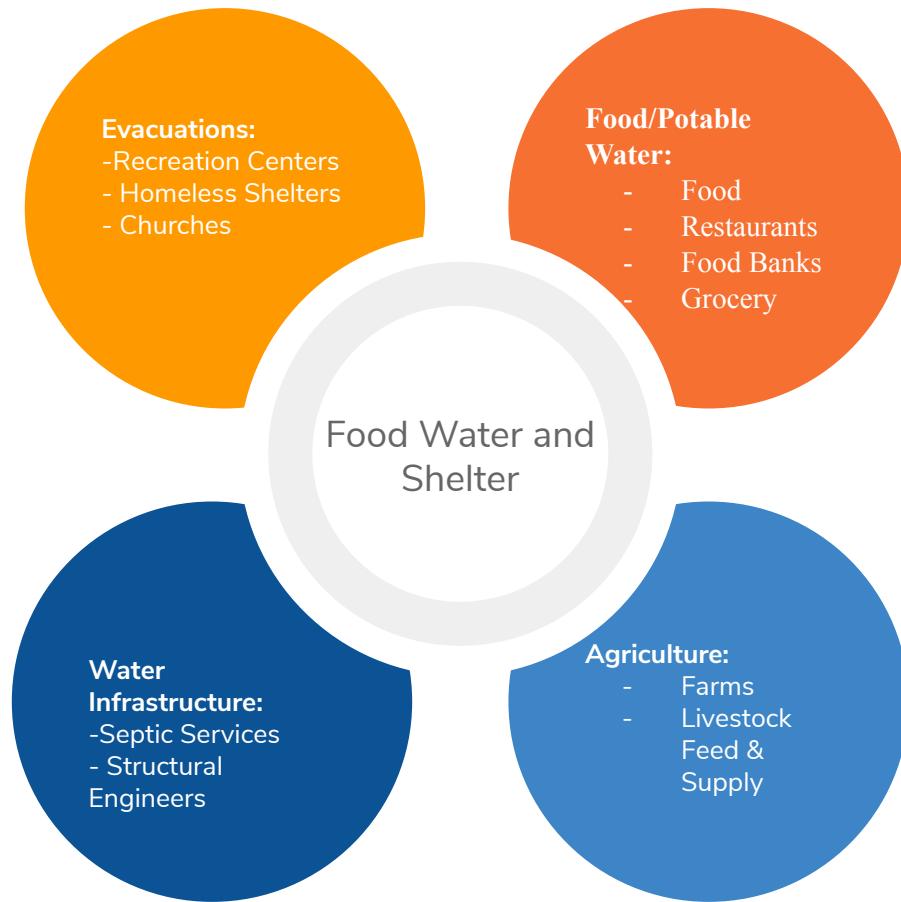
# FEMA LIFELINES:



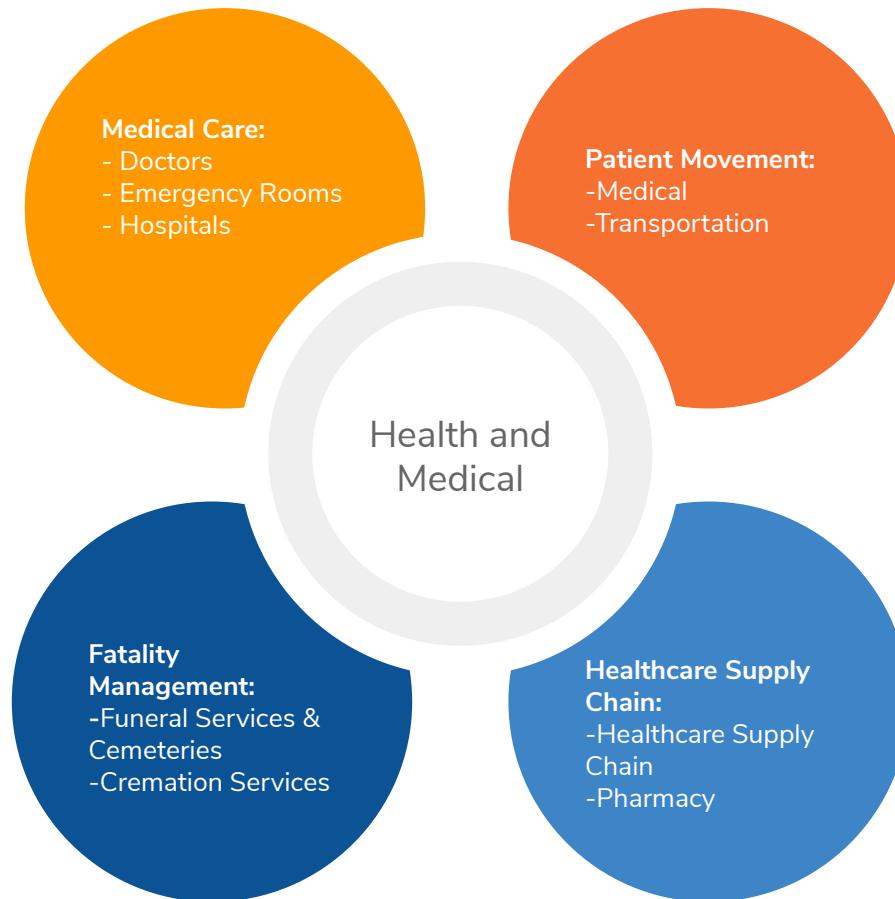
# Aligning Yelp Categories with FEMA LifeLines



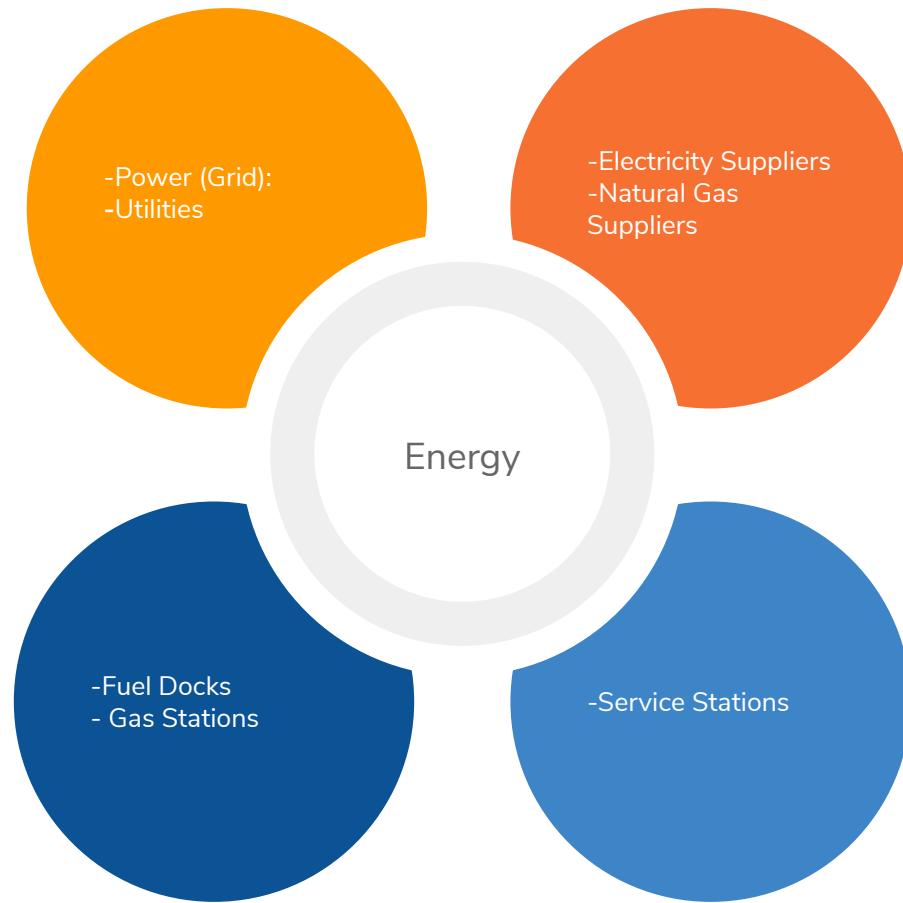
# Aligning Yelp Categories with FEMA LifeLines



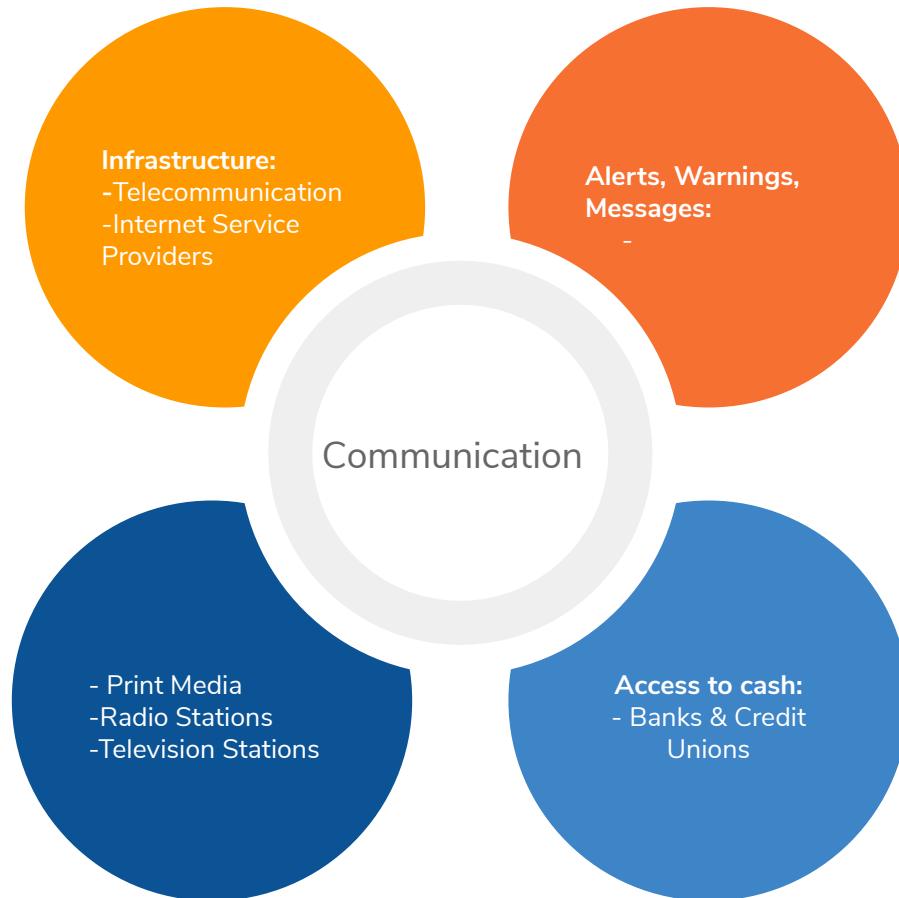
# Aligning Yelp Categories with FEMA LifeLines



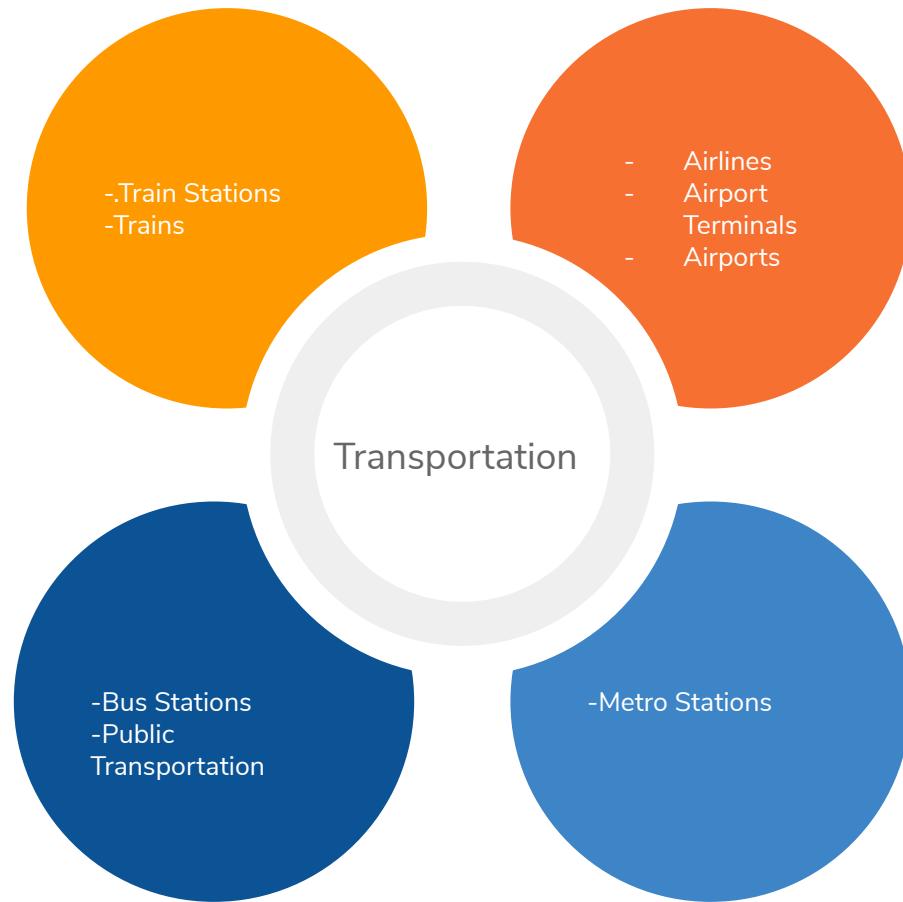
# Aligning Yelp Categories with FEMA LifeLines



# Aligning Yelp Categories with FEMA LifeLines



# Aligning Yelp Categories with FEMA LifeLines



# Aligning Yelp Categories with FEMA LifeLines



# Exploratory Data Analysis





# Houston by Zip Code

## Data Dictionary

A brief overview of all the columns, what they are, and their origins. **Bolded** for emphasis.

Column	Description	Source
<b>zip_code</b>	<b>Zip code</b>	<b>Yelp, zillow, splitwise</b>
num_yelp_bus	Number of businesses in zip code	Yelp Scrape
med_hh_income	Median household income	Kaggle Dataset
<b>est_num_damaged_bus</b>	<b>Estimated number of damaged businesses by zip code</b>	<b>Dun &amp; Bradstreet</b>
<b>2010_census_pop</b>	<b>Population per zip code</b>	<b>2010 census dataset</b>
state	State	Zillow Dataset
metro	Metro area	Zillow Dataset
county	County name	Zillow Dataset
city	City name	Zillow Dataset
<b>ZHVI</b>	<b>Zillow Housing Value index at time of data collection</b>	<b>Zillow Dataset</b>
YoY	Year Over Year zhvi	Zillow Dataset
5Year	5 year prediction	Zillow Dataset
10Year	10 year prediction	Zillow Dataset
PeakZHVI	Peak zhvi for zip code	Zillow Dataset



Mean Income  
across Zip Codes



Mean Population across Zip  
Codes



Average count of  
damaged businesses  
among zip codes after  
Harvey



Avg Zillow Housing Value  
Index among Zip codes

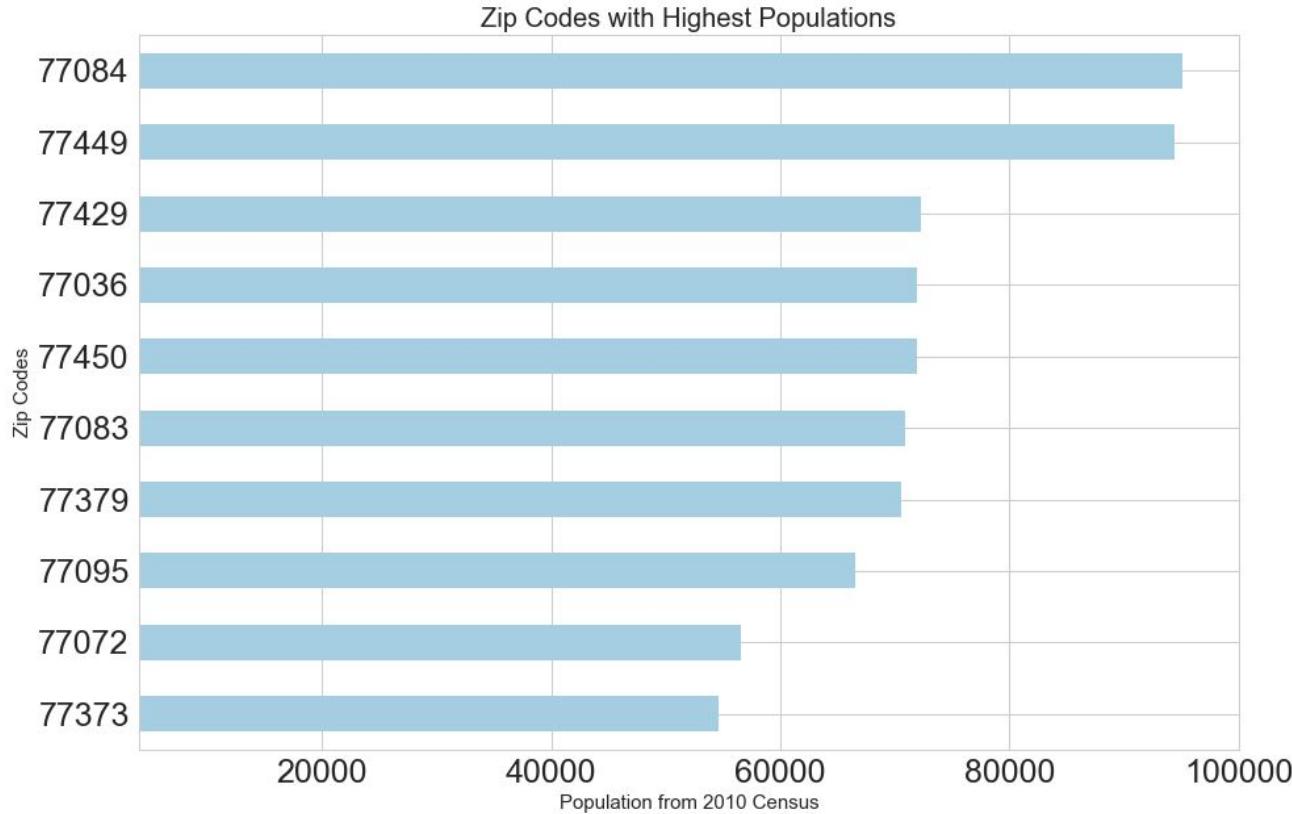
\$ 62, 633

31,916

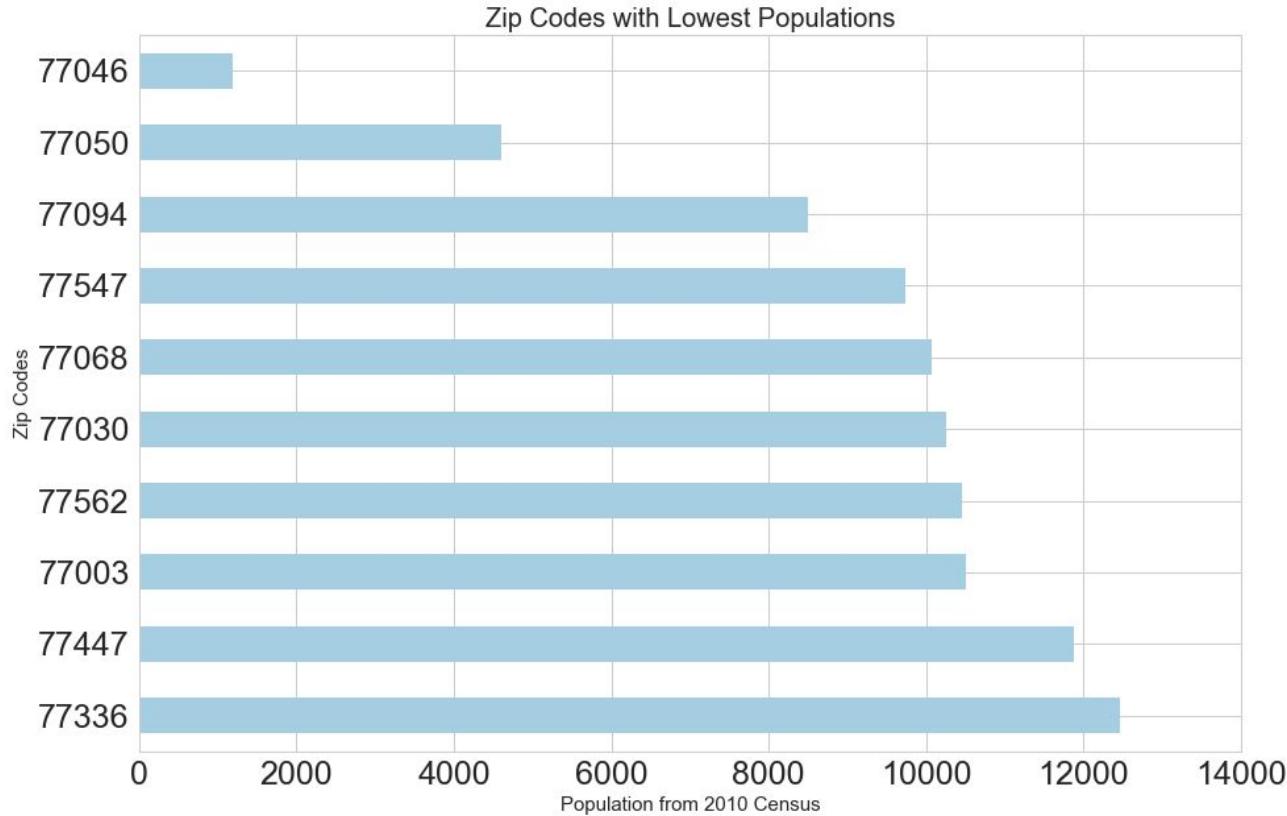
2,462

\$ 237, 129

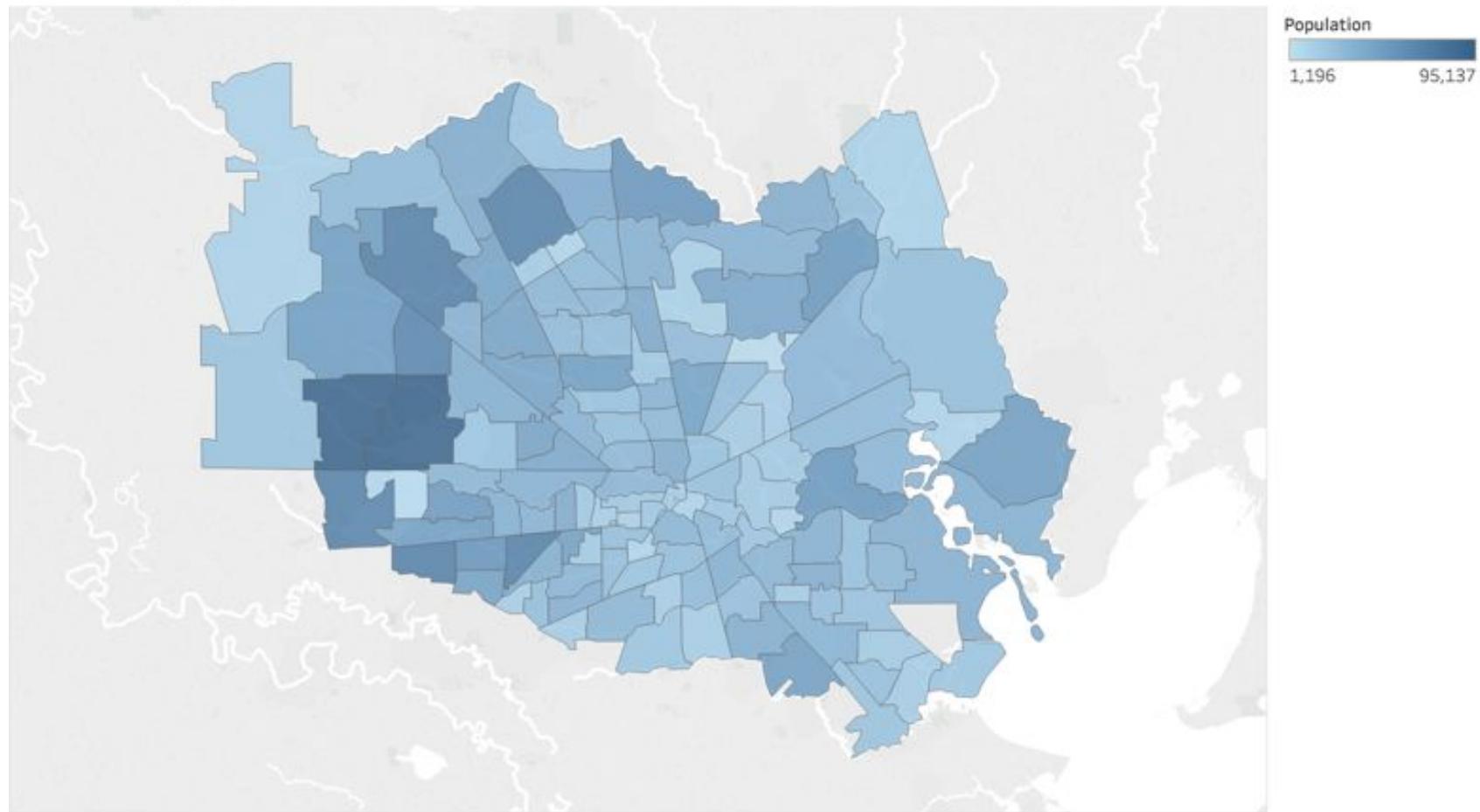
# Most Populated Zip Codes



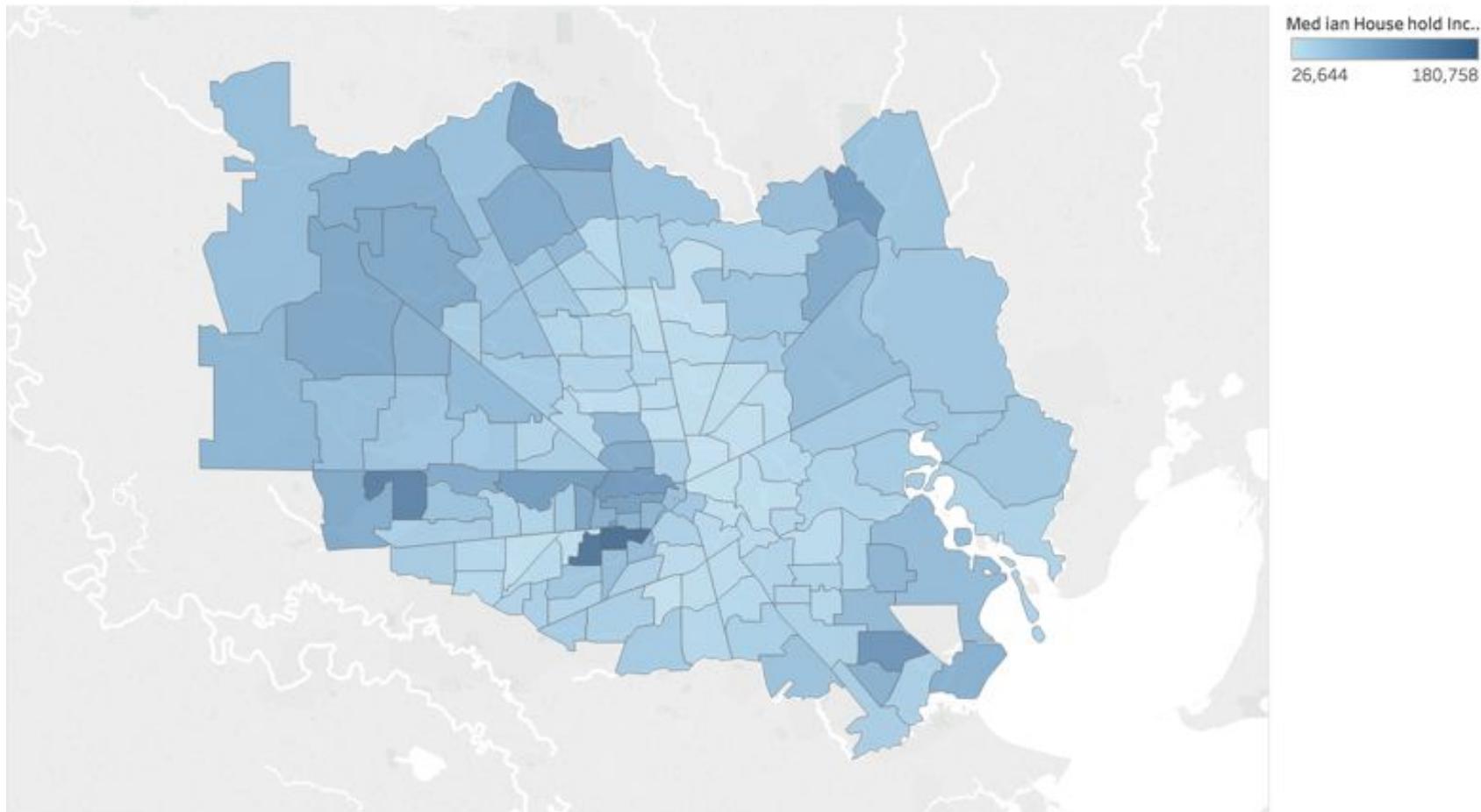
# Least Populated Zip Codes



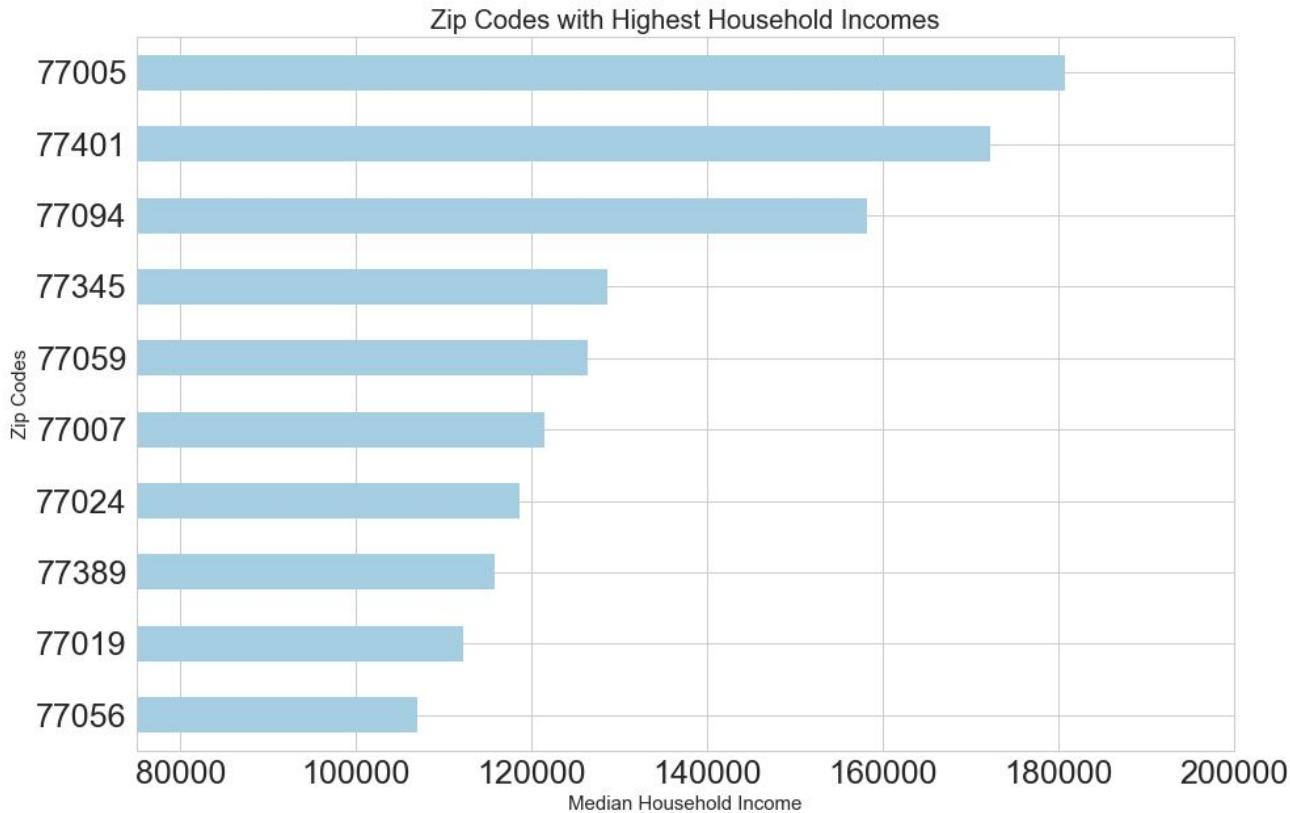
Population by Zipcode



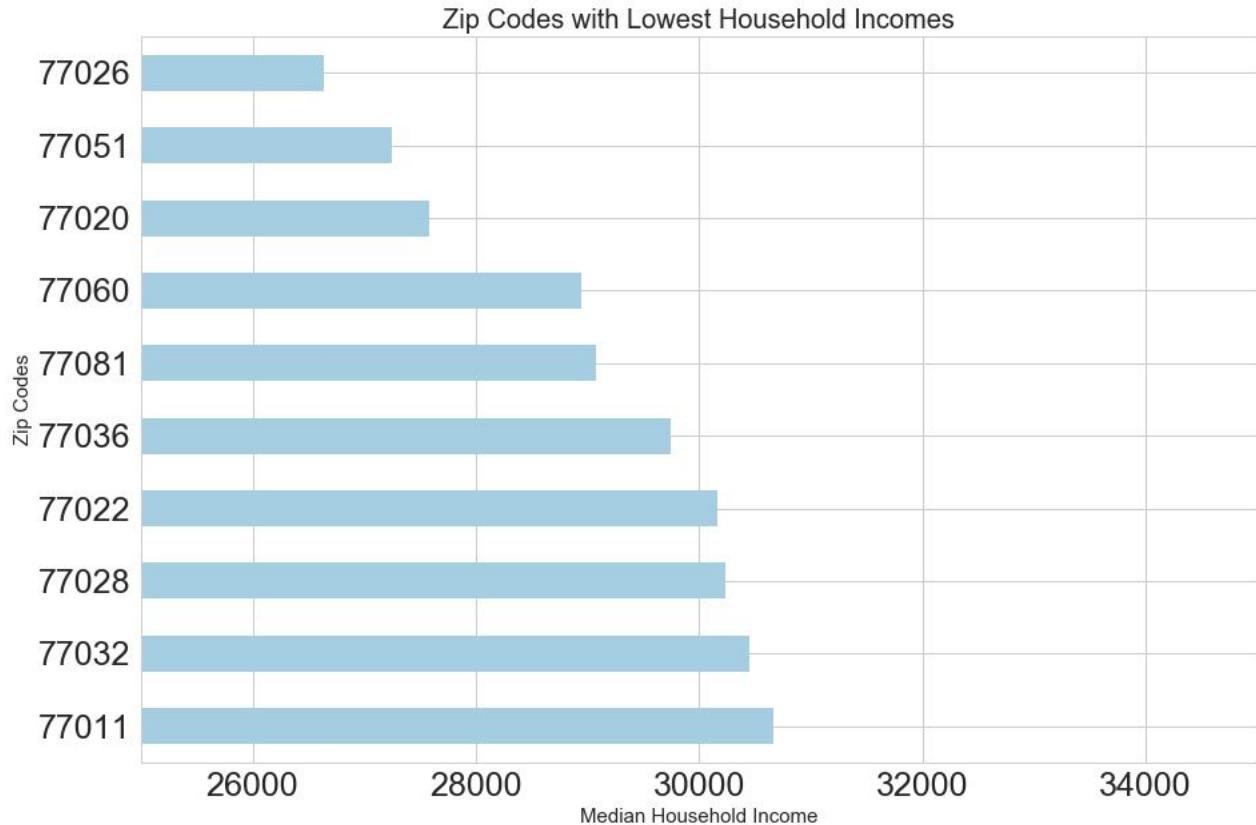
## Income by Zipcode



# Houston's Richest Zip Codes:



# Houston's Poorest Zip Codes:



A photograph showing a row of houses with different architectural styles and colors (blue, yellow, red) under a clear sky. Some trees are visible in the lower-left foreground.

# Zillow Housing Value Index

## Methodology

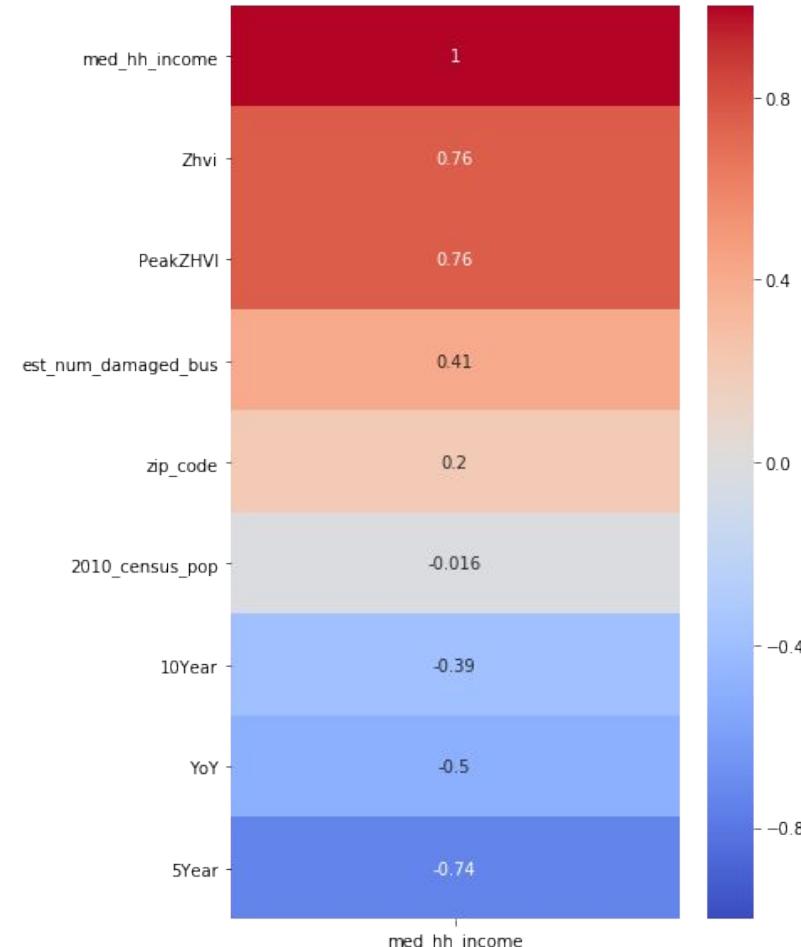
Using the estimated market value of every home as represented in the Zestimate, the main steps in the construction of the ZHVI are as follows:

1. Calculate Raw Median Zestimates
2. Adjust for Any Residual Systematic Error
3. Apply Henderson Moving Average Filter
4. Apply Seasonal Adjustment
5. Final Quality Control

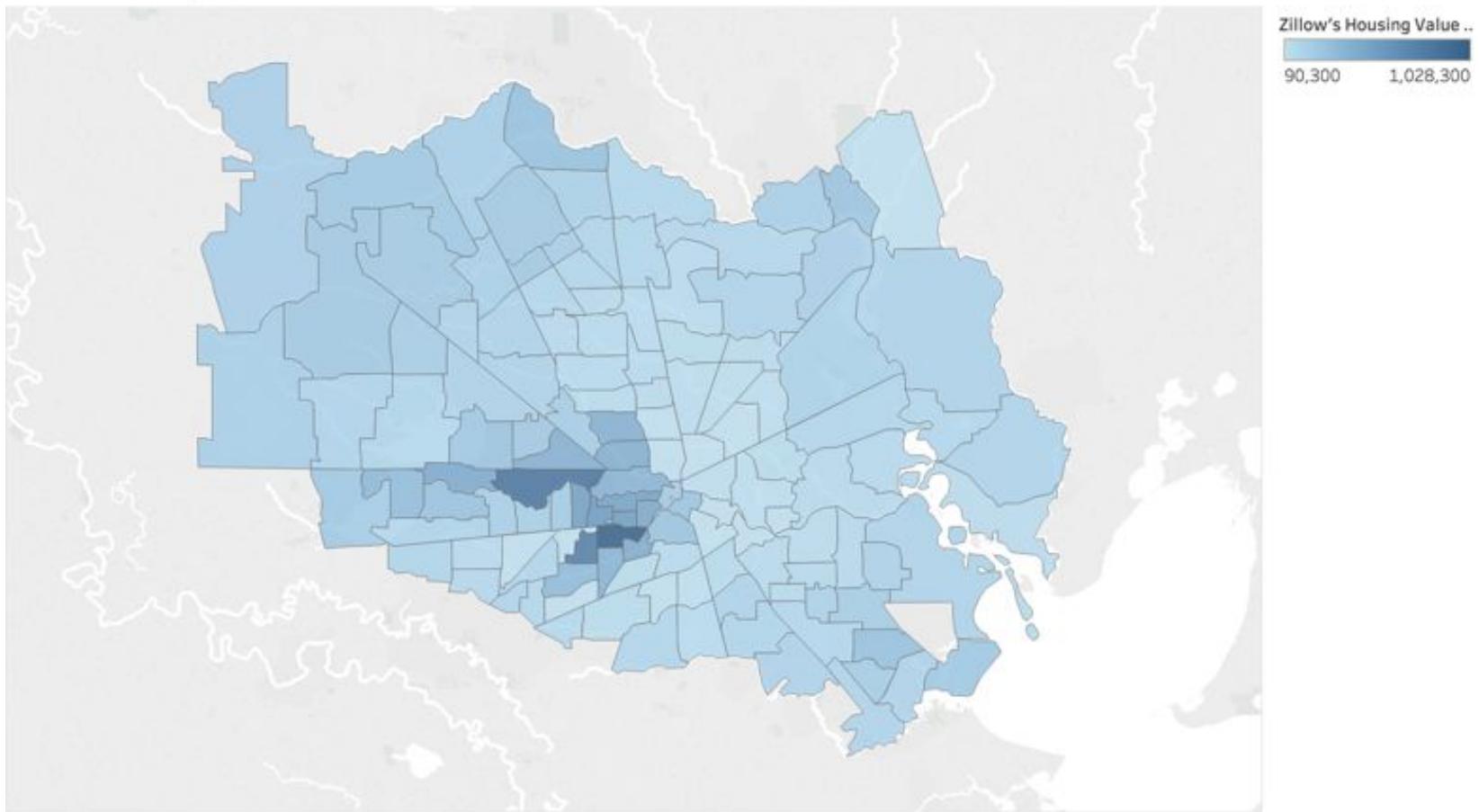
- Zillow's unique housing value index
- Used for time series
- Available in Zillow's public datasets

# Household Income Correlations

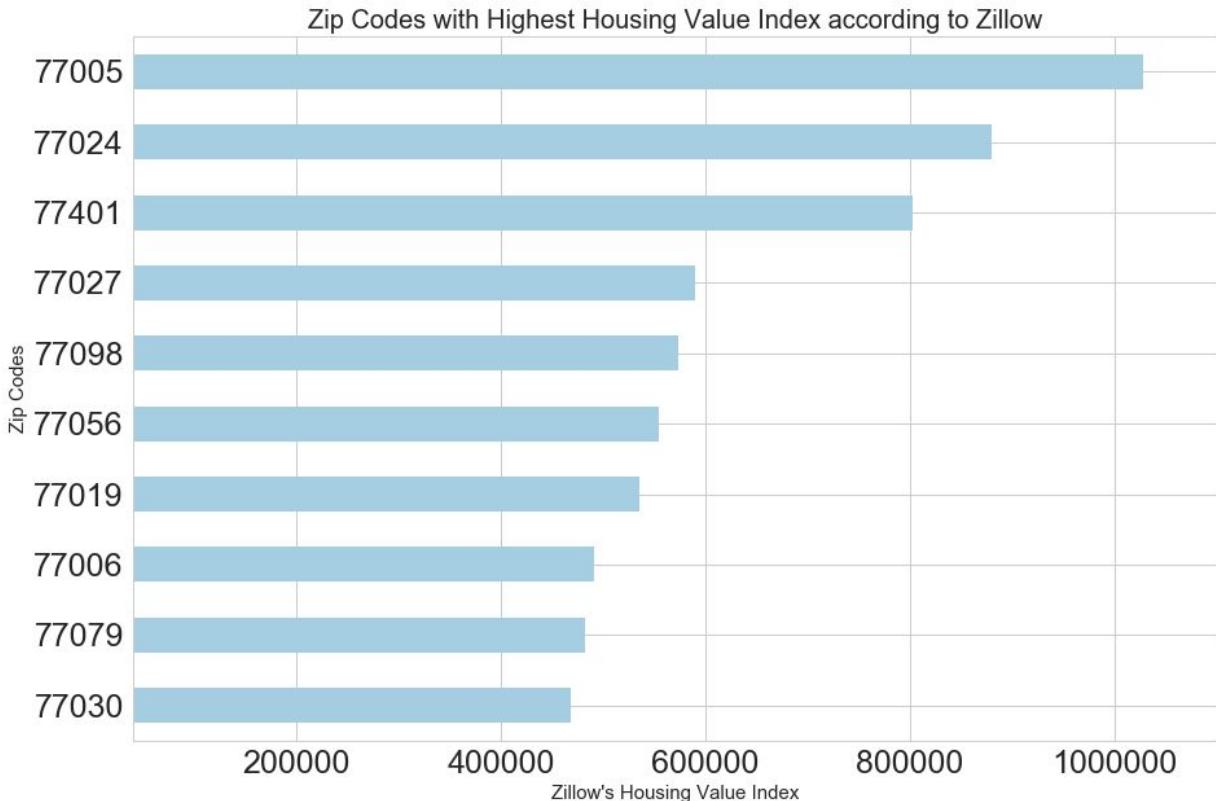
Correlations for median household income



## Zillow Housing Value Index



# Houston's Most Expensive Zip Codes:



# Houston's Most Expensive Zip Codes:

Zip	Information	Median Household Income	Estimated Number of Damaged Businesses
77005	Primarily white, wealthy, with good schools and low unemployment	180758	3561
77024	Extremely wealthy, older	118561	7542
77401	Primarily "Bellaire" - major housing market	172205	3213
77027	Little further from the nicest parts of the city	91296	5118
77098	Upper middle class	87820	3493
77056	Diverse, mostly families	107003	7334
77019	Primarily white, wealthy	112243	3281
77006	Diverse, upper middle class	82878	3050
77079	White, upper middle class	98774	4569
77030	Significant asian student population, upper middle class	75290	2160

1. The wealthiest neighborhoods are primarily a mix of retirement communities and families, meaning that even though they're well funded they might need specified attention during disasters and evacuations.
2. Rich neighborhoods do still have a significant portion of damaged businesses. Possibly due to reporting methods.

# Houston's Cheapest Zip Codes:



# Houston's Cheapest Zip Codes:

Zip	Information	Median Household Income	Estimated Number of Damaged Businesses
77028	Poor, mostly african-american	30233	978
77033	Poor, significant hispanic population	33750	866
77029	Primarily hispanic, lower-middle class, lots of families	32200	1473
77016	Awkwardly shaped, diverse area	32302	1067
77078	Primarily hispanic, lower-middle class	35770	418
77020	Primarily hispanic, lower	27587	1210
77093	Primarily hispanic, lower middle-class	30837	1784
77506	Diverse, lower-middle class	34838	1138
77050	Diverse, young, middle class	57163	200
77520	Diverse, middle class	46311	1775

1. The neighborhoods Zillow finds to be low on the housing index are primarily made up of families
2. Houston as a whole is diverse, and it will be important for evacuation instructions to be translated into multiple languages.



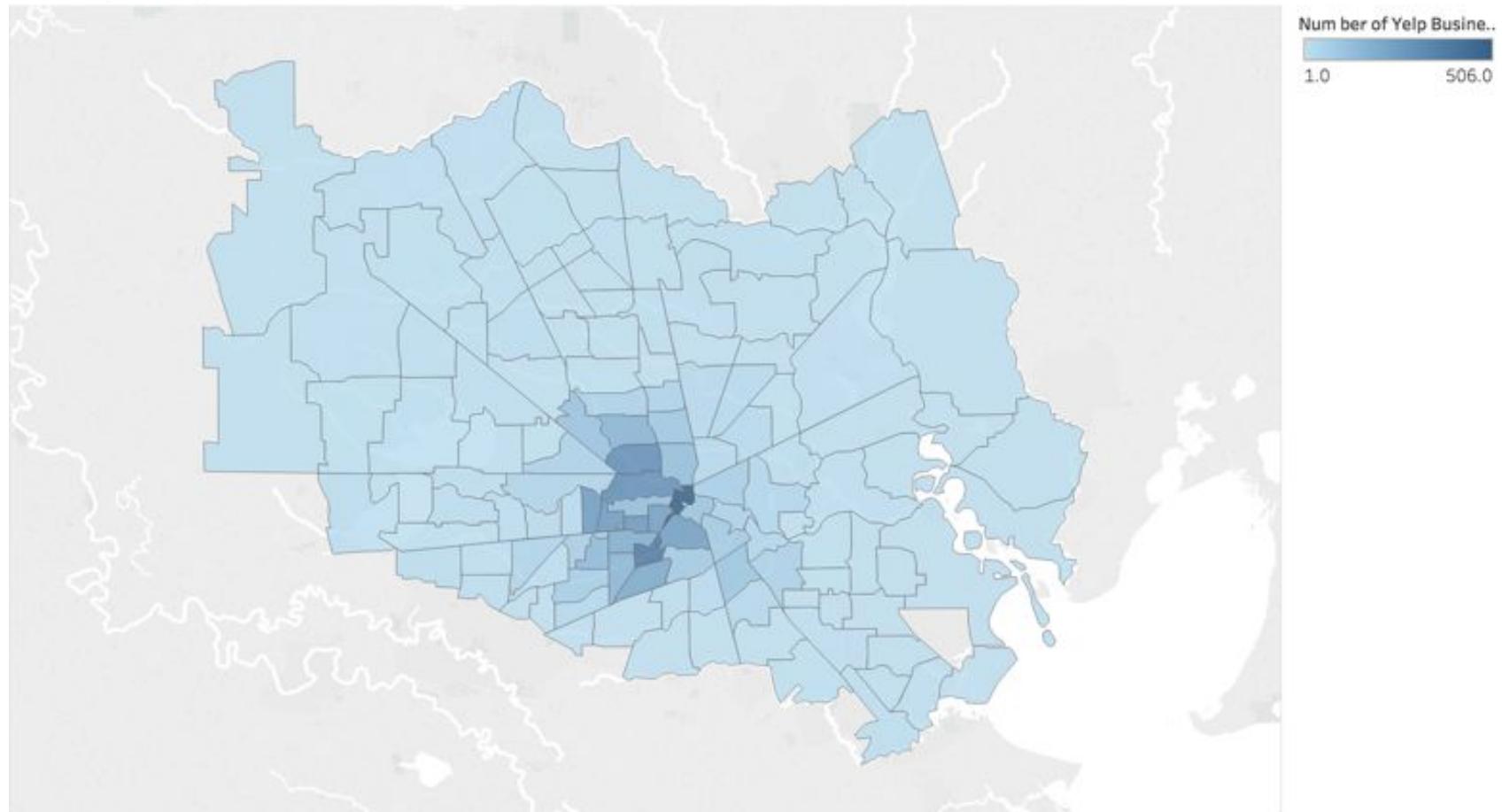
# Main Takeaways:

- 1. Understand community's unique needs.**
- 2. Plan disaster response around data and inference.**
- 3. Focus on clear communication.**

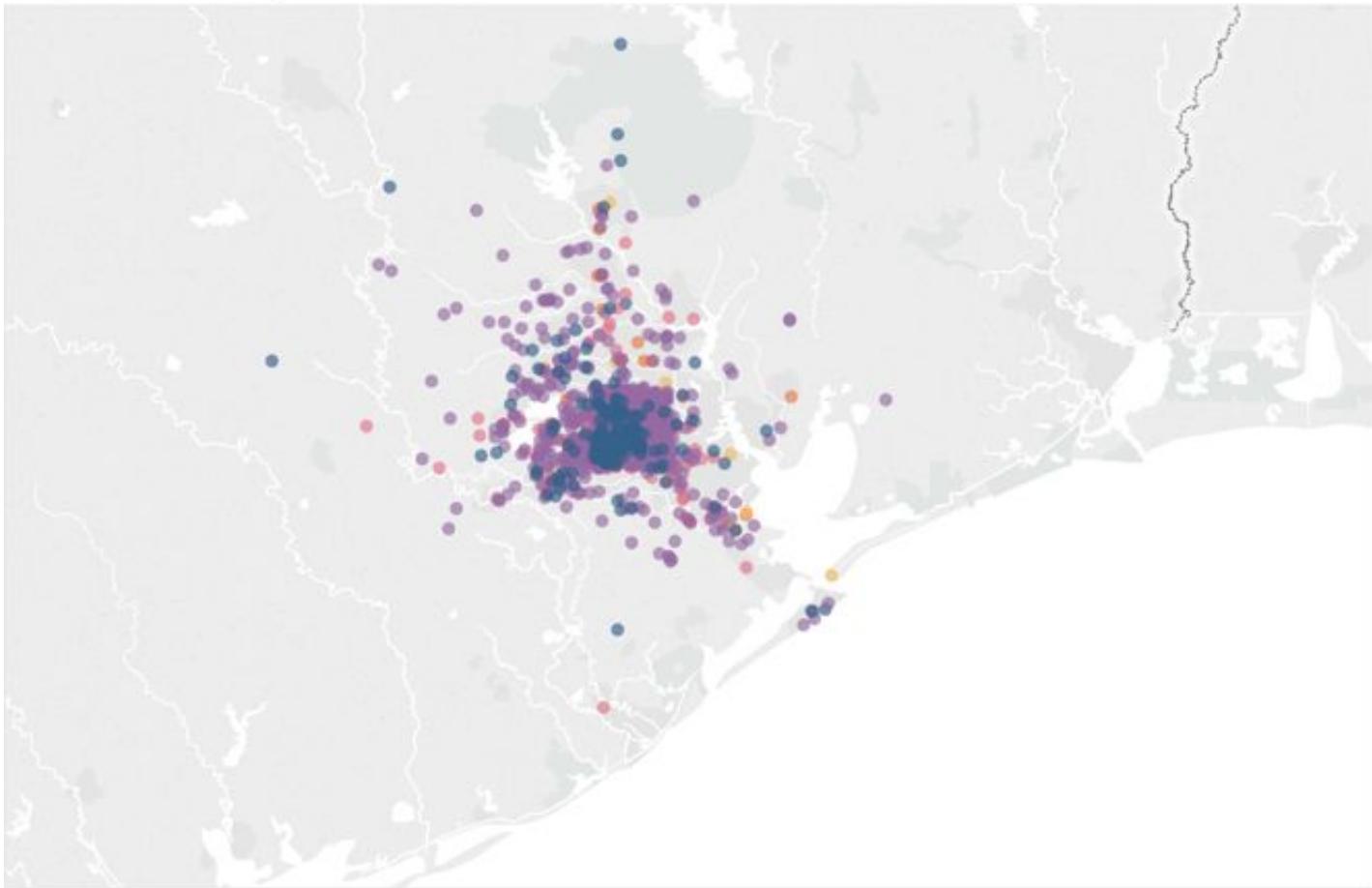
# Mapping the Lifelines



Density of Yelp-identified Businesses



## Yelp Businesses by Life Line



● Safety and Security

● Food, Water,  
Shelter

● Health and Medical

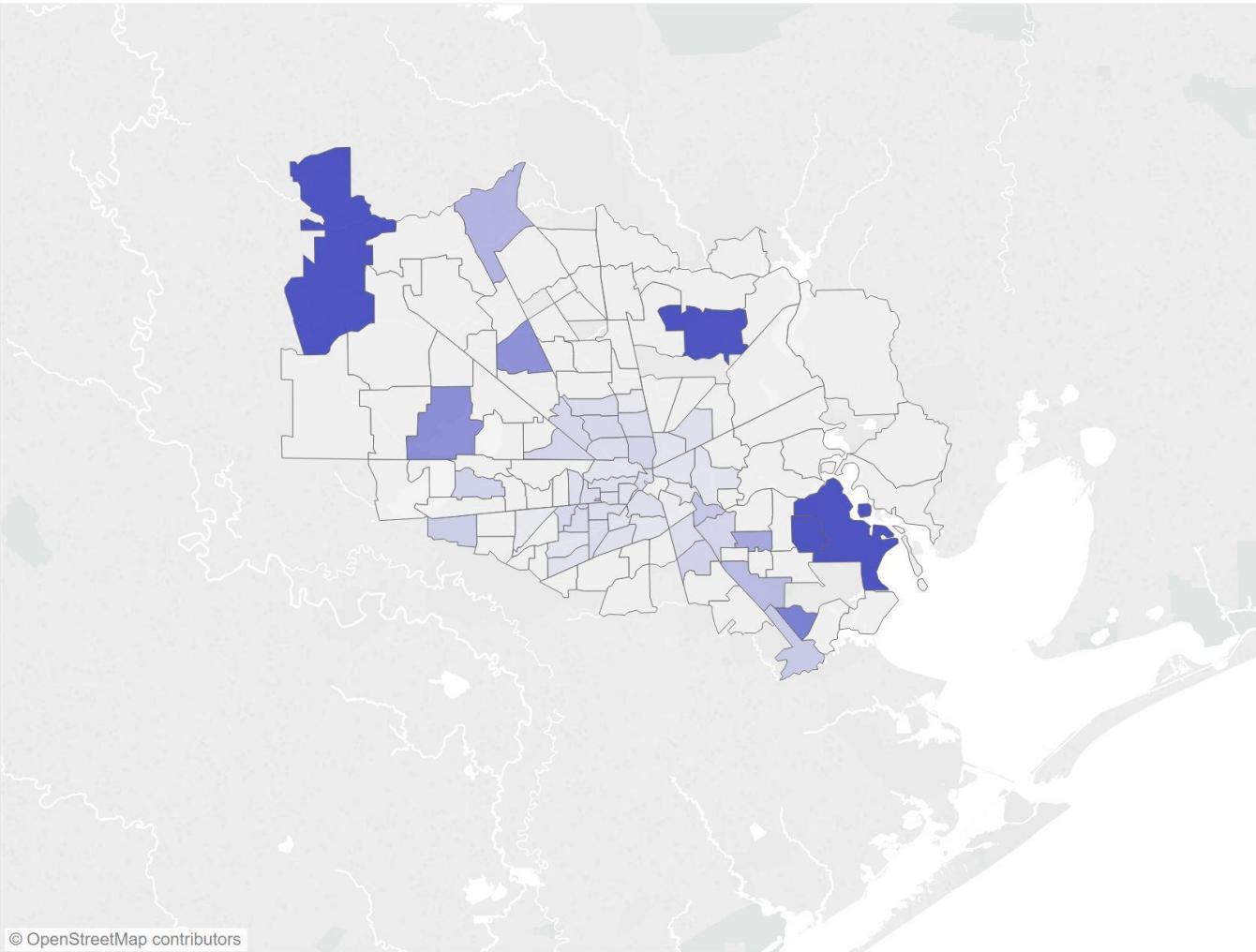
● Energy

● Communications

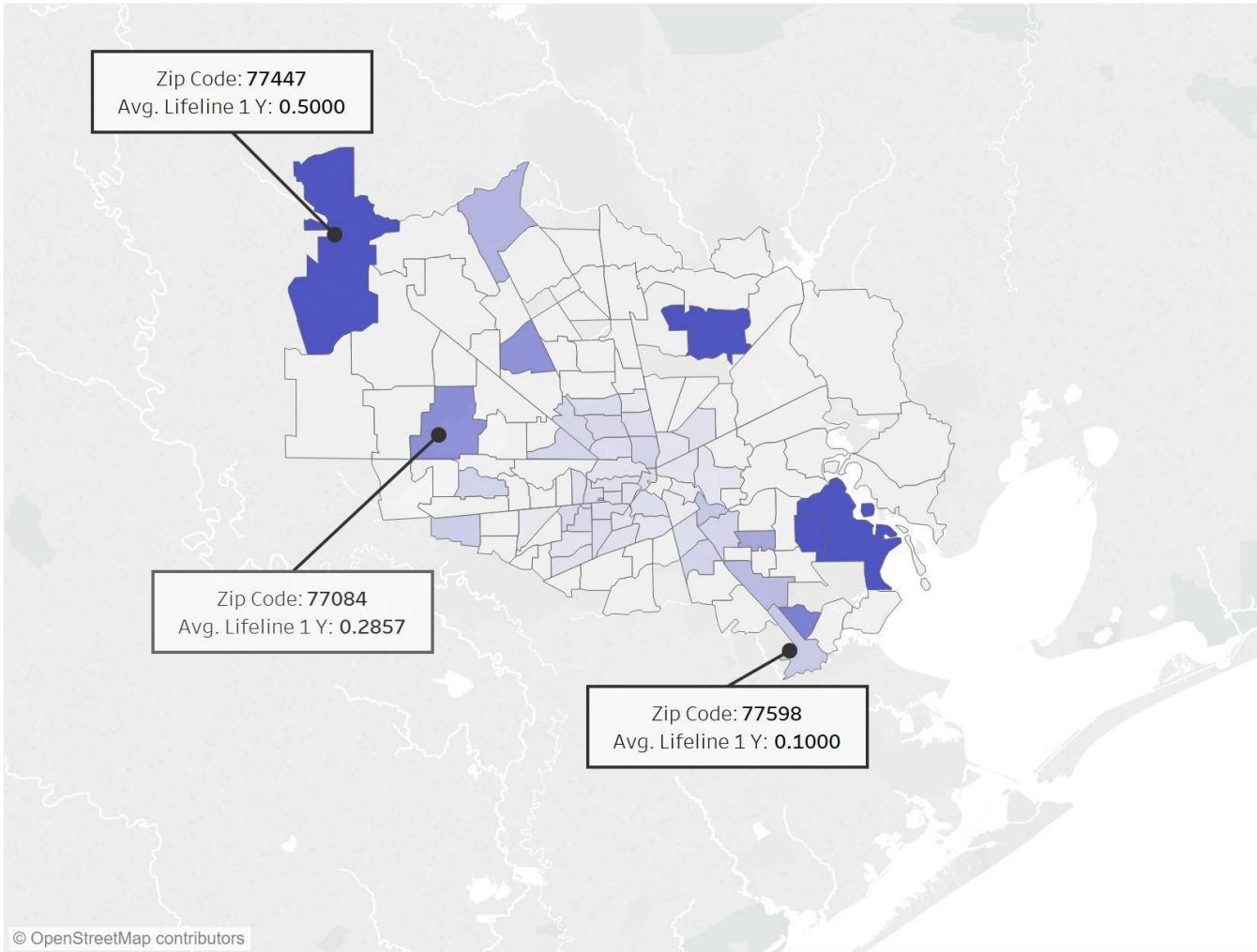
● Transportation

● Hazardous Material

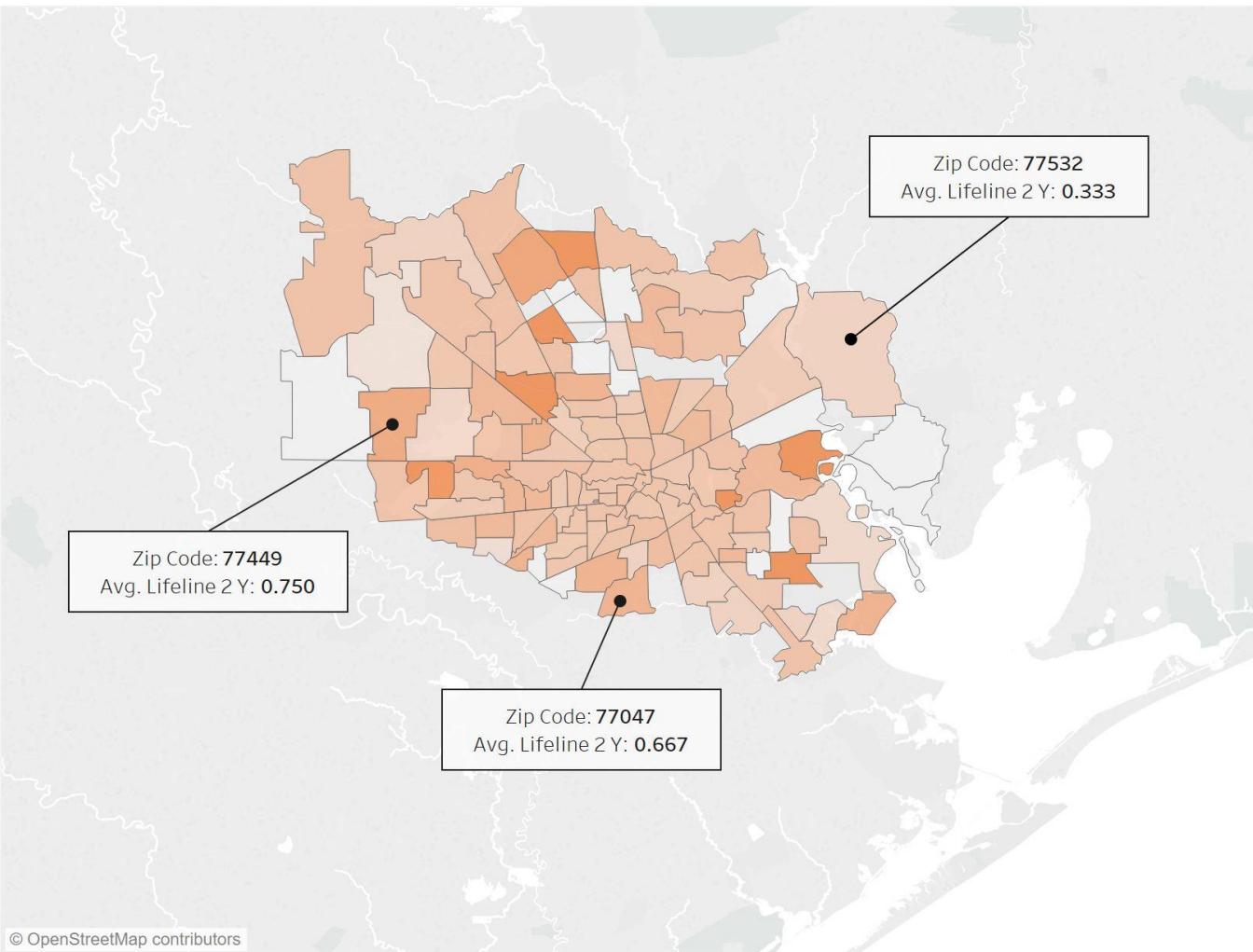
## Lifeline 1: Safety & Security Listings within Zip Code



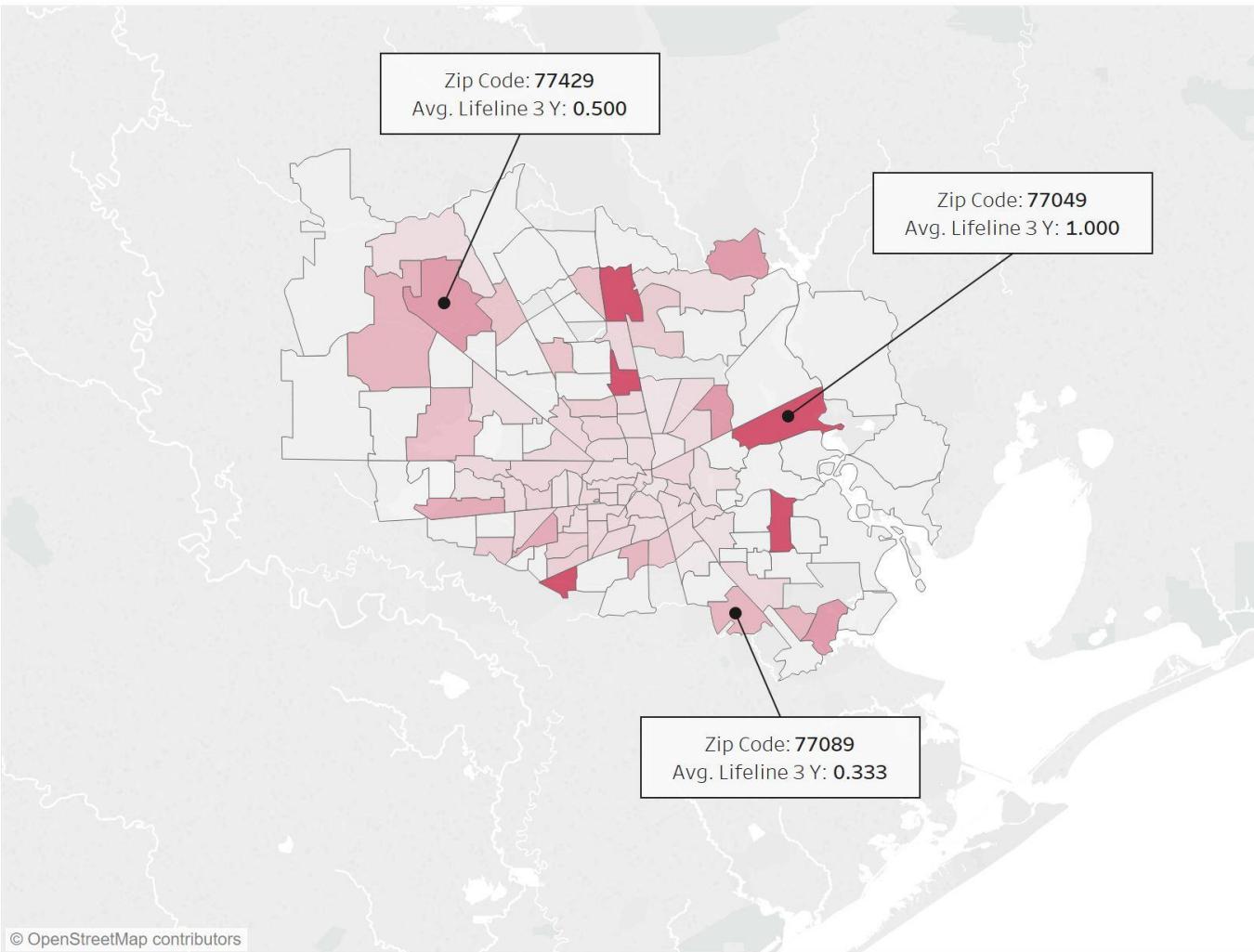
## Lifeline 1: Safety & Security Listings within Zip Code.



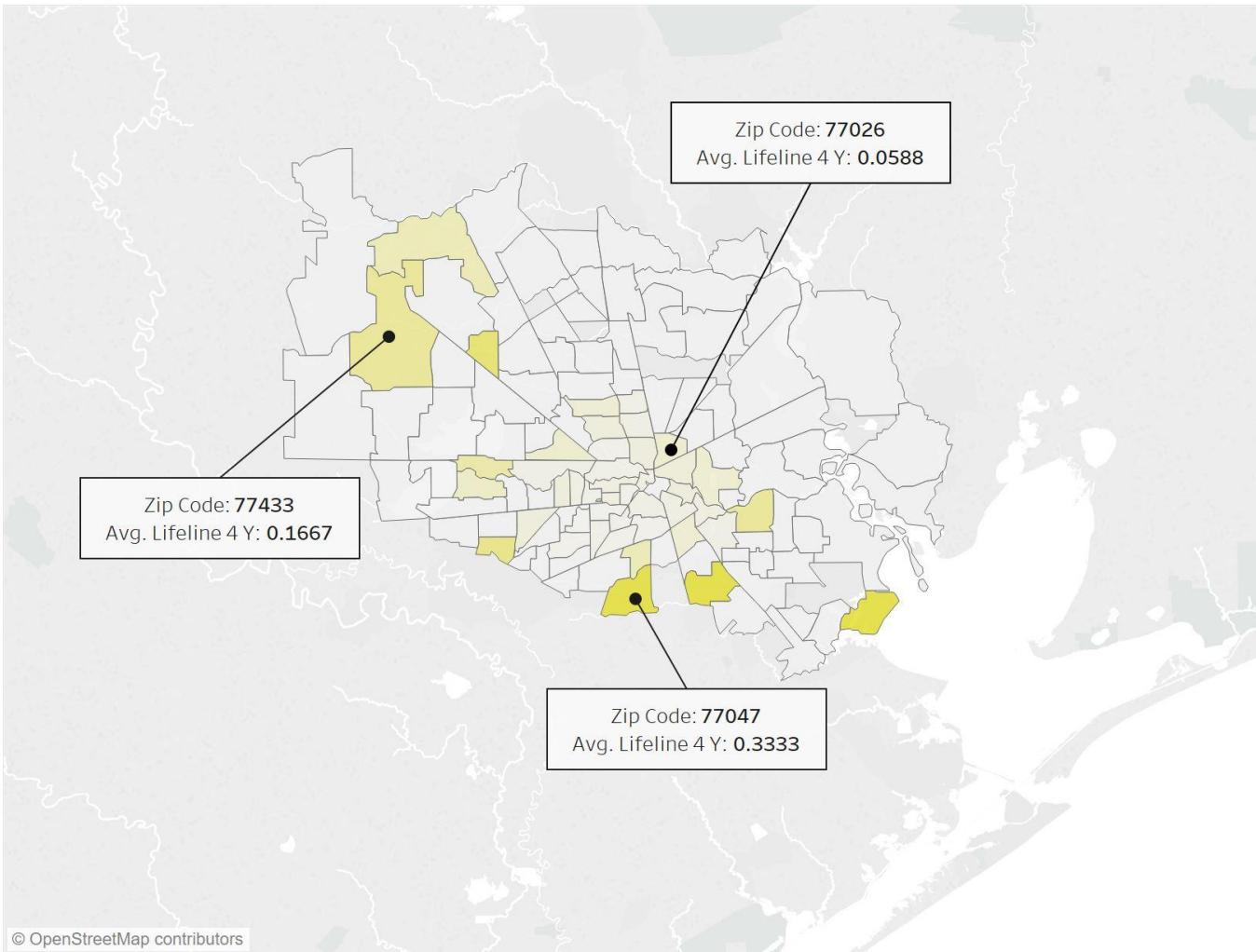
## Lifeline 2: Food, Water, Shelter Listings within Zip Code



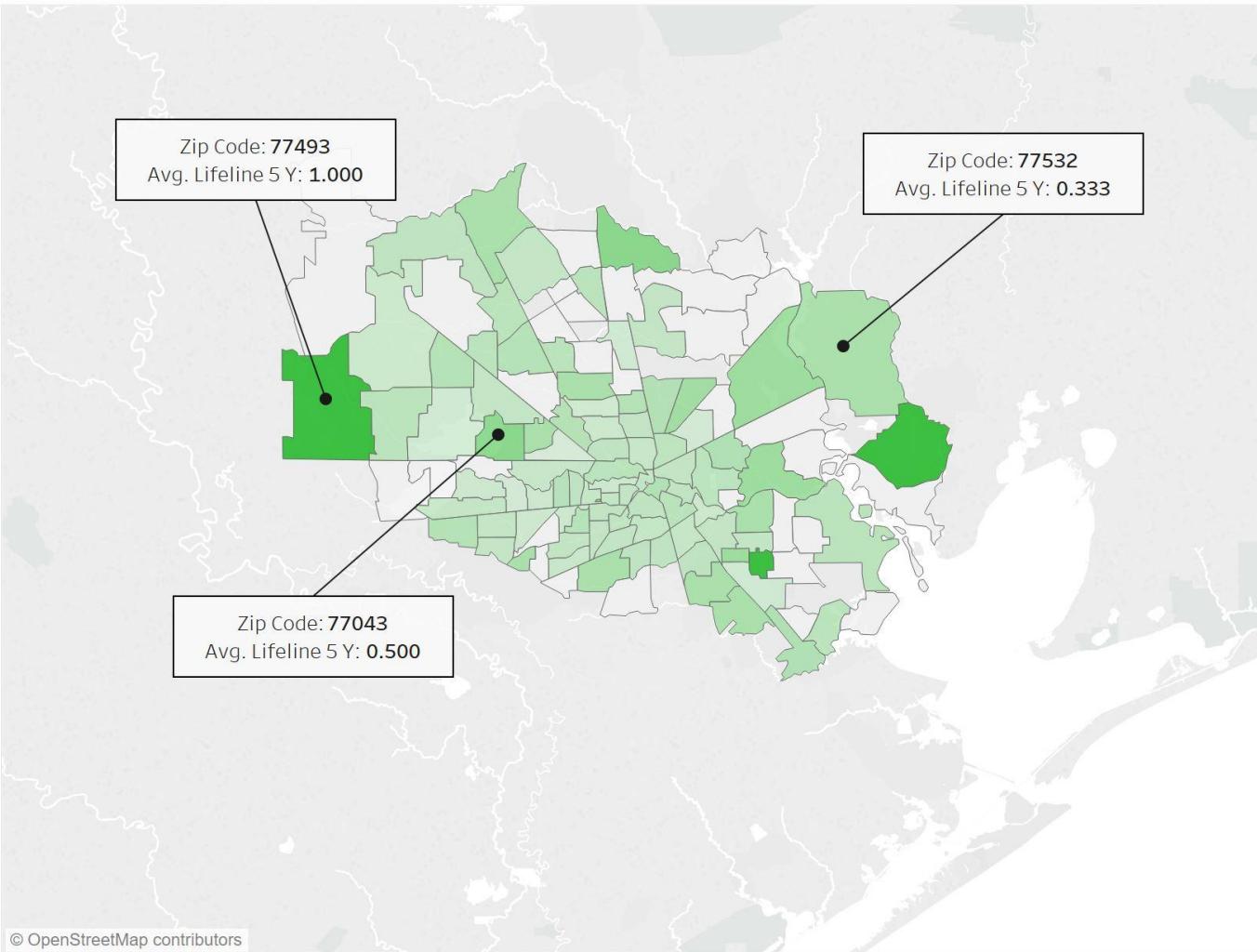
### Lifeline 3: Health & Medical Listings within Zip Code



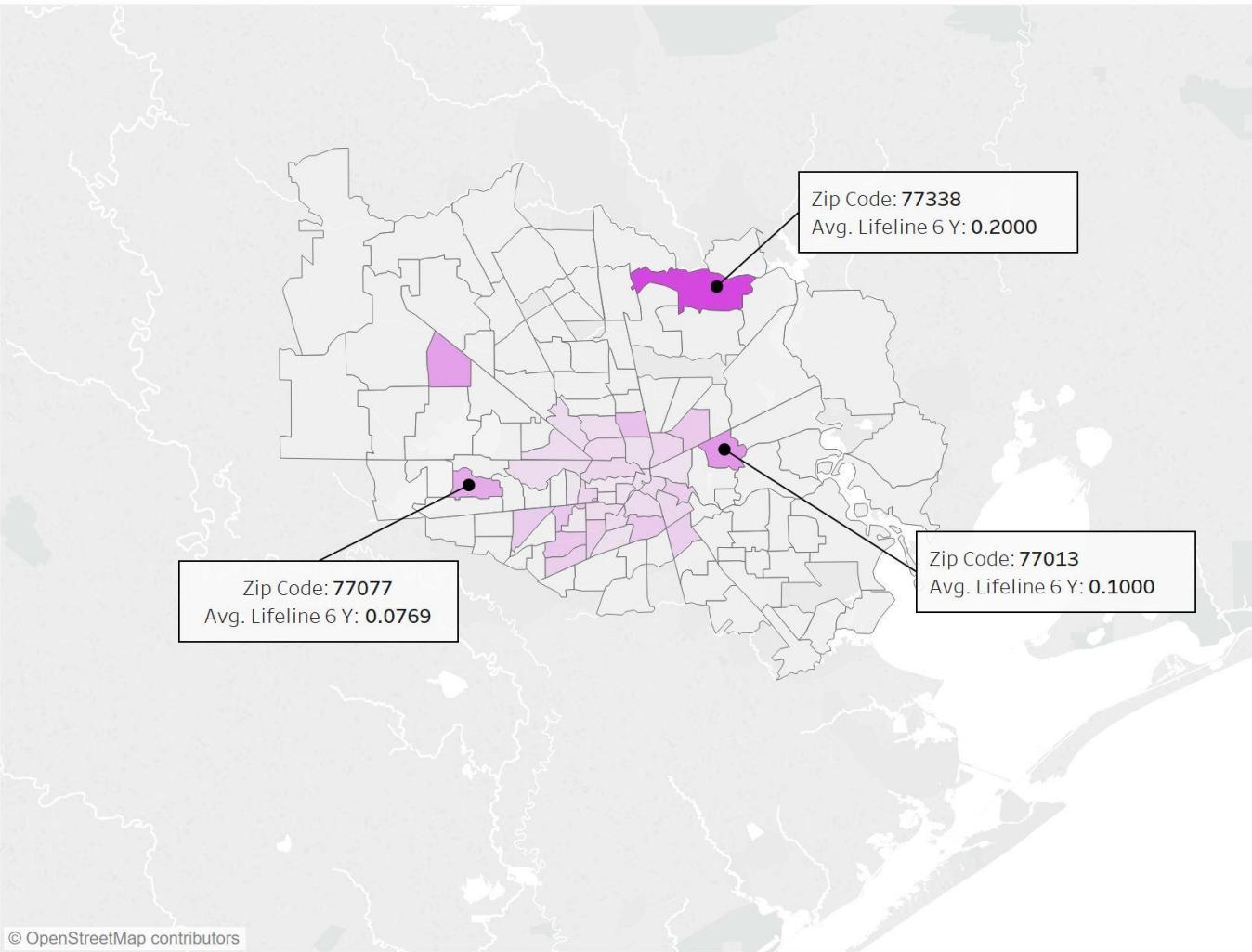
## Lifeline 4: Energy Listings within Zip Code



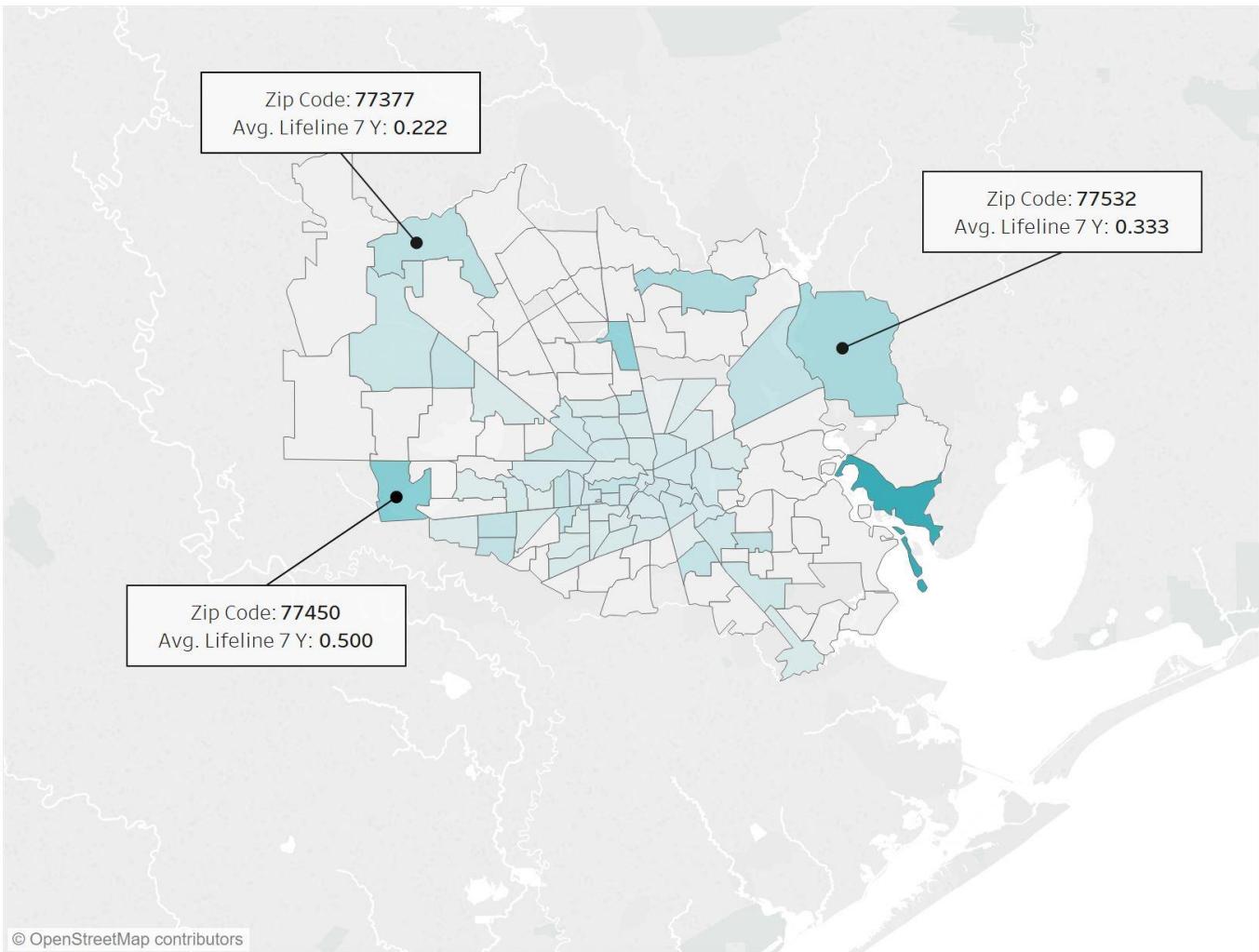
## Lifeline 5: Communication Listings within Zip Code



## Lifeline 6: Transportation Listings within Zip Code



## Lifeline 7: Hazardous Material Listings within Zip Code



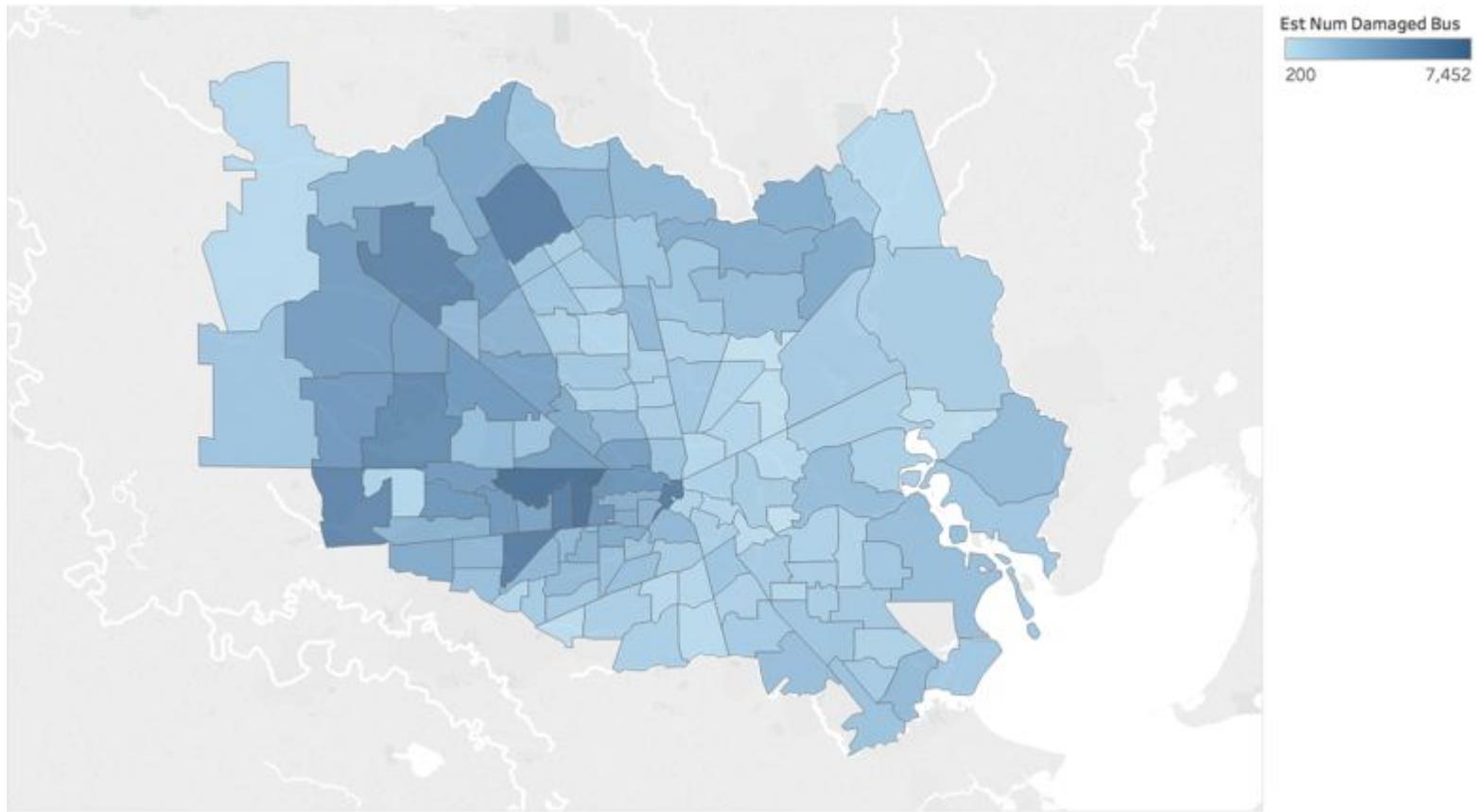


# HURRICANE HARVEY

# HURRICANE HARVEY FACTS:

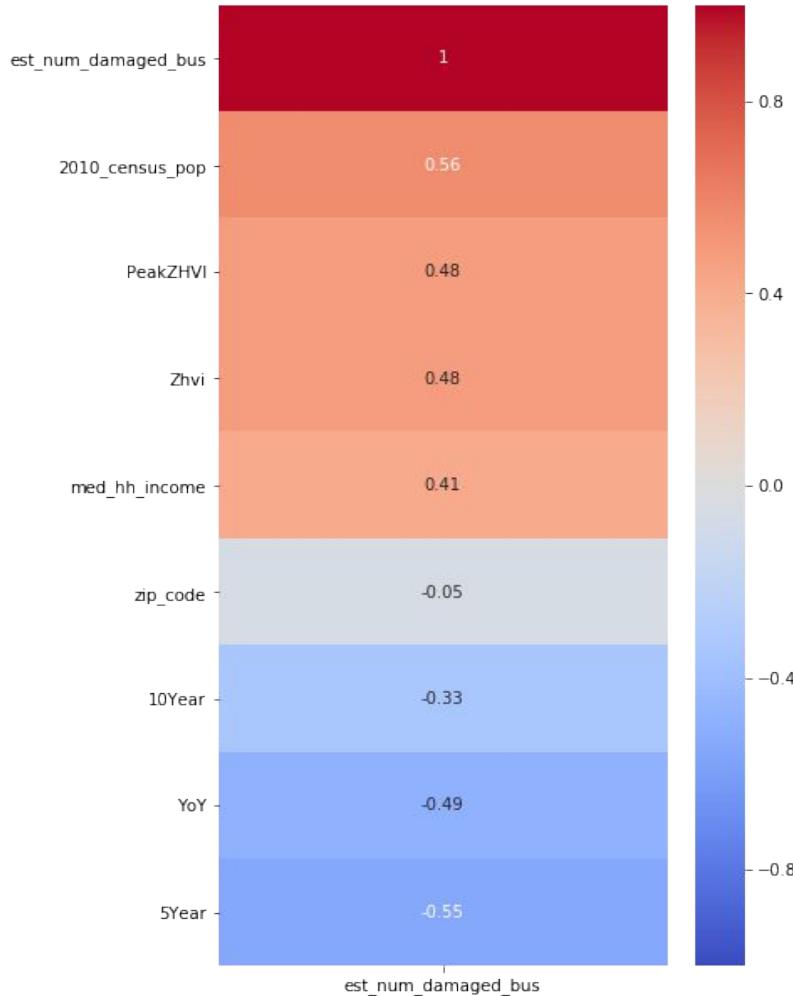
- **Category 4** storm that hit Texas on **August 25, 2017**
- Made landfall three times in six days.
- **Two feet of rain fell in the first 24 hours**, creating an unprecedented 1,000-year flood event.
- At its peak **33% of Houston was underwater** causing the city to sink two centimeters due to the water's weight
- **88 people died** from the storm.
- Forced 39,000 people out of their homes and into shelters.
- Left **200 million cubic yards of debris**.
- Caused **\$125 billion in damage** (National Hurricane Center)
- Damaged 204,000 homes.
- 738,000 people registered for assistance with **FEMA**, which paid out **\$378 million in direct payments**.
- In the Gulf area, **one million vehicles ruined** beyond repair (Black Book).

## Estimated Number of Damaged Businesses



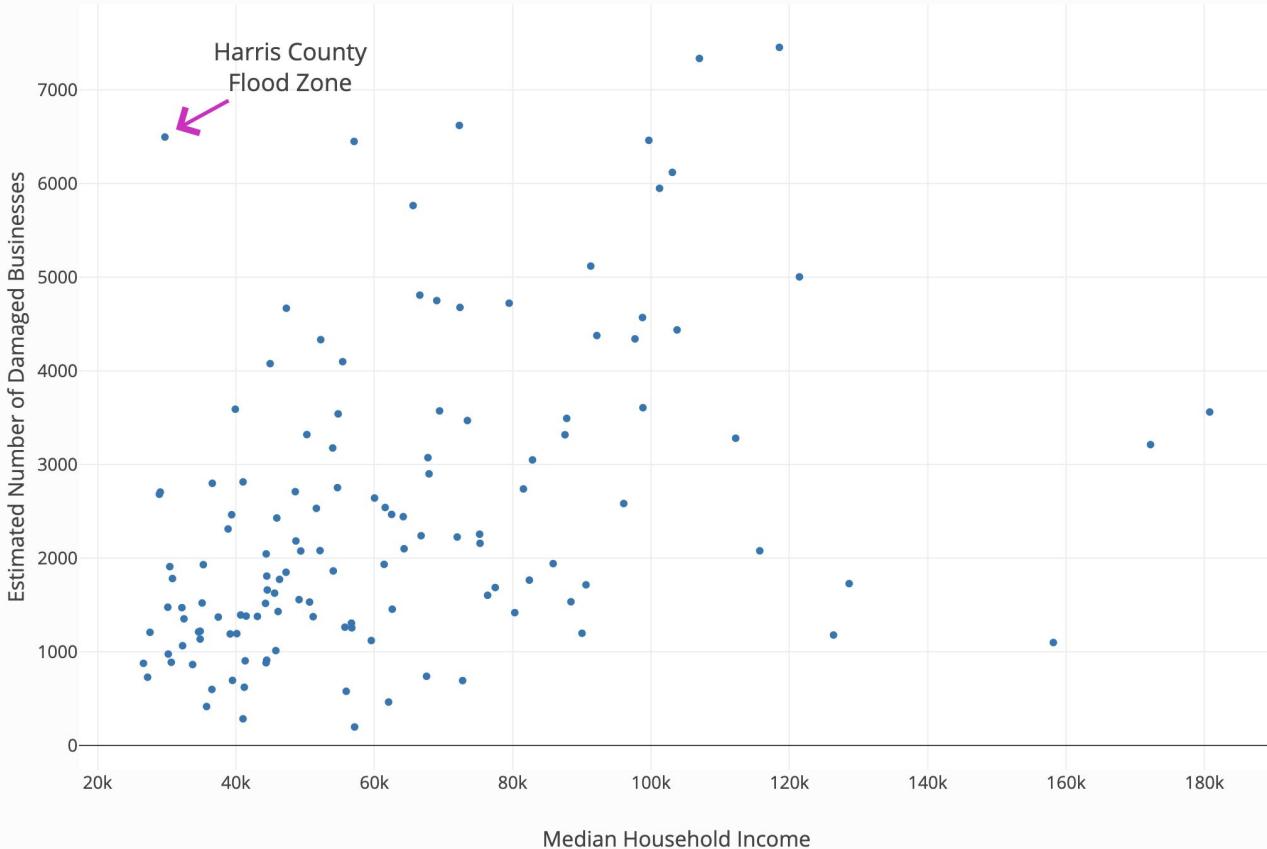
# Damaged Businesses Correlations

Correlations for damaged businesses



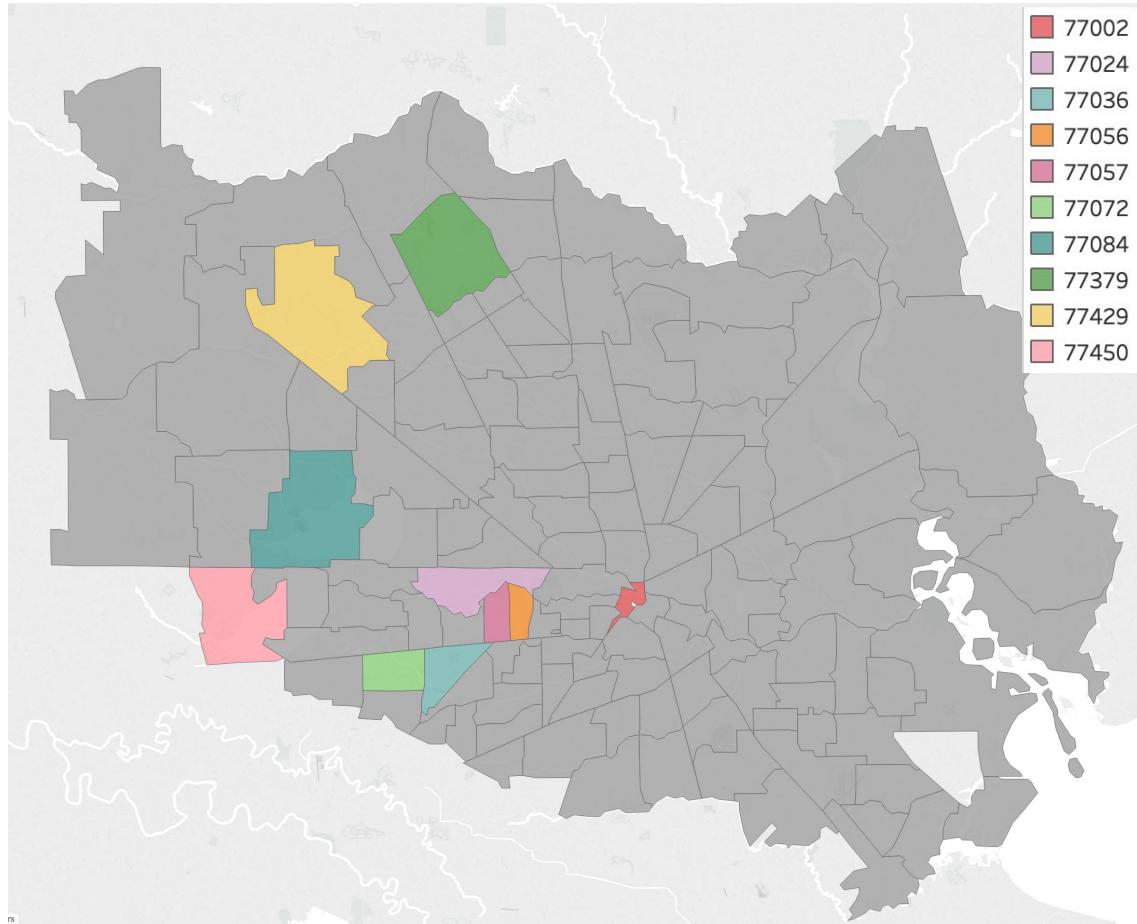
# Damaged Businesses Correlations

Median Household Income v.s. Estimated Number of Damaged Businesses



# Most Damaged Zipcodes

Top 10 Damaged ZipCodes



# Most Damaged Zipcodes by Lifeline

	Safety Security	Food Water Shelter	Health Medical	Energy	Comm.	Trans.	Hazard Waste
zip_code	lifeline_1	lifeline_2	lifeline_3	lifeline_4	lifeline_5	lifeline_6	lifeline_7
77024	0.00	52.27	13.64	2.27	18.18	2.27	11.36
77056	3.96	48.02	9.69	2.20	22.03	1.32	12.78
77002	5.01	47.60	11.27	2.92	20.25	1.67	11.27
77036	2.00	50.00	14.00	4.00	16.00	4.00	10.00
77379	0.00	80.00	0.00	0.00	20.00	0.00	0.00
77057	0.00	42.31	7.69	3.85	26.92	0.00	19.23
77429	0.00	50.00	50.00	0.00	0.00	0.00	0.00
77450	0.00	50.00	0.00	0.00	0.00	0.00	50.00
77084	28.57	28.57	28.57	0.00	14.29	0.00	0.00
77027	5.13	44.44	14.53	1.28	18.80	2.99	12.82

# Conclusions and Next Steps:

## **Yelp Data**

Major limitations with Yelp data  
Possible to categorize businesses by lifeline  
Yelp API only returns businesses with at least 1 review

## **Hurricane Harvey**

By passing in location information we are able to extract the most impacted Zip codes  
Utilizing the Yelp Data, we were able to rank zip codes by damage and provide proportions for the lifelines

## **FEMA**

Fema can use this mapping to inform how to distribute resources among lifelines and geographies  
And to have ongoing updated of the economic effects of disasters

## **Suggestion: Yelp**

The Yelp data needs to be supplemented or replaced with other sources  
Zillow or a local/national government resource for business registrations or property values may provide a more complete picture

## **Suggestion: Real time data**

Using real time social media data would be a great application of the Lifeline categorizations  
The Twitter API is perfectly suited to pull a live stream of geolocated data

## **Tool Deployment**

- Deploy the map feature using Django or other feature
- Add features to collect additional data in real time

# Lessons Learned:

- Group think: all settling on one project no one is passionate about
- Continuous focus on overall problem
- Question -> Data 
- Data -> Question 
- Twitter/Social Media Usage
- Task management and delegation

# References

- <https://www.yelp.com/developers/documentation/v3/business>
- <https://blog.splitwise.com/2013/09/18/the-2010-us-census-population-by-zip-code-totally-free/>
- <https://www.kaggle.com/goldenoakresearch/us-household-income-stats-geo-locations>
- <https://www.zillow.com/research/data/>
- <http://data.houstontx.gov/dataset>
- <http://pdata.hcad.org/GIS/>