The American Dream (Nightmare)

The future of housing and its impact from COVID

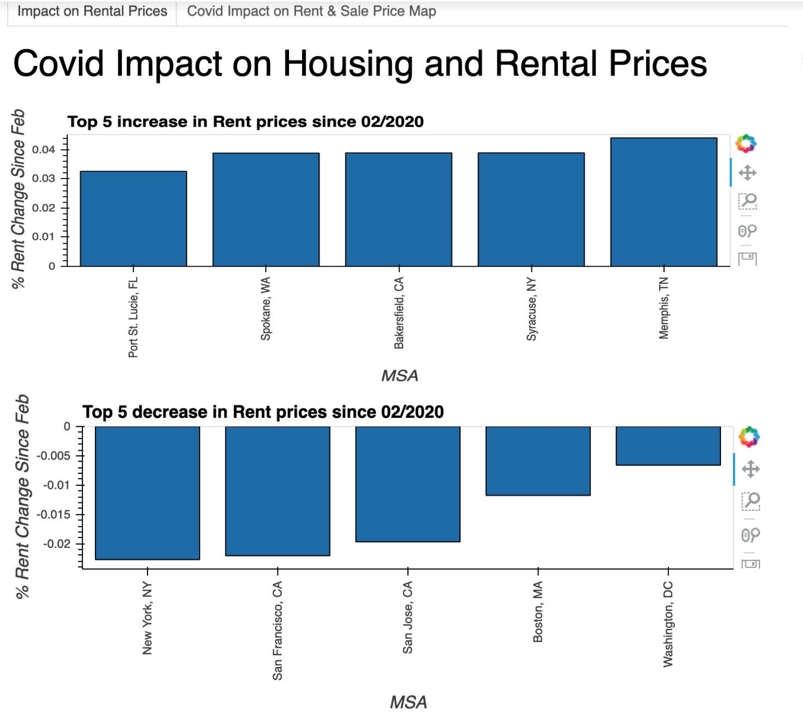
Introduction

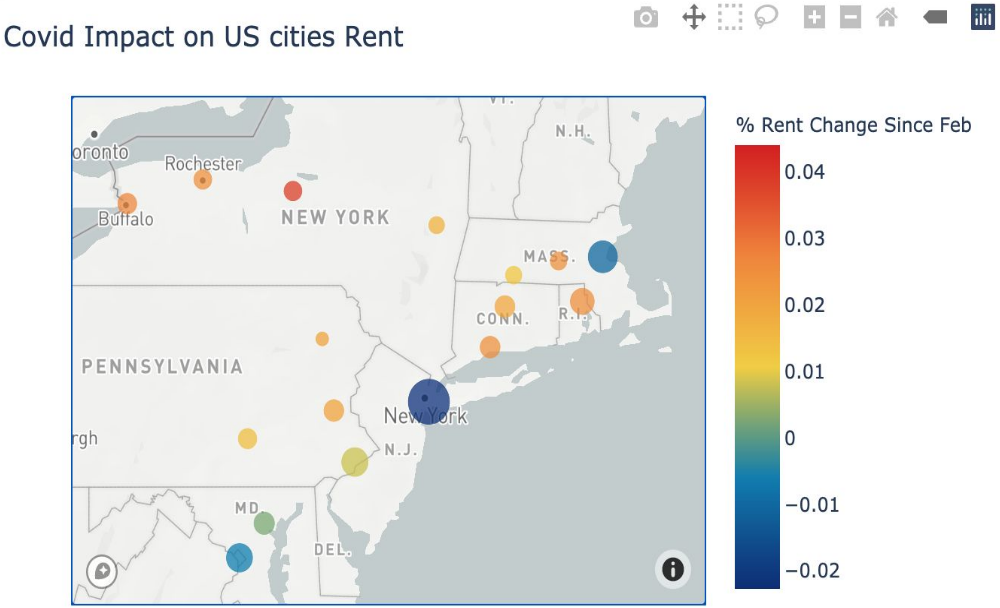
Contributors: Nika Chan, Matthew Musgrave, Carlos Tacchi, Alex Hall, and Eli Holden

The flight to the suburbs from COVID, aging millennials, and expensive costs of living have caused major changes in the housing market. This analysis examines the financial impacts, forecasts the future state of the housing market, and proposes where individuals and investors might find the best value for housing.

Covid Impact on Rental Prices

- COVID is causing a mass exodus from large metros such as NYC and San Francisco

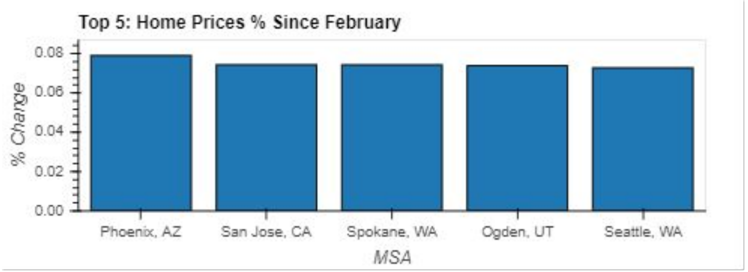


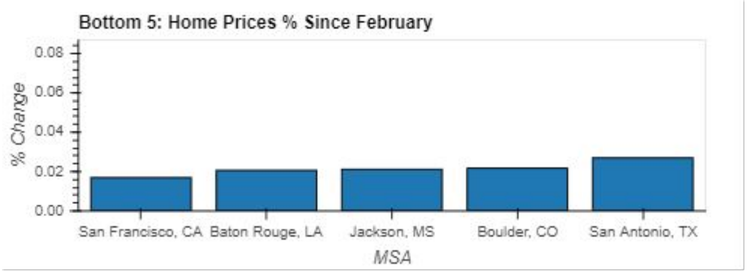


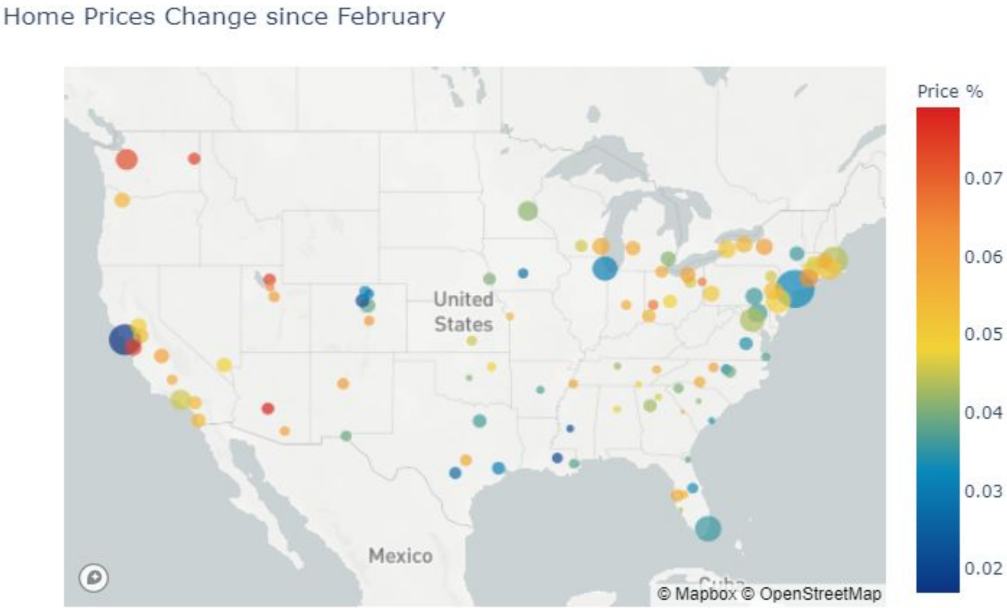
Sources: Zillow Rent Index (January 2020 - September 2020)

Nationwide Increase in Home Prices

- Lower price increase in larger cities (SF, Chicago, NY, Miami) but not only
- Largest price increase: less populated cities
- Possible explanations: low rates, more demand for larger houses, inflation hedge, no foreclosures







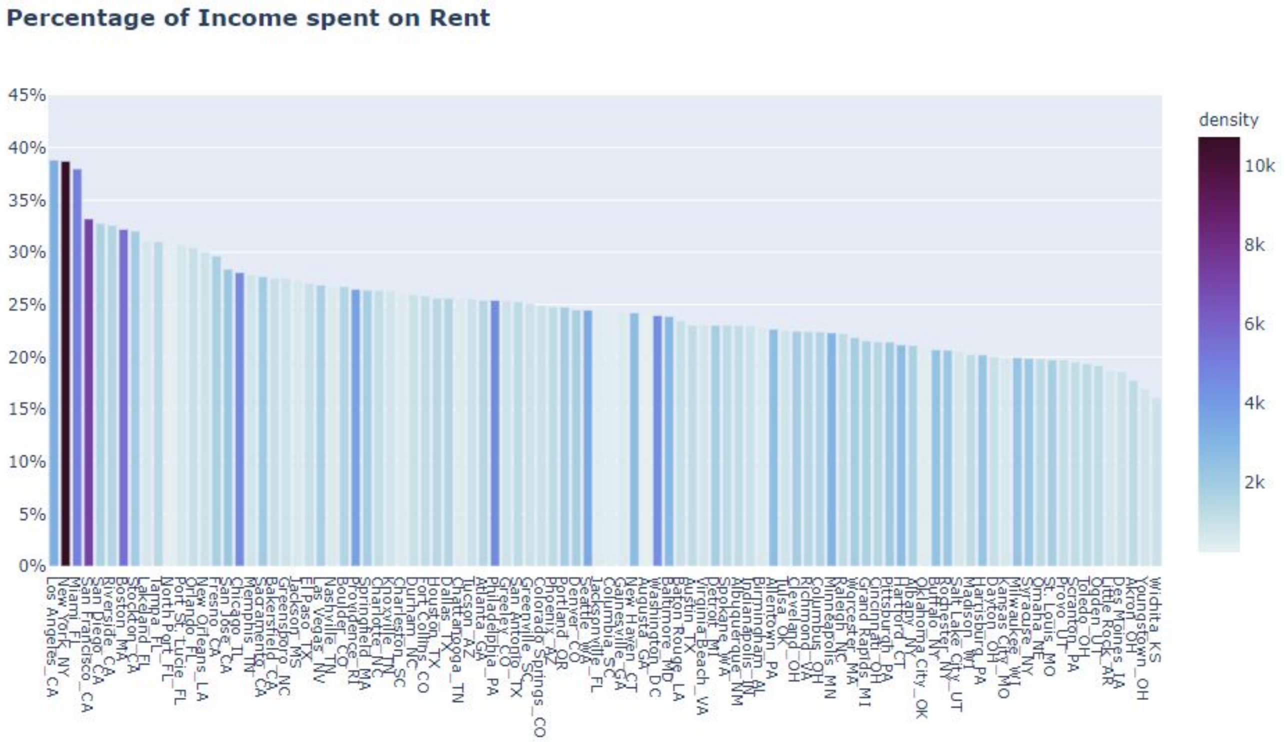
Sources: Zillow Home Prices (January 1996 - September 2020)

Housing Affordability

- Income and rent have a positive relationship
- Cities below the slope are relatively more affordable than the rest of the cities
- Residents in bigger cities tend to have higher rents and spend a greater proportion of income on rent

Sources: Zillow Rent Index (September 2020), BLS Median Income (2019)

Housing Affordability



<u>Most expensive cities</u>
1. LA 39%
2. NYC 38%
3. Miami 38%

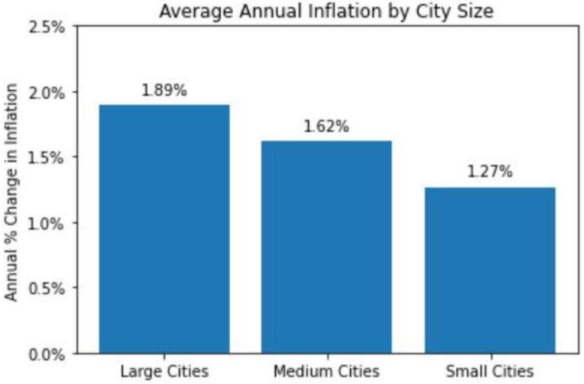
<u>Least expensive cities</u>
1. Wichita, Kansas 18%
2. Youngstown, Ohio 17%
3. Akron, Ohio 18%

<u>Affordable medium size cities</u>
-Chicago, IL 28%
-Providence, RI 26%
-Philadelphia, PA 25%
-Seattle, WA 24%
-New Haven, CT 24%
-Washington, DC 24%

Sources: Zillow Rent Index (September 2020), BLS Median Income (2019)

Housing Forecasts

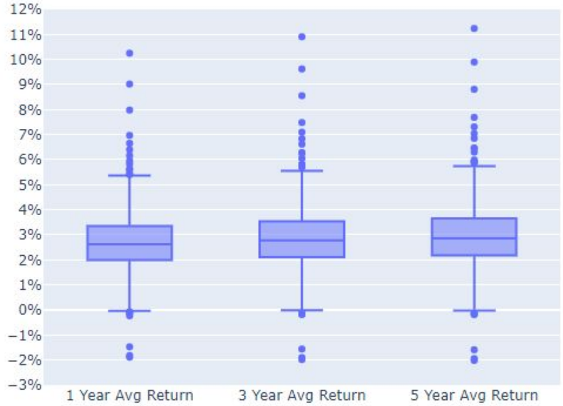
- Using over 20 years of housing price growth data we simulated the annualized future growth of MSAs to calculate an expected return over the next 5 years
- Average future house price growth exceeds recent inflation by 1-2%
  - Investors would also receive rental income (4-6% after expenses)
  - In comparison the stock market on average grows at approximately 7% annually



Average Returns After Removing Outliers
(small cities < 50K people)

	1 Year Avg Return	3 Year Avg Return	5 Year Avg Return	population
count	353.00	353.00	353.00	353
mean	2.84	3.01	3.09	521,585
std	1.07	1.13	1.16	1,359,038
min	-0.23	-0.20	-0.20	50,408
25%	2.19	2.31	2.39	77,609
50%	2.74	2.95	3.05	142,847
75%	3.48	3.67	3.77	378,732
max	6.22	6.65	6.87	18,713,220

Monte Carlo Forecast on Sale Prices



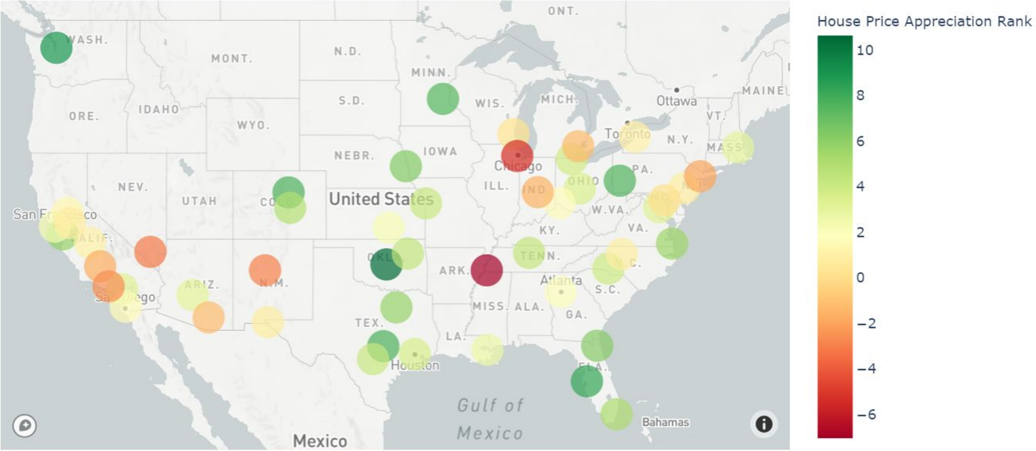
Sources: Zillow Home Prices (January 1996 - September 2020), BLS (CPI past 10 years)

Where should an investor or home buyer look?

- **Opportunities = High Total Return (Price + Yield) + High COVID Impact + High Population Growth + Affordable + Low Crime**
  - *Removes data with cities less than 50,000 people*
- Uses a rank system and customizable weights

MSA	House Price Appreciation Rank
Oklahoma City	10.6
Seattle	8.9
Tampa	8.3
Austin	7.7
Denver	7.7

House Price Appreciation Rank



API Data

- API’s used: Rapid API / APIDojo.com (Sales data from realtor.com)
- We constructed a tool to search for available properties based on zip code to find properties for sale within those desirable cities identified.

Example API Pull:

	City	Address	Postal Code	State Code	State	County	Lattitude	Longitude	Neighborhood
0	Lakewood	925 S Sheridan Blvd	80226	CO	Colorado	Jefferson	39.699291	-105.053638	South Alameda
0	Lakewood	633 S Depew St	80226	CO	Colorado	Jefferson	39.705008	-105.059019	South Alameda
0	Lakewood	954 S Miller St	80226	CO	Colorado	Jefferson	39.698715	-105.114126	Glennon Heights
0	Lakewood	7240 W Custer Ave Unit 304	80226	CO	Colorado	Jefferson	39.705949	-105.077061	South Alameda
0	Lakewood	888 S Johnson Ct	80226	CO	Colorado	Jefferson	39.700740	-105.108568	Belmar Park

Sources: Attom Data Solutions

Project Improvements

- Data frequency and granularity are not always the same
  - Using only the largest 50 MSA or more densely populated >50K people
  - Analyzing both sets of data for inconsistencies
- These factors may not be relevant to future house price appreciation
  - The next level to this analysis given more time would be a robust regression analysis
- House prices or Rental prices?
  - These variables are not always 100% correlated and respond at different rates to changes in the environment (example: buying homes can decrease rental prices as renters leave)

Data Exploration and Clean-Up

Raw Data and Exploration

Used data sets: house prices; rental prices; geolocation; income; Monte Carlo outputs
Not used: census data; crime
Key tasks:

- identifying key variables, and
- potential challenges when manipulating data: lack of unique identifier; multiple cities / zip codes in one cell; different cities in data base; same data from multiple years

RegionName R
United States
New York, NY
Los Angeles-Long Beach-Anaheim, CA

Aberdeen, WA Micro Area!!Families!!Estimate
19,664

Cleaning and Rearranging

2 Types of data frames:

- point in time
- Historical

Unique identifier: city, state (e.g. Los Angeles, CA)
Most used lines of code:

```
.str.split("-", n = 1, expand=True)
```

```
# Create unique identifier
house_price_df['MSA']=house_price_df['City']+', '+house_price_df['State']
```

```
# Merge - Puter as we want to give enough flexibility to drop data later
combined_df= pd.merge(sale_clean,mc_output_df,how='outer',on='MSA')
combined_df = pd.merge(combined_df,rent_clean,how='outer',on='MSA')
combined_df = pd.merge(lon_lat_density_df,combined_df,how='outer',on='MSA')
combined_df = pd.merge(median_income_df,combined_df,how='outer',on='MSA')
combined_df = pd.merge(census_clean,combined_df,how='outer',on='MSA')
combined_df = pd.merge(crime_df,combined_df,how='outer',on='MSA')
```

```
# Transpose dataframe
house_price_df=house_price_df.transpose()
# Drop the name that would look as the name of the index otherwise
house_price_df.columns.name = None
# set index
house_price_df.index.rename('Date',inplace=True)
```

```
# Two methonds
initial_population=census_df.iloc[0,:].apply(lambda x: float(x.split()[0].replace(',', '')))
initial_population=census_df.iloc[0,:].str.replace(',', '').astype(float)
```

Master Files

After merging the data, additional fields are generated (CAGR, ratios, etc.) and data frames are exported to a csv file

```
# Define a list of data frames, called df_list
df_list=[combined_df,rental_price_df,house_price_df]

# Define a list of data frames names, called df_name
df_name=['combined_df','rental_price_df','house_price_df']

# Create a for loop to iterate through each df and savethe files
for df, name in zip(df_list,df_name):
    file_name=name +'.csv'
    output_file = Path(f"../Data/Clean/{file_name}")
    df.to_csv(f"{output_file}")
```

Questions?

Appendix

Data Sources

- Zillow, U.S. Census Bureau, U.S. Bureau of Labor Statistics, U.S. Geological Survey, FBI government data
- RapidAPI (free) provides real-time property sales data and can consume various criteria in order to filter down available properties that are customized to individual buyers.
- The datasets vary in range and depth but are standardized around metropolitan statistical areas (MSAs) and typically monthly frequency
  - For example, housing price data has over 900 MSAs and goes through 1996
- For Zillow data there are two datasets:
  - Housing Prices
  - Zillow Rent Index (ZRI): is a dollar-valued index intended to capture typical market rent for a given segment (IE, multifamily or single-family units) and/or geography (IE for a given ZIP code, city, county, state or metro) by leveraging Rent Zestimates. (https://www.zillow.com/research/zillow-rent-index-methodology-2393)

New Python Library DataComPy (Helps with data cleaning and exploring)

```
DataComPy Comparison
--------------------

DataFrame Summary
-----------------

          DataFrame  Columns  Rows
0         Rental_DF        1   105
1  Median Income DF        1   518

Column Summary
--------------

Number of columns in common: 1
Number of columns in Rental_DF but not in Median Income DF: 0
Number of columns in Median Income DF but not in Rental_DF: 0

Row Summary
-----------

Matched on: primary msa
Any duplicates on match values: No
Absolute Tolerance: 0
Relative Tolerance: 0
Number of rows in common: 99
Number of rows in Rental_DF but not in Median Income DF: 6
Number of rows in Median Income DF but not in Rental_DF: 419

Number of rows with some compared columns unequal: 0
Number of rows with all compared columns equal: 99
```

```
Column Comparison
-----------------

Number of columns compared with some values unequal: 0
Number of columns compared with all values equal: 1
Total number of values which compare unequal: 0

Sample Rows Only in Rental_DF (First 10 Columns)
------------------------------------------------

         primary msa
88  Daytona Beach_FL
65        Ventura_CA
82     Fort Myers_FL
57       Stamford_CT
95      Melbourne_FL
42     Louisville_KY

Sample Rows Only in Median Income DF (First 10 Columns)
-------------------------------------------------------

         primary msa
258        Helena_MT
512   Williamsport_PA
147     Bluefield_WV
306     Lake City_FL
211          Enid_OK
399    Pine Bluff_AR
409   Poughkeepsie_NY
226      Florence_SC
453   Sevierville_TN
492         Ukiah_CA
```

Data Exploration a.k.a. Charts that Did Not Make It