Impact of COVID on real estate data in Houston and beyond

By Ophelia Heard, Iryna Petrovska Marchaino, Frank Perez

Rice University Analytics & Visualization Bootcamp

Houston, TX
2021

Did something happen?!

— — —

- These days Houston's housing market is going through some major changes.

- It is not unheard of to see >10-20 offers bidding on the same property, leading to at times over-inflation of sale price beyond that of the listing.

- We also keep hearing about all those "New Yorkers and Silicone Valley folks" moving down to "cheaper" states, now that remote work in tech sector is much more common.

So…

Project Objective

———

Our key objective is to understand what impact, if any, did COVID have on the real estate market in the US, in particular:

- in Houston TX vs other cities and
- in Harris County vs other counties in TX

Methodology & Data

— — —

Three key sources of information we used were:

- US Census data
  https://txcip.org/tac/census/
- Real estate data
  https://www.zillow.com/research/data/
- Johns Hopkins COVID data
  https://coronavirus.jhu.edu/about/how-to-use-our-data

Data Cleaning: Census data

— — —

The Texas County Information Program provided census data for all 254 counties. The data contained in CIP's database are obtained from official sources (https://www.census.gov/), but allowed us to parse the info needed more easily for our analysis.

TX Counties (5):

- Harris
- Fort Bend
- Montgomery
- Galveston
- Trinity

Data Retrieved:

- Population
- Unemployment Rate
- Annual Income (2019)
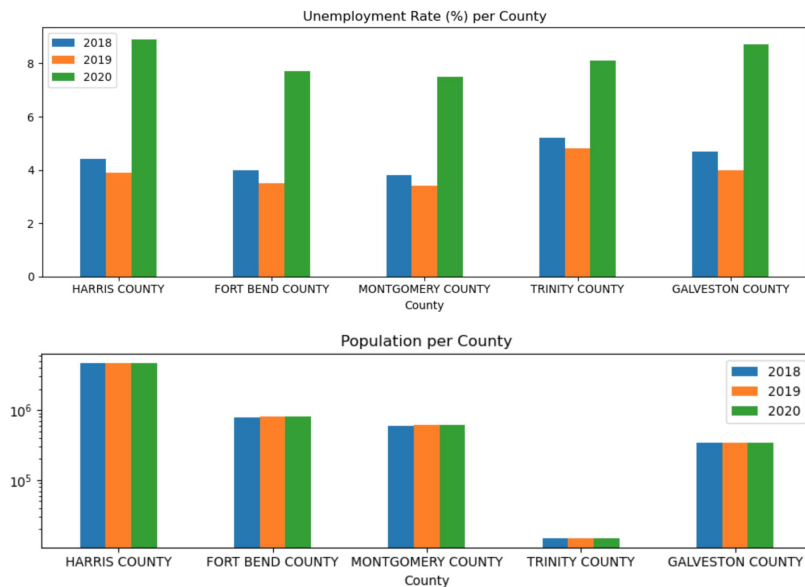
Data Cleaning: Census Data (cont'd)

Data Frames from web scraping data URLs

County	Median Household Income - 2019 (Census Bureau):	Per Capita Income - 2019 (BEA):	Total Personal Income - 2019 (BEA):
HARRIS COUNTY	$61,638	$60,002	$282,809,166,000
FORT BEND COUNTY	$101,361	$59,653	$48,419,712,000
MONTGOMERY COUNTY	$88,833	$63,424	$38,523,113,000
TRINITY COUNTY	$43,788	$36,062	$528,344,000
GALVESTON COUNTY	$73,214	$54,250	$18,561,154,000

County	Unemployment Rate - 2018 (%)	Unemployment Rate - 2019 (%)	Unemployment Rate - 2020 (%)
HARRIS COUNTY	4.4	3.9	8.9
FORT BEND COUNTY	4.0	3.5	7.7
MONTGOMERY COUNTY	3.8	3.4	7.5
TRINITY COUNTY	5.2	4.8	8.1
GALVESTON COUNTY	4.7	4.0	8.7

Transposed version of data frame
for merging purposes, obtained using pd.melt()

	County	Year	Estimated Population
0	HARRIS COUNTY	2018	4676913
1	FORT BEND COUNTY	2018	789269
2	MONTGOMERY COUNTY	2018	590127
3	TRINITY COUNTY	2018	14672
4	GALVESTON COUNTY	2018	337639
5	HARRIS COUNTY	2019	4709243
6	FORT BEND COUNTY	2019	812737
7	MONTGOMERY COUNTY	2019	607583
8	TRINITY COUNTY	2019	14675
9	GALVESTON COUNTY	2019	341541
10	HARRIS COUNTY	2020	4709243
11	FORT BEND COUNTY	2020	812737
12	MONTGOMERY COUNTY	2020	607583
13	TRINITY COUNTY	2020	14675
14	GALVESTON COUNTY	2020	341541

Data Cleaning: Real Estate Data

— — —

- Zillow offered a variety of pretty well organized, clean *.csv files which were used to read in "live" data

- Following was the data available for City-level information:
  - Medial listing price for all properties (provided on per month basis)
  - Medial sale price for all properties (provided on per month basis)
  - The count of unique listings that were active at any time in a given month.
  - How long it takes homes in a region to change to pending status on Zillow.com after first being shown as for sale.

- Following was the data available for County-level information:
  - ZHVI (Zillow Housing Value Index) value for all single-family homes in a given county, for condo/coops ($), for all homes with 1, 2, 3, 4 and 5+ bedrooms ($)

Data Cleaning: Real Estate Data (cont'd)

--- ---

- Typical raw DataFrame from Zillow looks like this:

- After filtering by a given county we get a DF with single row

- We then use melt() function to transform Date labeled columns into a "Property vs Value" two column format

Category columns *Value columns*

	RegionID	SizeRank	RegionName	RegionType	StateName	2017-11-30	2017-12-31	2018-01-31	2018-02-28	2018-03-31	...	2020-07-31	2020-08-31	2020-09-30	2020-10-31
0	102001	0	United States	Country	NaN	274900.0	269900.0	270000.0	279000.0	289000.0	...	340000.0	342400.0	344000.0	343400.0
1	394913	1	New York, NY	Msa	NY	524000.0	519000.0	519900.0	539000.0	549000.0	...	630000.0	649000.0	649900.0	655000.0
2	753899	2	Los Angeles-Long Beach-Anaheim, CA	Msa	CA	795000.0	799000.0	799000.0	809000.0	840000.0	...	954000.0	975000.0	959000.0	975000.0
3	394463	3	Chicago, IL	Msa	IL	300000.0	297990.0	299000.0	319499.0	338999.0	...	349900.0	349900.0	350000.0	349000.0
4	394514	4	Dallas-Fort Worth, TX	Msa	TX	339990.0	340990.0	345000.0	350000.0	359990.0	...	359900.0	359900.0	359999.0	355000.0

	RegionID	SizeRank	RegionName	RegionType	StateName	2017-11-30	2017-12-31	2018-01-31	2018-02-28	2018-03-31	...	2020-07-31	2020-08-31	2020-09-30	2020-10-31
6	394692	6	Houston, TX	Msa	TX	314407.0	317084.0	317337.0	321341.0	327192.0	...	327202.0	331988.0	336969.0	337703.0

	RegionID	SizeRank	RegionName	RegionType	StateName	Date	MedianListPrice
0	394692	6	Houston, TX	Msa	TX	2017-11-30	314407.0
1	394692	6	Houston, TX	Msa	TX	2017-12-31	317084.0
2	394692	6	Houston, TX	Msa	TX	2018-01-31	317337.0
3	394692	6	Houston, TX	Msa	TX	2018-02-28	321341.0
4	394692	6	Houston, TX	Msa	TX	2018-03-31	327192.0

Data Cleaning: COVID data (1 of 4)

— — —

1. **Data Source**:

   - Johns Hopkins Coronavirus Resource Center
   - Link: https://coronavirus.jhu.edu/about/how-to-use-our-data

2. **Data Selected:**

   CASES AND DEATHS - U.S. (Updated Daily) (CSSE GitHub)

   - Covid19 confirmed cases US (csv files)
   - Covid19 deaths US (csv files)

   U.S. STATE LEVEL POLICY TRACKER - U.S. State Level Policy Data (Updated Regularly) (CCI GitHub)

   - Texas Policy (csv file)

Data Cleaning: COVID data (2 of 4)

— — —

3. **Data Exploration**

- **Cases and Deaths data**
    - CSV files generated daily
    - Contains data of (confirmed) cases and deaths by day reported, by county and state
- **U.S. State Level Policy Tracker** -
    - CSV files updated regularly
    - Contains data of announcements, ordinances, laws, regulations and service announcements

4. **Data Challenges**

- CSV files created daily (cases and deaths) and named differently
- Cases and deaths were provided accumulated over days (as opposed to daily cases)

— — —

5. Data Cleanup process

5.1 Before Jupyter Notebook

- **Get number of cases & deaths per day**
- **Identify TX policies impacting housing market**
- **Testing data using excel graphs**
- **Transposing data and merging data into one file**

Admin2	Province	Country	1/22/20	1/23/20	3/16/20	3/17/20	3/18/20
Fort Bend	Texas	US	0	0	9	9	12
Galveston	Texas	US	0	0	1	2	4
Harris	Texas	US	0	0	13	16	23
Montgom	Texas	US	0	0	4	4	7
Trinity	Texas	US	0	0	0	0	0

date	policy
6-Mar-20	Governor Abbott issued a statement regarding price-gouging of medical supplies, pointing out that should the
11-Mar-20	Governor Abbott and the Texas Department of Insurance requested insurersand HMOs to waive copayments,
11-Mar-20	Governor Abbott and the Texas Department of State Health Service surged Texans to take preventative measure
12-Mar-20	Governor Abbott hosted a call with Texas legislators, mayors, and judges, reiterating state support in the form o
13-Mar-20	The Governor declared a state of disaster in Texas due to COVID-19.
16-Mar-20	The Governor announced that he has waived the State of Texas Assessments of Academic Readiness testing req
16-Mar-20	The Governor announced that Texas is waiving certain rules relating to vehicle registration, parking placards for
16-Mar-20	The Governor granted the Office of the Attorney Generals request for suspension of certain open-meeting statu
17-Mar-20	The Texas Department of State Health Services confirmed the first death as a result of COVID-19.

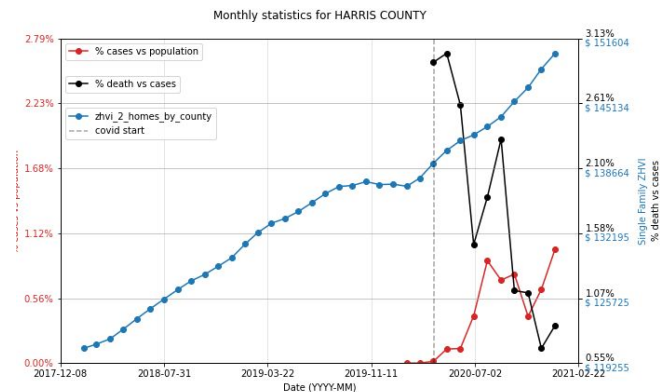| Date	County	Population	Cases	Deaths	Policies Announcements (Texas Government)
3/17/2020	FORT BEN	811688	0	0	The Texas Department of State Health Services confirm
3/19/2020	FORT BEN	811688	7	0	The Governor issued a series of Executive Orders to ac
3/20/2020	FORT BEN	811688	5	0	The Texas Supreme Court issued an emergency order s
3/28/2020	FORT BEN	811688	0	0	The Department of Health and Human Services waived
4/1/2020	FORT BEN	811688	0	0	The Governor and the Department of Housing and Com
4/24/2020	FORT BEN	811688	0	0	The Governor and the Texas Department of Housing an
4/29/2020	FORT BEN	811688	16	2	The Governor suspended certain statutes to temporari
5/8/2020	FORT BEN	811688	45	1	The Governor, the Texas Education Agency, and the Da

Daily cases rate per 100,000 by county

5. **Data Cleanup process**
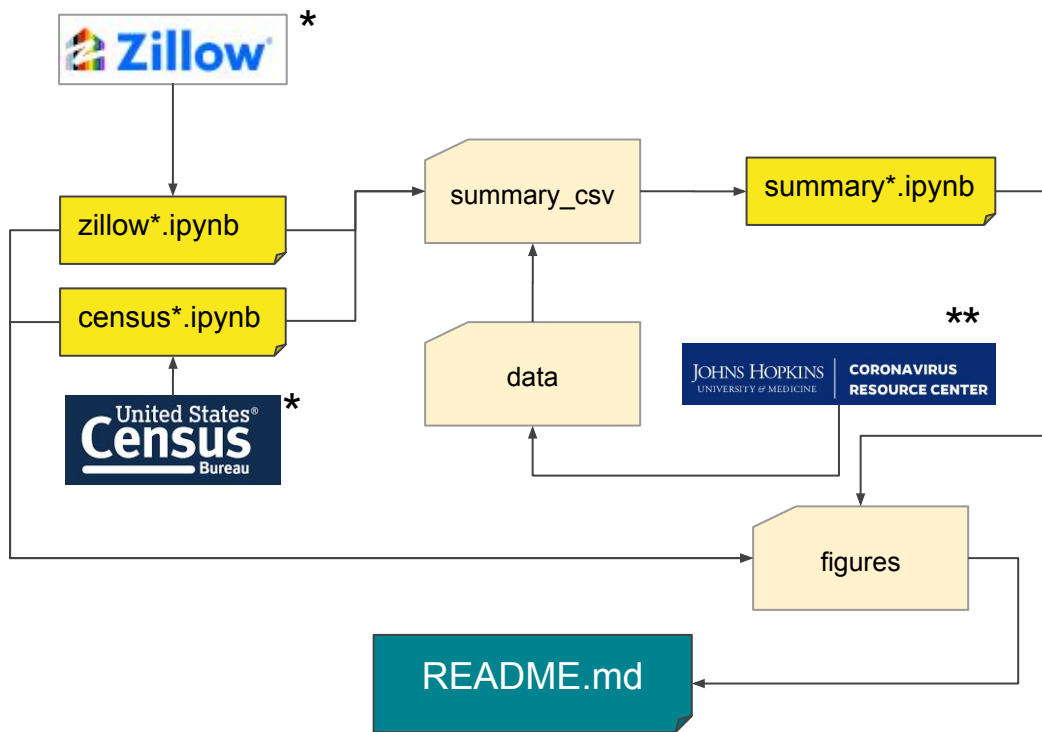
   5.2 Using Jupyter Notebook

   - **Open data**
   - **Shared with team and merge with housing data**

	Date	County	Population	Cases	Deaths	Policies & Announcements (Texas Government)
0	1/22/2020	FORT BEND COUNTY	811688	0	0	None
1	1/23/2020	FORT BEND COUNTY	811688	0	0	None
2	1/24/2020	FORT BEND COUNTY	811688	0	0	None
3	1/25/2020	FORT BEND COUNTY	811688	0	0	None
4	1/26/2020	FORT BEND COUNTY	811688	0	0	None
...	...	...	...	...	...	...
986	5/27/2021	GALVESTON COUNTY	342139	36	0	None
987	5/28/2021	GALVESTON COUNTY	342139	25	0	None
988	5/29/2021	GALVESTON COUNTY	342139	29	0	None
989	5/30/2021	GALVESTON COUNTY	342139	0	1	None
990	1/22/2020	HARRIS COUNTY	4713325	0	0	None

991 rows × 6 columns

Monthly statistics for HARRIS COUNTY

- % cases vs population
- % death vs cases
- zhvi_2_homes_by_county
- -- covid start

gitHub Repo Structure

- Data cleaning takes place in each respective notebook

- Individual summary DataFrames are saved to CSV

- All this summary CSVs are then merged together in the summary*.ipynb notebook where final analysis takes place

*   live data
* *   static/downloaded & pre-processed data
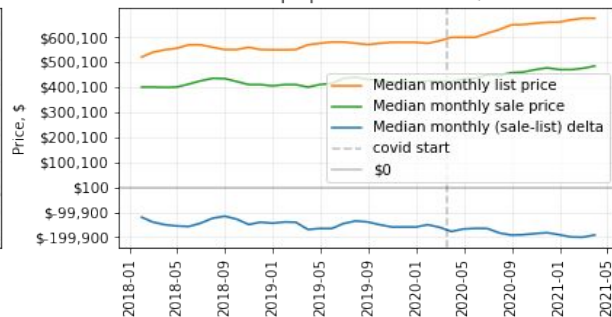
Listed vs Sold Price for Houston, TX

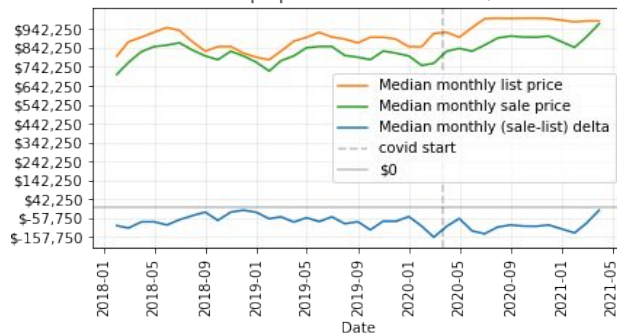Change in median list vs sale price for all properties in Houston, TX
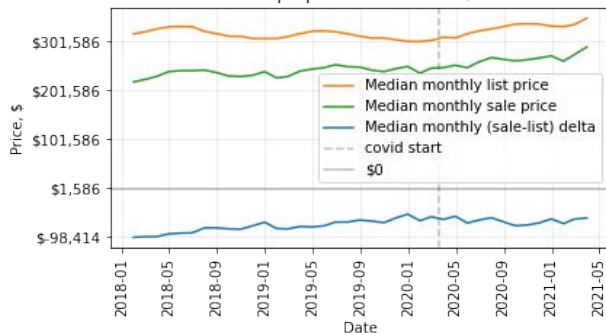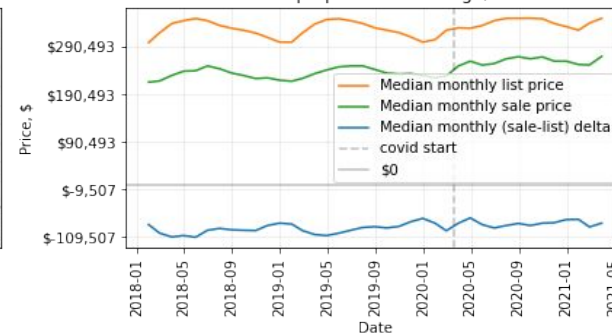

Change in median list vs sale price for all properties in Seattle, WA
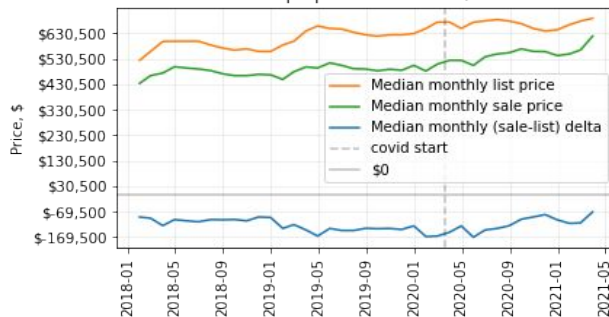
Change in median list vs sale price for all properties in New York, NY

Change in median list vs sale price for all properties in San Francisco, CA

Change in median list vs sale price for all properties in Chicago, IL

Key observations:

- San Francisco is almost hitting the point where properties are to be sold at listed price
- Implication is when list&sold converge - that people are willing to pay more

Real estate inventory FOR SALE & DAYS TO PENDING in Houston vs other metropolitan cities across US

— — —

Inventory count "For Sale" for all property types

Median count of days on the market prior to property satus chage from For Sale to Pending for all property types
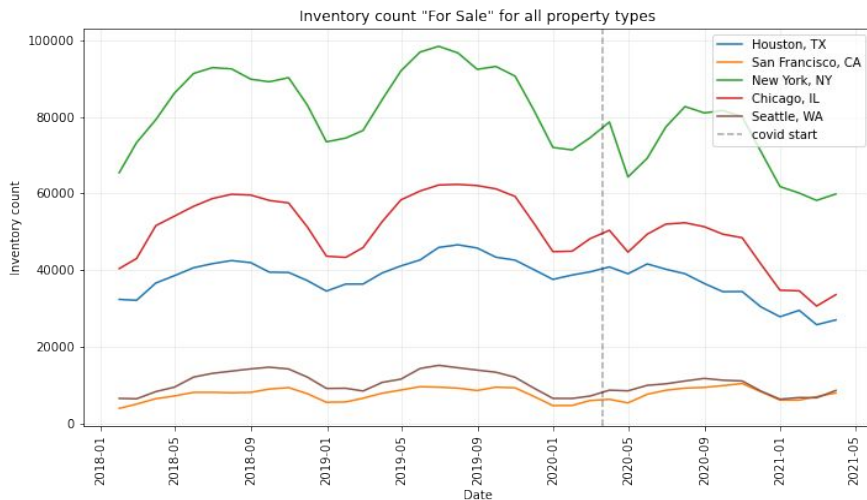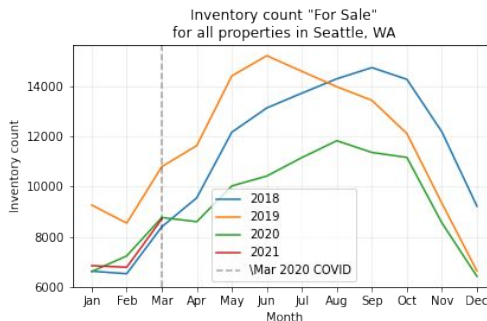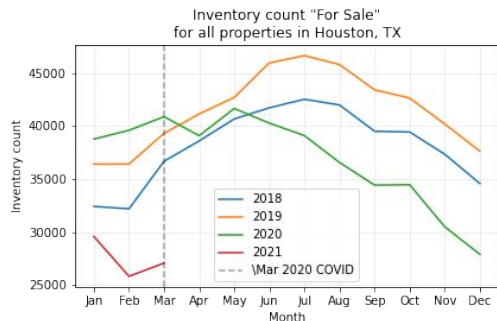
- Zillow data available through March'21 for all cities
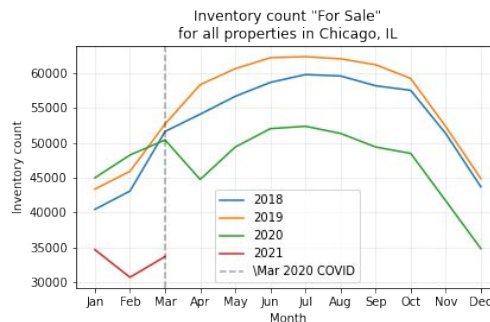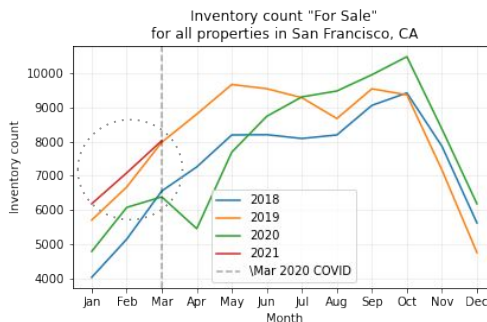- Number of houses for sale in San Francisco caught up with Seattle for the first time in 3 years
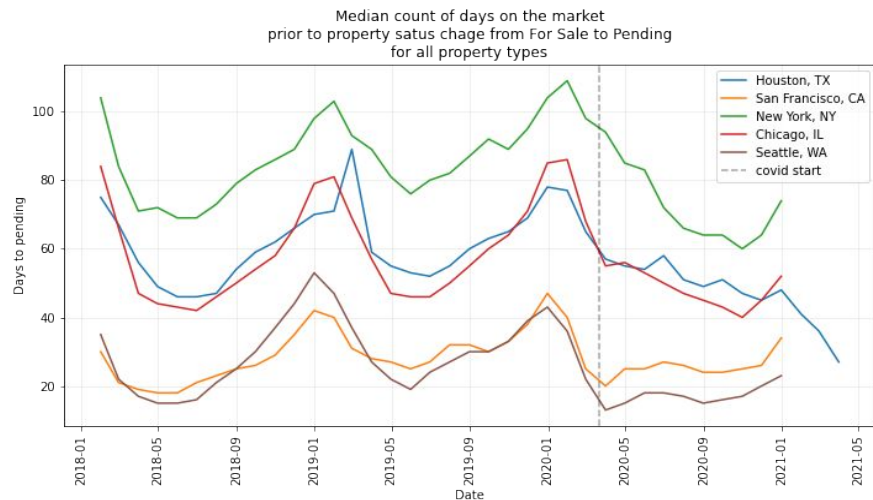
- Zillow data available through 2021 only for Houston, TX
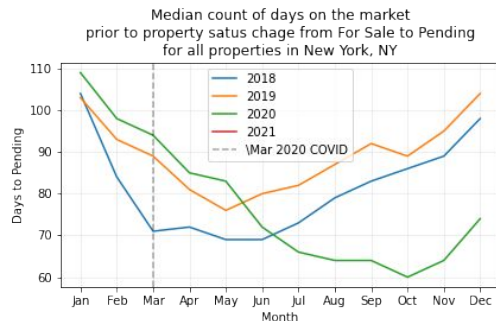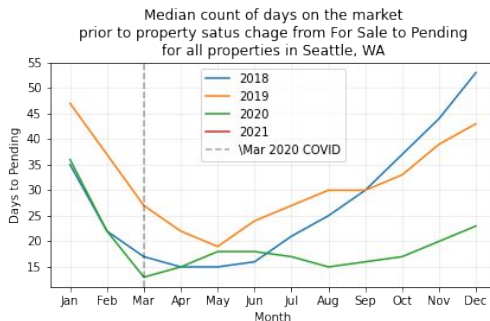- It is clear to see that for all markets, COVID had instantaneous impact on real estate market
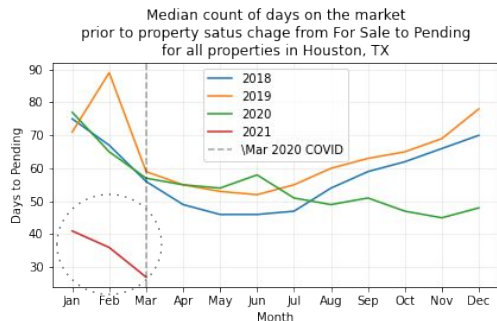
Real estate inventory FOR SALE (seasonal) in Houston vs other metropolitan cities across US
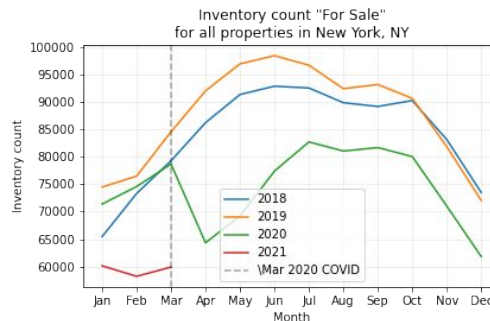

- Zillow data available through March'21

- Houston, New York and Chicago all have substantially lower inventory of homes for sale compared to prior years and start of 2020 specifically

- For the west coast, in Seattle inventory to date in 2021 has recovered to 2020 pre-COVID level

- San Francisco is the only city where 2021 inventory for sale exceeds 2020 levels (>1000 more listings)

- It is clear to see that for all markets, COVID had instantaneous impact on real estate market
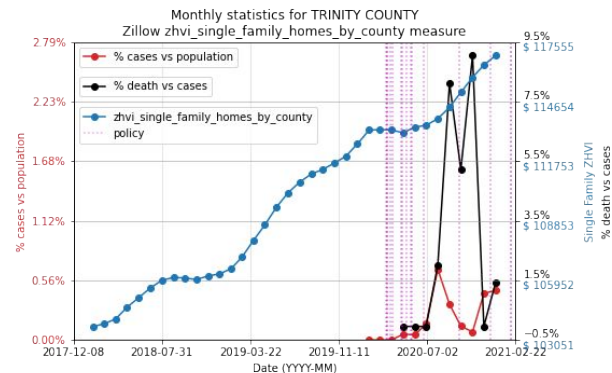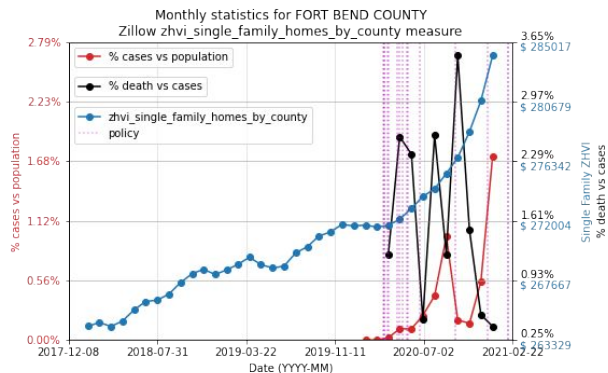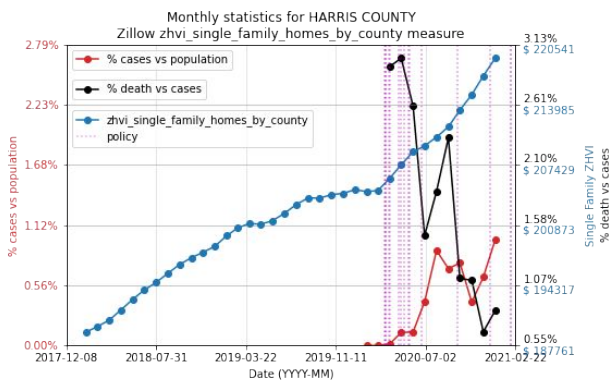
Real estate DAYS TO PENDING (seasonal) in Houston vs other metropolitan cities across US

Median count of days on the market prior to property satus chage from For Sale to Pending for all properties in Houston, TX

Median count of days on the market prior to property satus chage from For Sale to Pending for all properties in Seattle, WA

Median count of days on the market prior to property satus chage from For Sale to Pending for all properties in New York, NY

Median count of days on the market prior to property satus chage from For Sale to Pending for all properties in San Francisco, CA

Median count of days on the market prior to property satus chage from For Sale to Pending for all properties in Chicago, IL

- Zillow data available through 2021 only for Houston, TX

- A more than 50% drop in DAYS to PENDING can be seen from 2020 to 2021

- Across all reviewed cities we can see that DAYS to PENDING decreased throughout 2020, with only San Francisco showing values at the end of 2020 closer to those at the beginning of the year

- It is clear to see that for all markets, COVID had instantaneous impact on real estate market
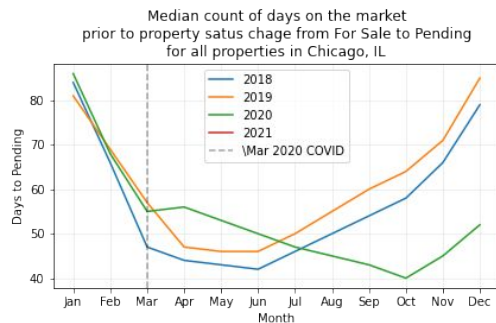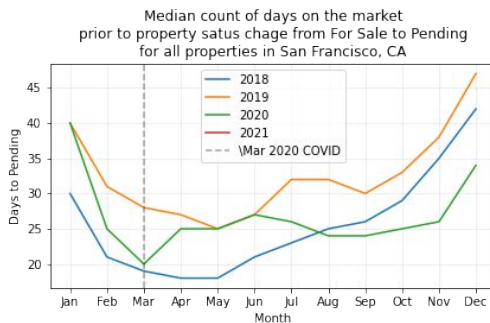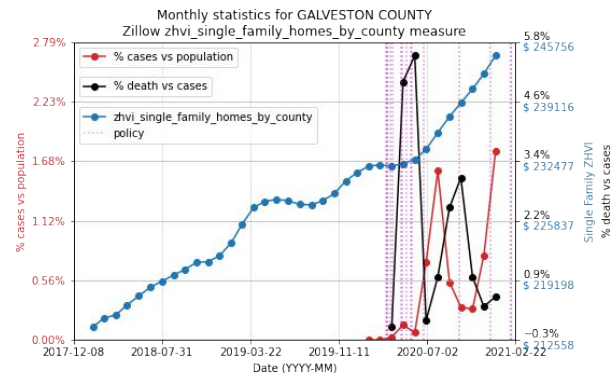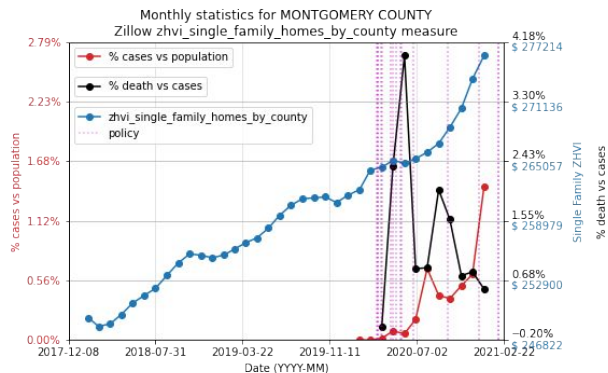
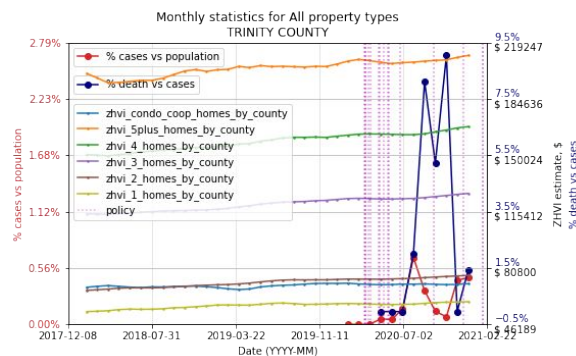# Impact of COVID cases\death and COVID related legislation on single family home values in different counties

Monthly statistics for HARRIS COUNTY
Zillow zhvi_single_family_homes_by_county measure

Monthly statistics for FORT BEND COUNTY
Zillow zhvi_single_family_homes_by_county measure

Monthly statistics for TRINITY COUNTY
Zillow zhvi_single_family_homes_by_county measure

Monthly statistics for MONTGOMERY COUNTY
Zillow zhvi_single_family_homes_by_county measure

Monthly statistics for GALVESTON COUNTY
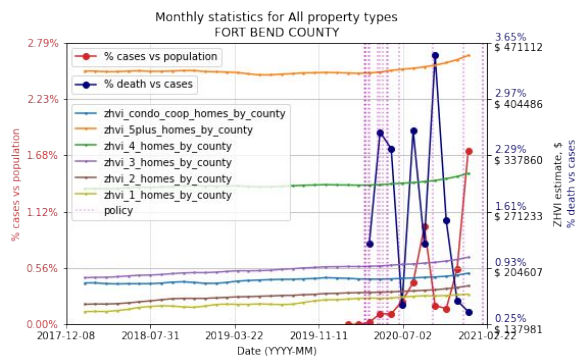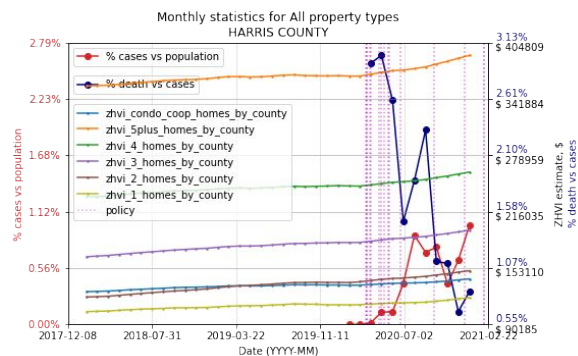Zillow zhvi_single_family_homes_by_county measure

Key observations:

- Spike in mortality typically associated with a "slow down" in ZHVI growth
- ZHVI grows post-covid is apparent across all counties
- COVID-related legislation seems to be associated with mortality spikes

Impact of COVID cases\death and COVID related legislation In Harris County across different property types

Monthly statistics for All property types
HARRIS COUNTY

Monthly statistics for All property types
FORT BEND COUNTY

Monthly statistics for All property types
TRINITY COUNTY

Monthly statistics for All property types
MONTGOMERY COUNTY

Monthly statistics for All property types
GALVESTON COUNTY

Key observations:

- Though cost of 2bed homes in Harris county was surpassing that of condos prior to COVID, the gap really widened during the pandemic
- Are people willing to pay for extra bedroom to be used as home office?...
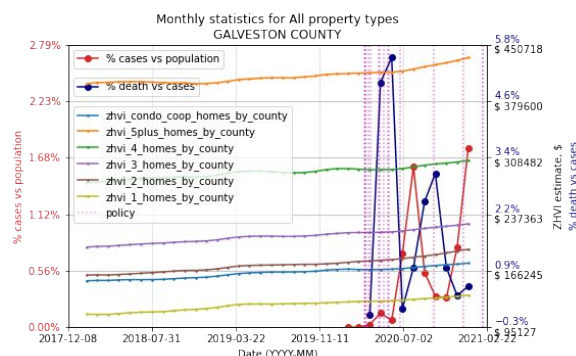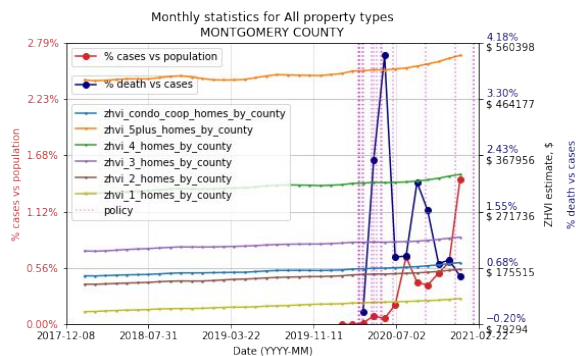
If we had more time...

— — —

- Use google maps API and produce graphics with markers for selected cities/counties (and maybe a heatmap of % ZHVI change)
- Run regression on ZHVI vs Deaths/Cases/Vaccinations count and numerically identify cities/counties with highest correlation between the two
- Calculate % increase in ZHVI for various property types and administrative entities and visualize those as a relative measure, potentially run regression vs COVID stats
- Split data in time chunks based on intervals between key COVID milestones:
    - 1st case detected
    - 1st death
    - Date of highest deaths count
    - Mask mandate in effect
    - Vaccination start date

Conclusions

———

- We were able to successfully pull and integrate data from 3 different sources

- We saw that COVID had differing impact on different metropolitan areas, such that for sale inventory in San Francisco increased beyond pre pandemic levels while that of the other cities reviewed mostly remained substantially lower compared to the start of 2020

- We also saw that in Harris county 2bed homes increased in value compared to condos much faster than in other Texas counties reviewed.
  We speculate that demand for larger (more bedrooms) housing options was driven by the need of using one of the bedrooms as home office.