GDP and Real Estate Trends

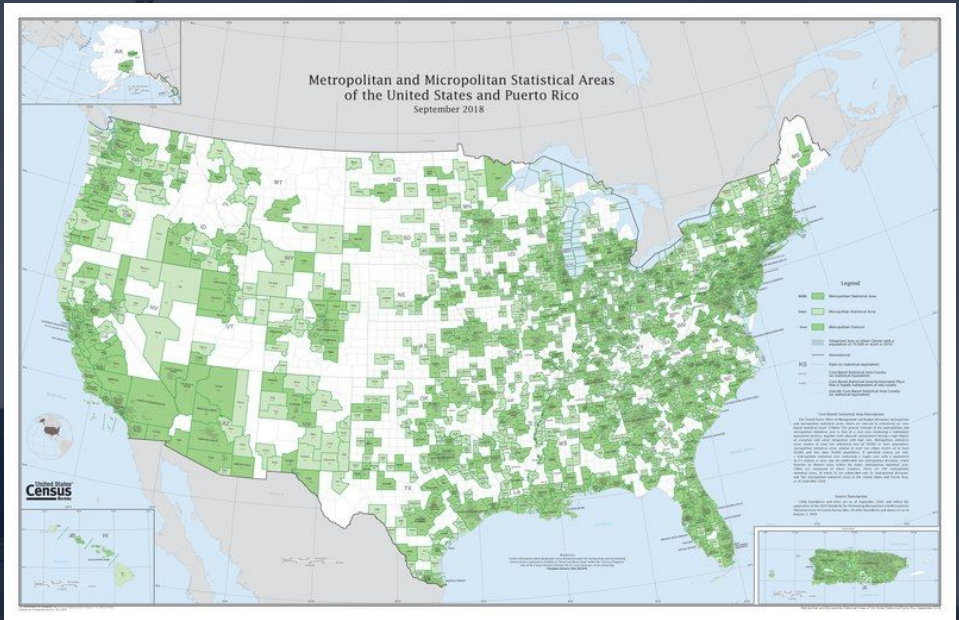
Metropolitan Statistical Areas

Team Members:
Anmol Madan
Davi Aragao
Shenglin Ye

EXECUTIVE SUMMARY

Real Estate pricing and economic conditions are vastly different depending on location. The success of a business or the livelihood of a family can be severely affected by trends in these measures.

In order to quickly assess potential destinations we have created a system to load, store and extract meaningful insights from various data sources, that can be used by individuals, governments and businesses.



Business Use Case

Individuals

Narrow down potential moving destinations that feature low cost homes or rentals and high potential earning opportunities.

Businesses

Finding regions to expand one's business based on high measures of productivity in your industry and low personnel cost.

Government

State and Federal Government can more effectively direct resources to cities experiencing economic decline.

What is an MSA?

Office of Management and Budget

- Metropolitan statistical areas (MSA) are delineated by the U.S. OMB as having at least one urbanized area with a minimum population of 50,000.
- MSA is the formal definition of a region that consists of a city and surrounding communities that are linked by social and economic factors.
- MSAs serve to group counties and cities into specific geographic areas for population censuses and compilations of related statistical data.

The U.S. has a total of 384 MSAs

Source: https://www.investopedia.com/terms/m/msa.asp

Data

Data Sources

CAGDP1 Gross Domestic Product (GDP) summary by metropolitan area

- Current-dollar Gross Domestic Product (Thousands of dollars)

CAINC30 Economic Profile:

- Total employment (Number of jobs)
- Population (Number of persons)
- Personal income (Thousands of dollars)

MARPP - Regional Price Parities by MSA

- RPPs: Services: Rents (Index)

Zillow Home Value Index (ZHVI): A smoothed, seasonally adjusted measure of the typical home value and market changes

- Composite
- One Bedroom
- Two Bedroom
- Three Bedroom
- Four Bedroom
- Five+ Bedroom
- Condo

Zillow Observed Rent Index (ZORI): A smoothed measure of the typical observed market rate rent across a given region.

Zillow For-Sale Inventory

Data Sources

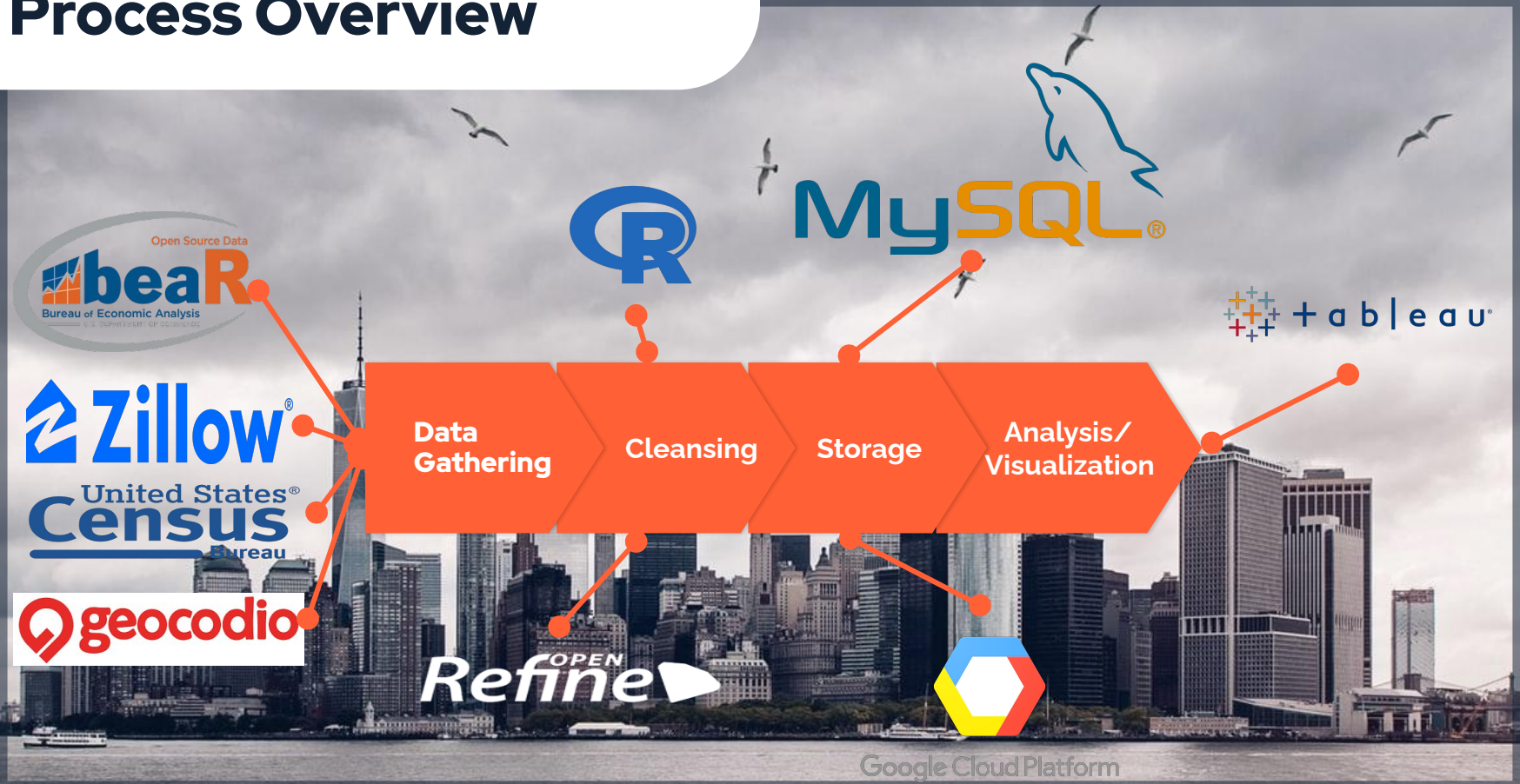
Metropolitan Area Census Data

- Total Area
- Land Area
- Water Area

Geocodio API used to source location coordinates based on MSA name

- Longitude
- Latitude

Process Overview



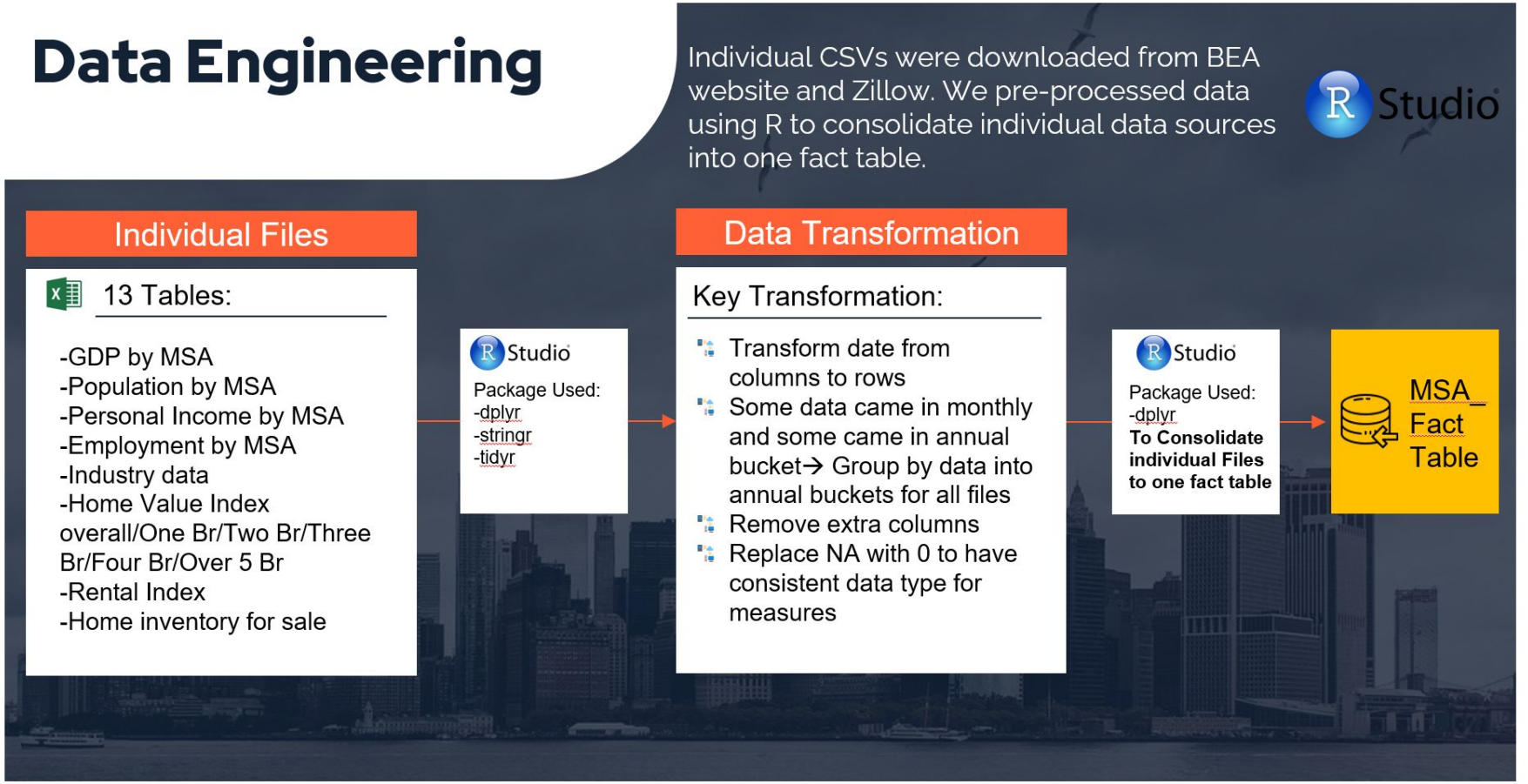
Data Cleaning

Between all of the data sets the MSA names varied slightly and made joining the tables difficult, for example, Washington DC was labeled the following ways given the data source:

To solve this we chose the BEA labels to be our default and modified the other labels using OpenRefine

Data Engineering

Individual CSVs were downloaded from BEA website and Zillow. We pre-processed data using R to consolidate individual data sources into one fact table.



Data Transformation

We chose a yearly grain for our fact table given trends in economic production and real estate pricing develop slowly over time.

BEA Data was compiled yearly while Zillow data was compiled monthly. Using OpenRefine we averaged the monthly data to return yearly values.

DDL Script Snapshot

Foreign Keys were created to join two Fact tables to two shared Dimension tables.

Integer variables were used in most cases except for GDP and Personal Income which got very large at the US level.

Home value indices were converted into Measures in the Fact Table instead of creating a new table.

```
CREATE TABLE IF NOT EXISTS `msa_snowflake`.`fact_msa` (
  `fact_id` INT(10) NOT NULL AUTO_INCREMENT,
  `msa_id` INT(8),
  `date_id` INT(8),
  `population` INT(10),
  `gdp` BIGINT,
  `employment` INT (10),
  `personal_income` BIGINT,
  `regional_price_parity` INT(10),
  `zillow_home_value_index` INT(10),
  `zillow_rent_index` INT(10),
  `zillow_for_sale_inventory` INT(10),
  `zillow_one_br_home_value_index` INT(10),
  `zillow_two_br_home_value_index` INT(10),
  `zillow_three_br_home_value_index` INT(10),
  `zillow_four_br_home_value_index` INT(10),
  `zillow_five_br_home_value_index` INT(10),
  `zillow_condo_home_value_index` INT(10),
  PRIMARY KEY (`fact_id`),
  CONSTRAINT `date_id`
    FOREIGN KEY (`date_id`)
    REFERENCES `msa_snowflake`.`dim_year` (`date_id`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT `msa_id`
    FOREIGN KEY (`msa_id`)
    REFERENCES `msa_snowflake`.`dim_location` (`msa_id`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION
    )
ENGINE = InnoDB
```

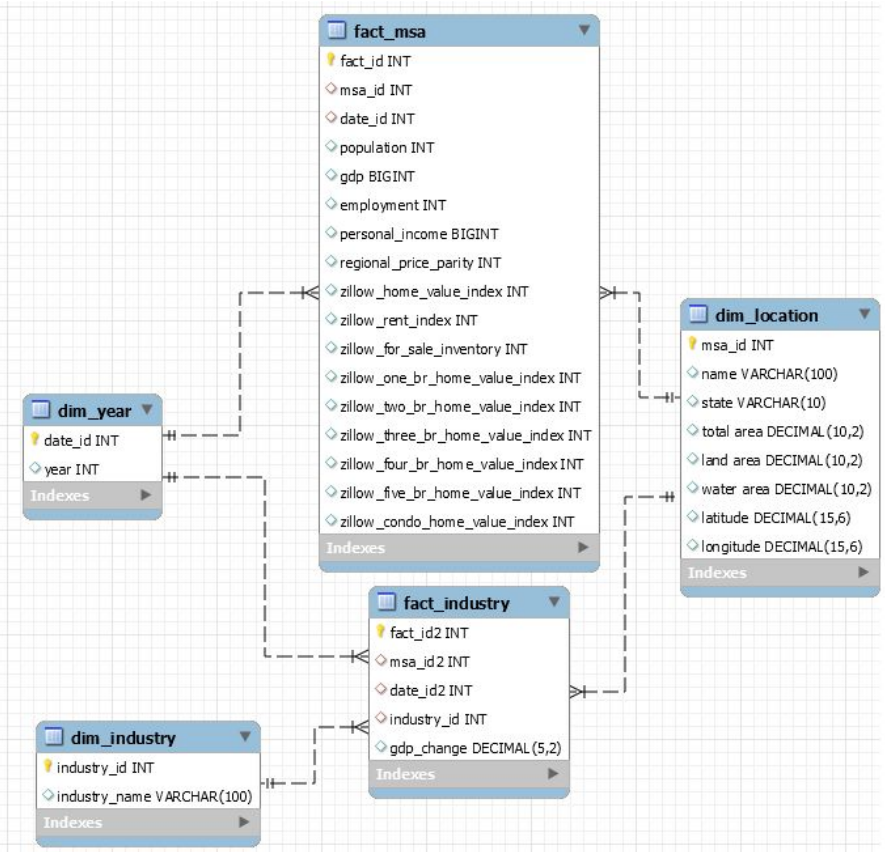
DML Script Snapshot

Data insertion was done through Values as uploading large csv files was slower and required more effort to make changes.

We were able to fix data errors quickly through shared csv files over google drive and using a values based CONCATENATE function in EXCEL.

```
-- -----------------------------------------------------
-- Populate Fact MSA
-- -----------------------------------------------------

INSERT INTO `fact_msa`
(`msa_id` , `date_id` , `population` ,`gdp` , `employment` , `personal_income` , `regional_price_parity`,`
`zillow_rent_index` , `zillow_for_sale_inventory` , `zillow_one_br_home_value_index`,`zillow_two_br_home_v
`zillow_three_br_home_value_index`, `zillow_four_br_home_value_index`, `zillow_five_br_home_value_index`,
VALUES
(1,6,158917,3638709,89511,3753367,0,0,0,0,0,0,0,0),
(1,7,159012,3746156,90338,3905004,0,0,0,0,0,0,0,0),
(1,8,158810,3929394,91132,4137066,0,0,0,0,0,0,0,0),
(1,9,160156,4174103,91998,4325061,0,0,0,0,0,0,0,0),
(1,10,160761,4423342,92937,4552568,0,88563,0,0,55728.1666666667,60648.9166666667,97084.4166666667,161206.2
(1,11,161389,4991087,94097,4850810,0,95465,0,0,57372.3333333333,63998.0833333333,105042.916666667,173908.8
(1,12,162023,5379170,96484,5119572,0,101096,0,0,58951.5833333333,68091.6666666667,112157.333333333,184397.
(1,13,162508,5666044,99264,5632799,73.4,102650,0,0,61229,70270.0833333333,114059.75,186867.333333333,26189
(1,14,163888,5541862,97269,5415851,72.1,102219,0,0,63824.4166666667,70352.3333333333,112906.5,187313.75,26
(1,15,165585,5839625,96809,5705250,72.8,101428,0,0,61649.3333333333,68438.75,112023.666666667,187151.08333
(1,16,166634,5996498,98193,6052939,74.6,97955,0,0,56542,63129.8333333333,109453.75,185446.25,264853.833333
```

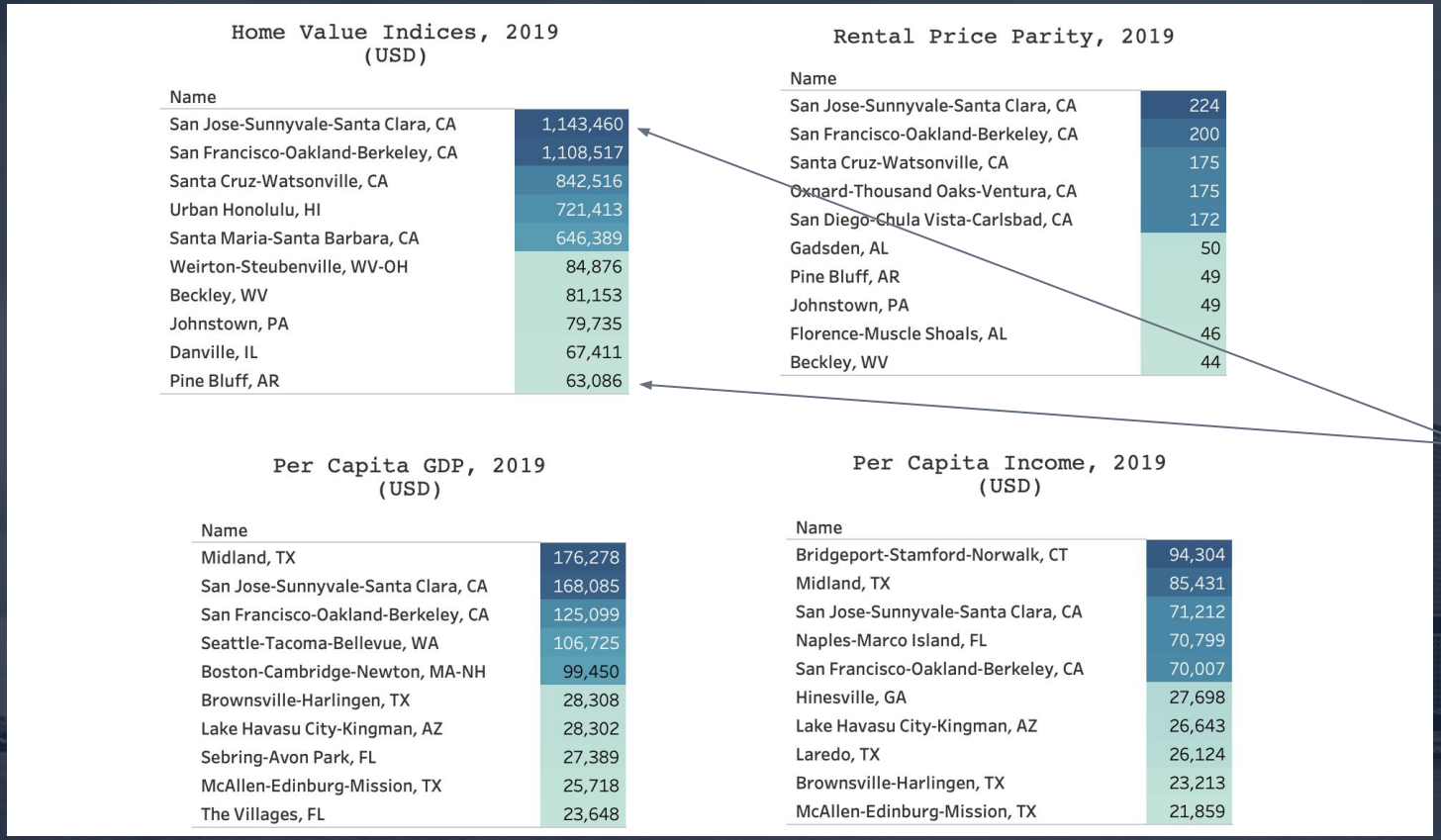


Dimensional Model

Our final EER model was modified to incorporate two Fact tables as data by industry used a different measure than the main dataset. We joined the Fact tables to common Dimension tables and created a new Dimension table for future expansion on industries.

Descriptive Analytics

Range of Home Values, Rental Parity, GDP and Income



Home Value Indices, 2019 (USD)

Name	
San Jose-Sunnyvale-Santa Clara, CA	1,143,460
San Francisco-Oakland-Berkeley, CA	1,108,517
Santa Cruz-Watsonville, CA	842,516
Urban Honolulu, HI	721,413
Santa Maria-Santa Barbara, CA	646,389
Weirton-Steubenville, WV-OH	84,876
Beckley, WV	81,153
Johnstown, PA	79,735
Danville, IL	67,411
Pine Bluff, AR	63,086

Rental Price Parity, 2019

Name	
San Jose-Sunnyvale-Santa Clara, CA	224
San Francisco-Oakland-Berkeley, CA	200
Santa Cruz-Watsonville, CA	175
Oxnard-Thousand Oaks-Ventura, CA	175
San Diego-Chula Vista-Carlsbad, CA	172
Gadsden, AL	50
Pine Bluff, AR	49
Johnstown, PA	49
Florence-Muscle Shoals, AL	46
Beckley, WV	44

Per Capita GDP, 2019 (USD)

Name	
Midland, TX	176,278
San Jose-Sunnyvale-Santa Clara, CA	168,085
San Francisco-Oakland-Berkeley, CA	125,099
Seattle-Tacoma-Bellevue, WA	106,725
Boston-Cambridge-Newton, MA-NH	99,450
Brownsville-Harlingen, TX	28,308
Lake Havasu City-Kingman, AZ	28,302
Sebring-Avon Park, FL	27,389
McAllen-Edinburg-Mission, TX	25,718
The Villages, FL	23,648

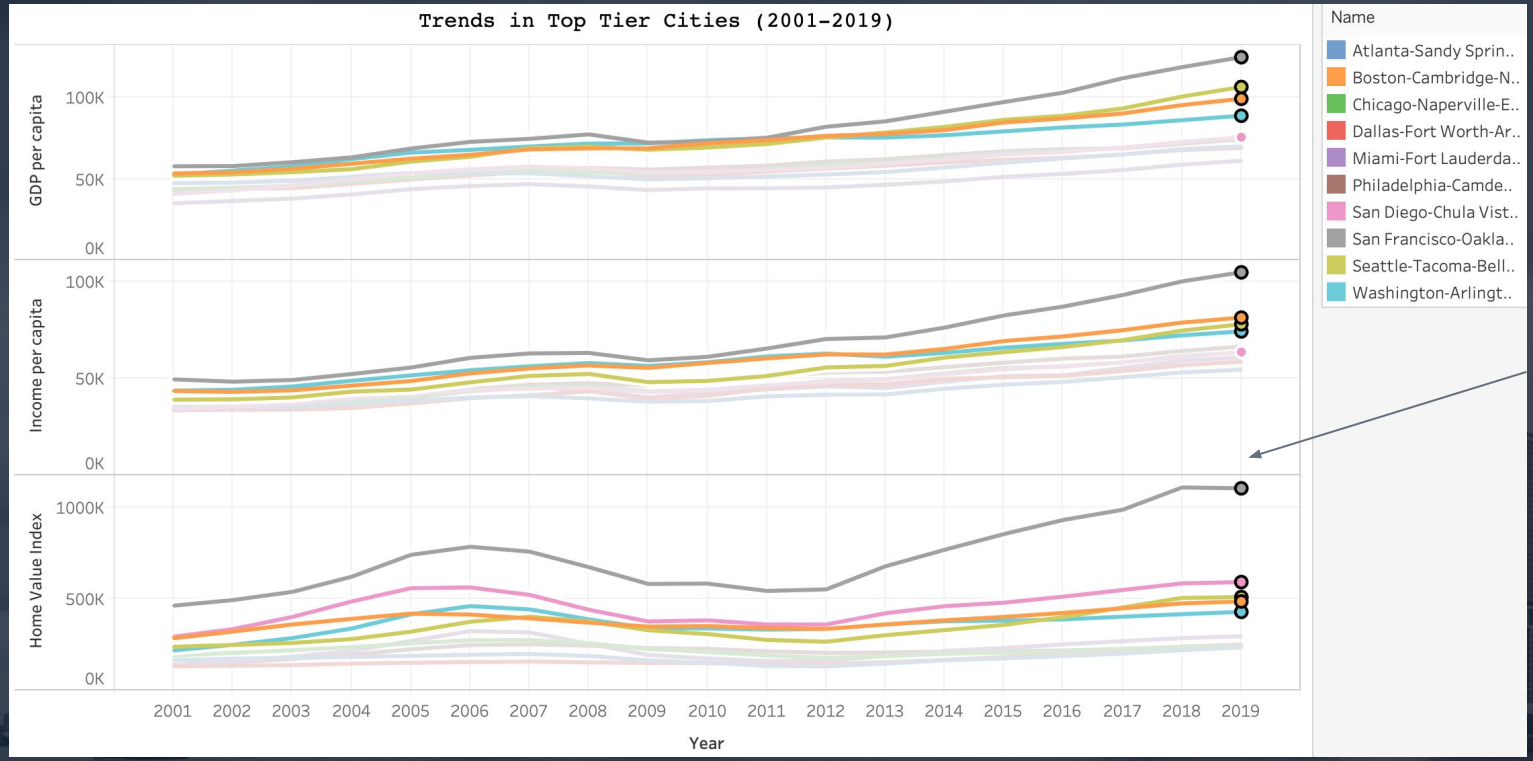
Per Capita Income, 2019 (USD)

Name	
Bridgeport-Stamford-Norwalk, CT	94,304
Midland, TX	85,431
San Jose-Sunnyvale-Santa Clara, CA	71,212
Naples-Marco Island, FL	70,799
San Francisco-Oakland-Berkeley, CA	70,007
Hinesville, GA	27,698
Lake Havasu City-Kingman, AZ	26,643
Laredo, TX	26,124
Brownsville-Harlingen, TX	23,213
McAllen-Edinburg-Mission, TX	21,859

This dashboard shows the range of home values, rental price parity, per capita GDP and Income of MSAs.

San Jose, CA had Home Values 18x over Pine Bluff, AR in 2019.

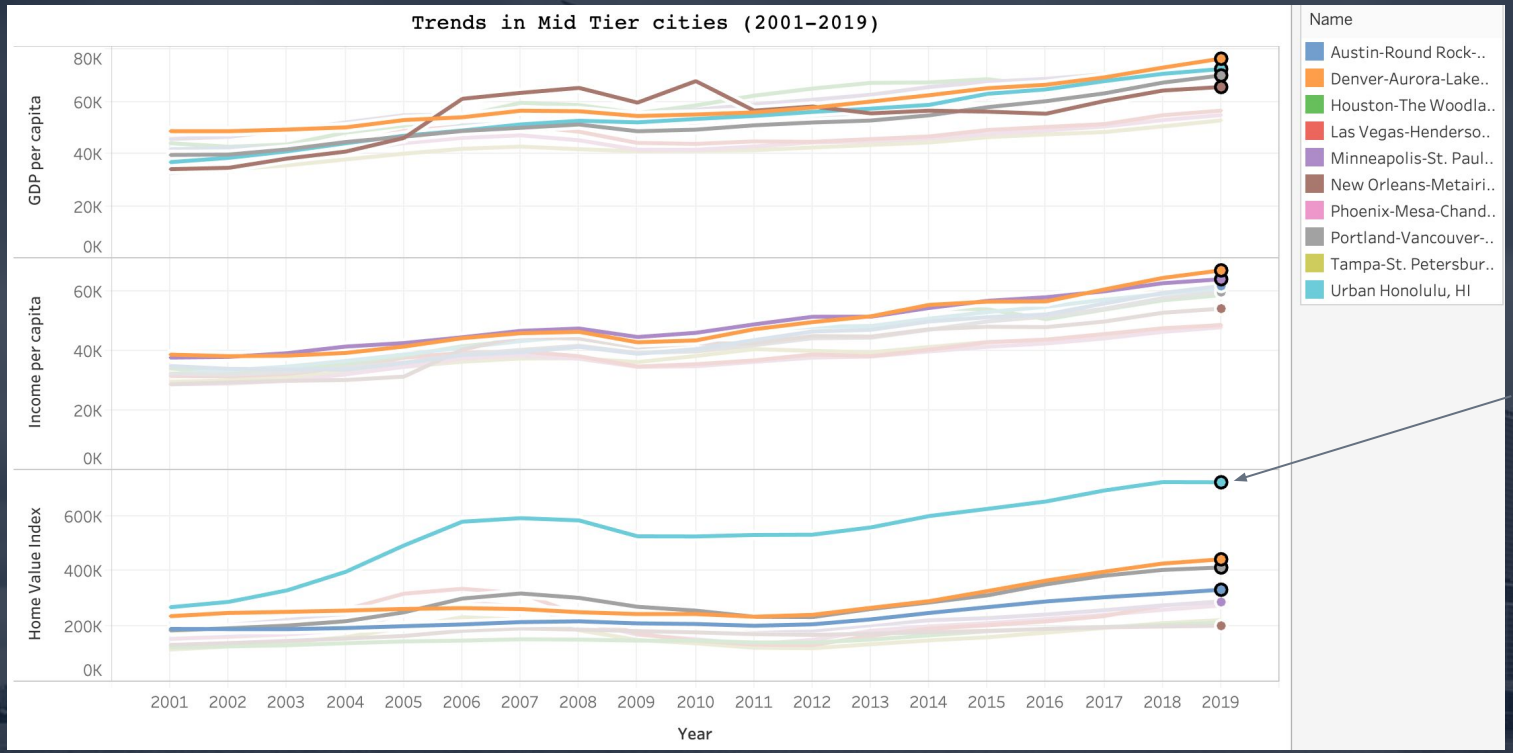
Divergence in GDP, Income and Home Values by top-tier MSAs



This shows the widening gap in top tier MSAs in per capita GDP and Personal Income, and Home Values.

San Francisco pulled ahead in all 3 measures, while Seattle, Boston and San Diego are in the middle.

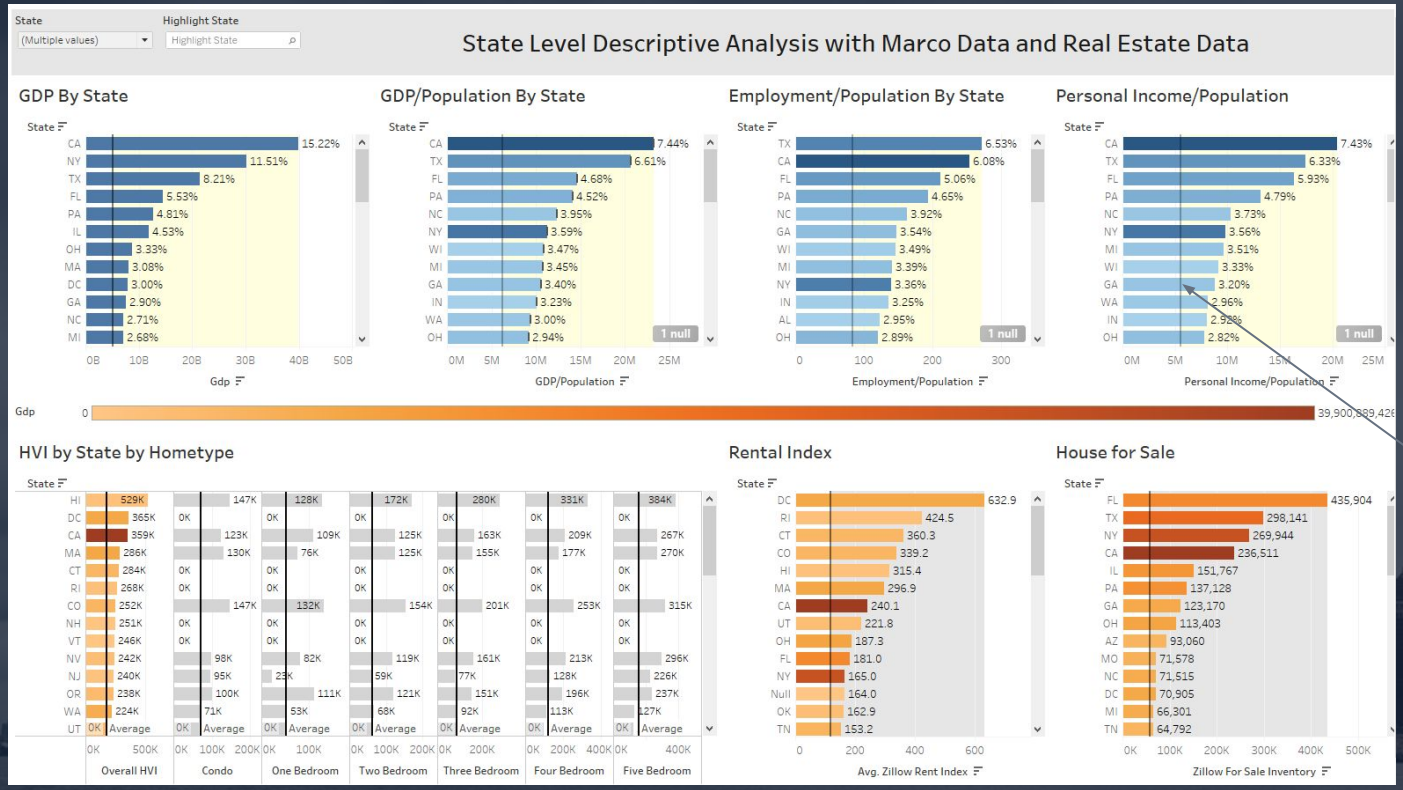
Divergence in GDP, Income and Home Values by mid-tier MSAs



Mid-tier MSAs have stayed close in per capita GDP and Personal Income, but diverged in Home Values

Honolulu pulled ahead on home values, while Denver and Portland are in mid-range.

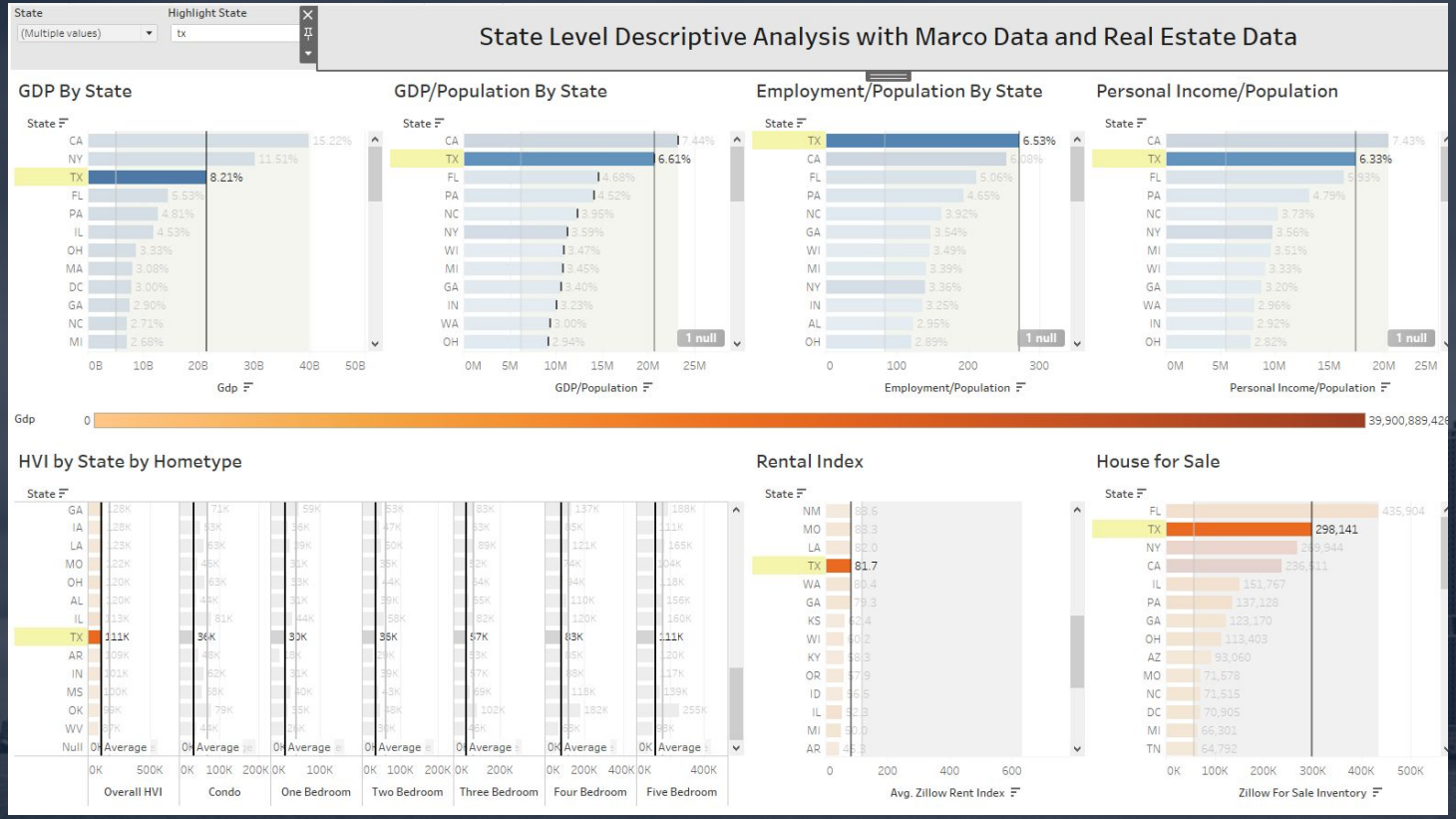
State Level Overview- To get overview of marco related data and real estate related data for each state and where state stands among all states.



Purpose of this Dashboard is to see if states have consistent ranking for both macro related measures and real estate measures to derive insights.

Note: Vertical line indicates national average value. States crossing the line exceeds average

State Level Findings- Through interactive dashboard, we were able to identify states that have different rankings across different measures. **Example 1: State TX (highlighted in dashboard)**

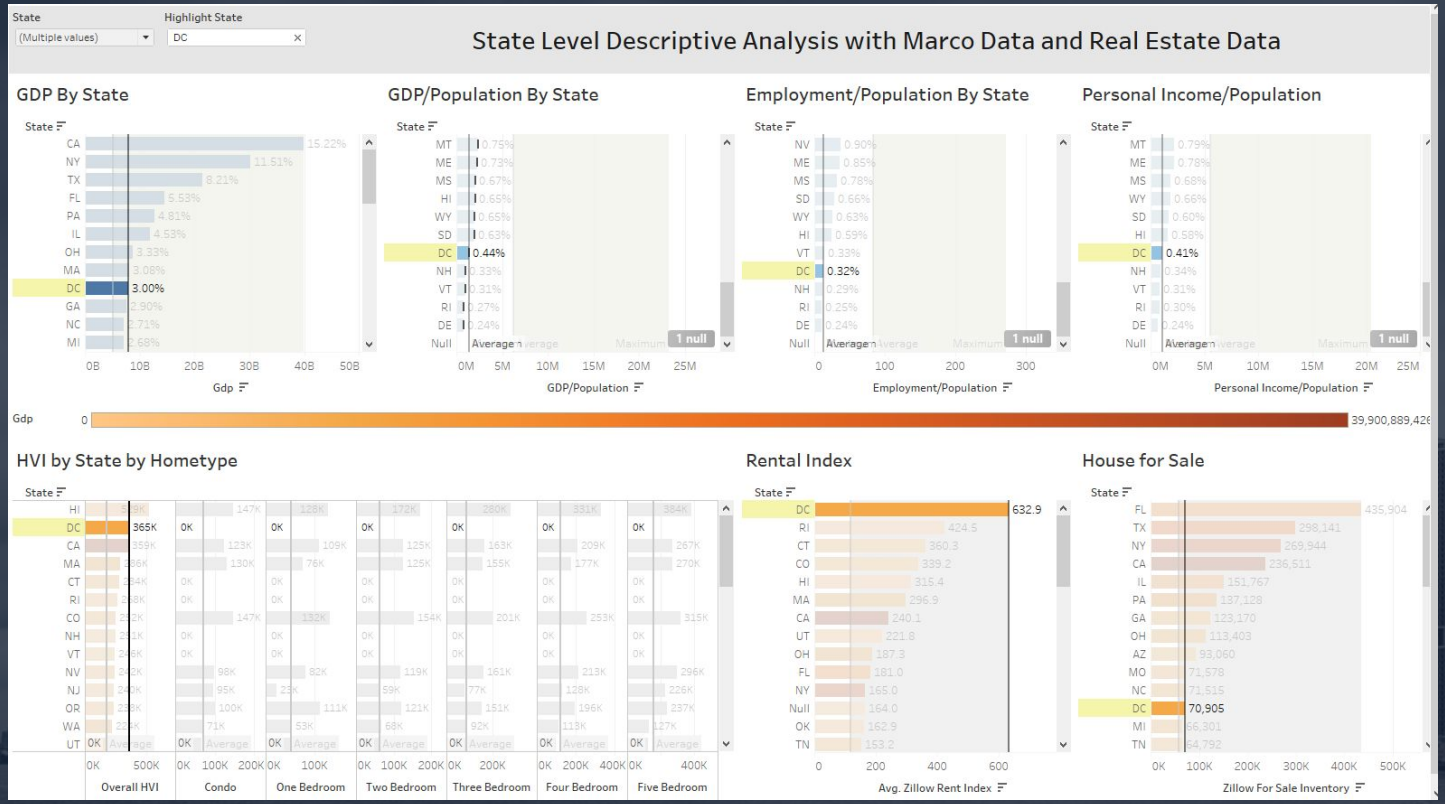


Insights:

Texas (TX) Ranked top in terms of macro measures: GDP, personal income per capita and Employment. However, TX home value index and Rental index ranked very low and merely reach national average level. At same time, TX has good amount of house for sale.

This indicates that TX can be a affordable place for living with high employment, income but low housing cost.

State Level Findings- Through interactive dashboard, we were able to identify states that have different rankings across different measures. **Example 2: State DC (highlighted in dashboard)**



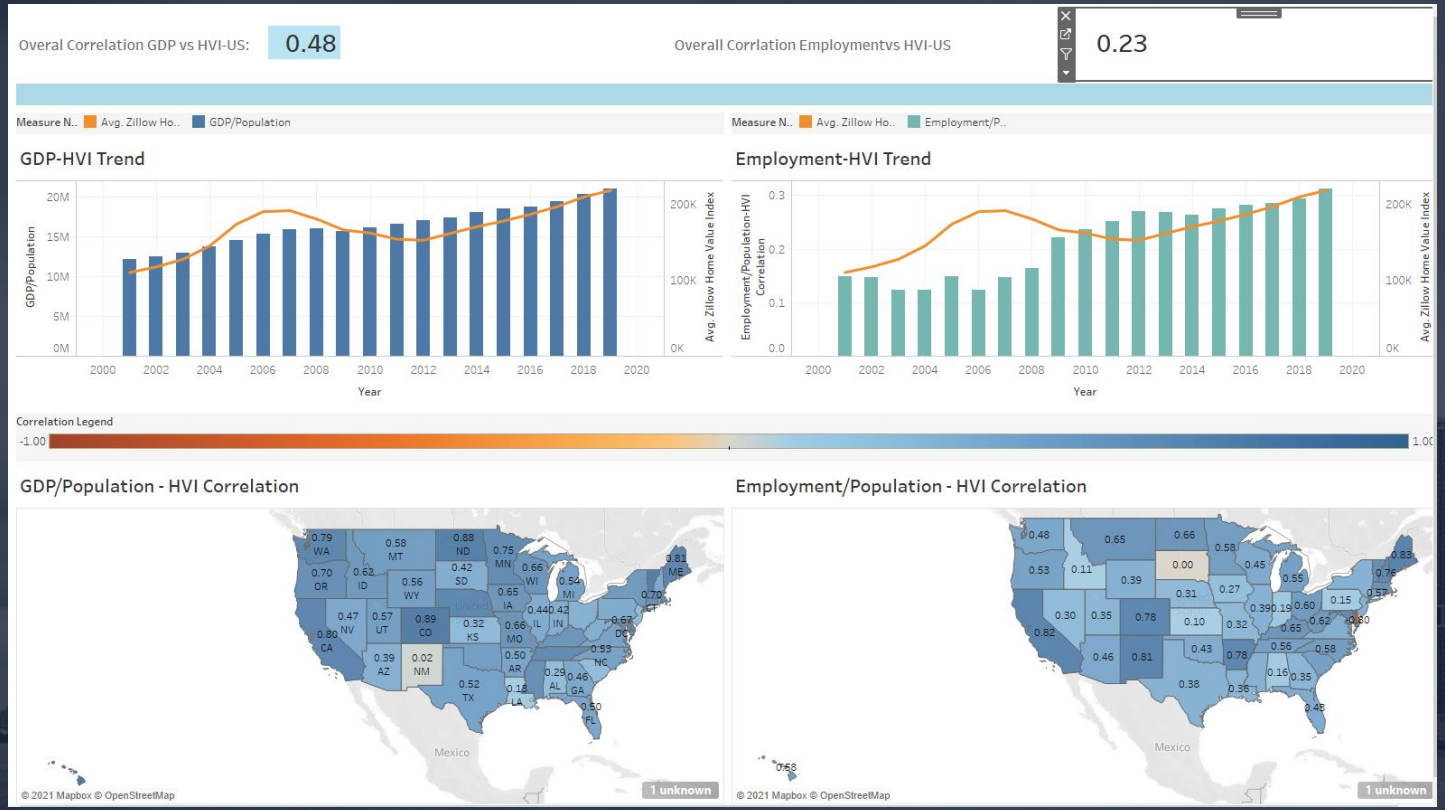
Insights:

On the contrary, DC ranked very low in macro economic measure with only national average level. However, DC ranked top in home value and rental index.

This indicates that DC might not be a first choice for living due to high housing cost and low employment as well as low income.

*Note: DC is unique since it only has one MSA and does not have low cost MSAs to average out the number.

Correlation between Macro Data and Real estate data- Dashboard to analyze correlation between GDP and HVI as well as Employment and HVI.



Purpose of this Dashboard is to see if macro measures move in the same direction as real estate measures. If strong correlation exists, we can predict future home value movement using macro economic data.

Key findings:

-GDP moves in same direction with home value. Same correlation exists across all states

-Employment vs HVI has weaker positive correlation compared with GDP vs HVI.

Business Analytics

Growth MSAs with Inexpensive Real Estate

```
SELECT a.*, CONCAT('$', FORMAT(b.zillow_home_value_index,0)) AS 'Home Value Index' FROM
(SELECT
        name AS MSA, industry_name as Industry, avg(gdp_change) AS 'Average Change in GDP'
FROM fact_industry f
        JOIN dim_industry i ON f.industry_id = i.industry_id
                JOIN dim_location l ON l.msa_id = f.msa_id2
                        JOIN dim_year y on y.date_id = f.date_id2
WHERE industry_name IN
                ('Natural resources and mining','Trade','Transportation and utilities','Government and government enterprises','Manufacturing','Information'
        , 'Professional and business services','Arts, entertainment, recreation, accommodation, and food services', 'Finance and insurance', 'Real estate and
rental and leasing'
        , 'Educational services', 'Health care and social assistance' )
            AND year = 2019
GROUP BY 1 , 2
ORDER BY 3 DESC
LIMIT 100) AS a
JOIN
(SELECT name,  zillow_home_value_index
FROM fact_msa f
        JOIN dim_location l ON f.msa_id = l.msa_id
                JOIN dim_year y on y.date_id = f.date_id
WHERE year = 2019
             AND zillow_home_value_index != 0
ORDER BY zillow_home_value_index) AS b
on a.MSA = b.name
ORDER BY 3 DESC, 4;
```

MSA	Industry	Average Change in GDP	Home Value Index
Midland, TX	Natural resources and mining	17.450000	$269,674
Homosassa Springs, FL	Transportation and utilities	7.550000	$160,286
Barnstable Town, MA	Real estate and rental and leasing	6.590000	$439,208
Ocean City, NJ	Real estate and rental and leasing	6.050000	$389,060
Lake Charles, LA	Manufacturing	5.260000	$174,253
Sebring-Avon Park, FL	Real estate and rental and leasing	5.060000	$138,296
Homosassa Springs, FL	Real estate and rental and leasing	4.000000	$160,286
Punta Gorda, FL	Real estate and rental and leasing	3.950000	$217,975
San Angelo, TX	Natural resources and mining	3.740000	$148,225
Boulder, CO	Professional and business services	3.650000	$566,144
Duluth, MN-WI	Natural resources and mining	3.640000	$167,170
Redding, CA	Real estate and rental and leasing	3.370000	$276,741
San Jose-Sunnyvale-S…	Information	3.170000	$1,143,460
Santa Cruz-Watsonvill…	Professional and business services	3.110000	$842,516
Evansville, IN-KY	Manufacturing	3.100000	$140,405
Utica-Rome, NY	Real estate and rental and leasing	2.910000	$135,069
Madera, CA	Real estate and rental and leasing	2.860000	$274,649
Binghamton, NY	Real estate and rental and leasing	2.780000	$120,934
Sebastian-Vero Beach…	Real estate and rental and leasing	2.730000	$227,265
Provo-Orem, UT	Professional and business services	2.710000	$350,428

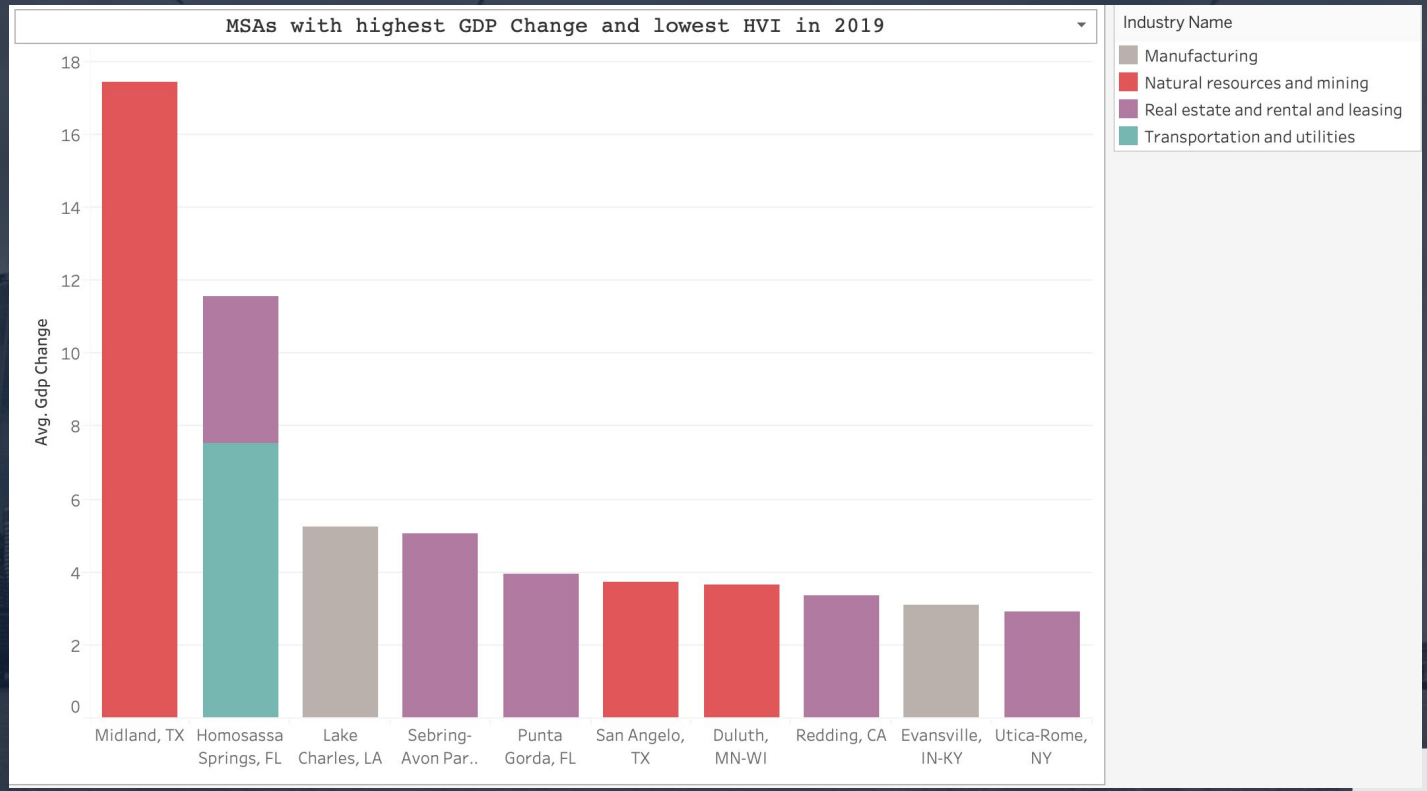
Growth MSAs with Inexpensive Real Estate

Tableau:

2019 growth in GDP by industry 2.9 - 20%

AND

2019 Home Value Index < $220,000

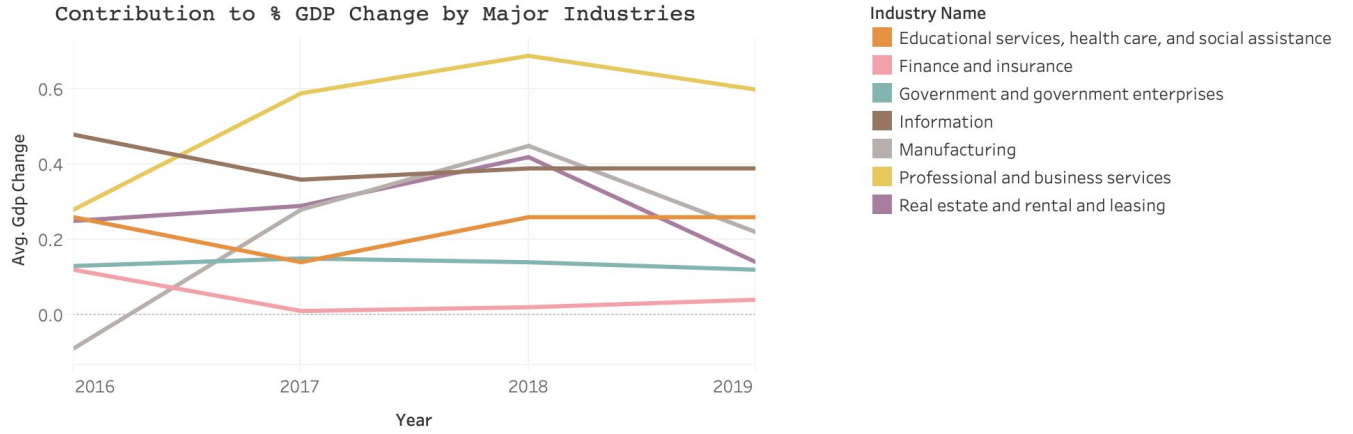


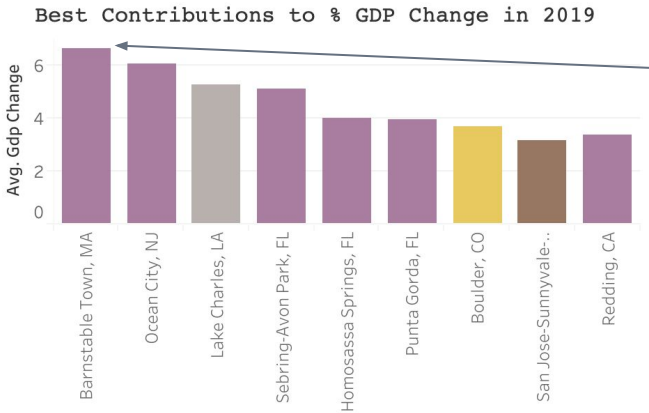
Highest Differential between GDP and HVI Percentile

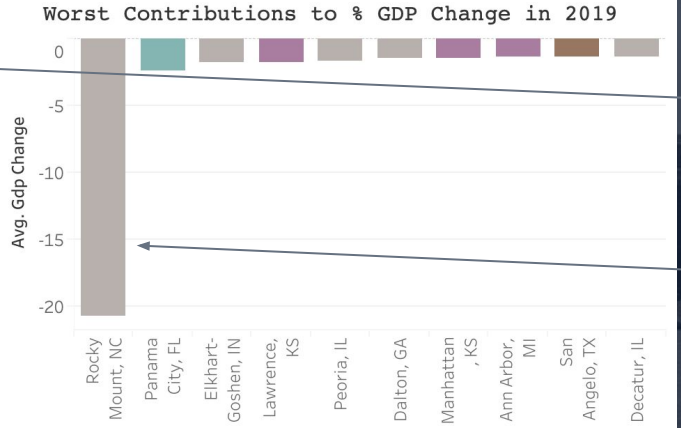
```
SELECT
    s.msa_id,
    s.name,
    f.zillow_home_value_index as HVI,
    f.gdp as GDP,
    ROUND(percent_rank() OVER (ORDER BY f.zillow_home_value_index), 2) as PercRankHVI,
    ROUND(percent_rank() OVER (ORDER BY f.gdp), 2) as PercRankGDP,
    ROUND(percent_rank() OVER (ORDER BY f.gdp) - percent_rank() OVER (ORDER BY  f.zillow_home_value_index), 2) as Diff
FROM
    fact_msa AS f
        INNER JOIN
    dim_location AS s ON f.msa_id = s.msa_id
                INNER JOIN
        dim_year AS y On y.date_id = f.date_id
WHERE y.year = 2019 and f.zillow_home_value_index>0
ORDER BY Diff DESC LIMIT 10;
```

msa_id	name	HVI	GDP	PercRankHVI	PercRankGDP	Diff
347	Toledo, OH	119688	38379623	0.11	0.79	0.68
92	Dayton-Kettering, OH	130013	45689081	0.16	0.82	0.66
383	Youngstown-Warren-Boardman, OH-PA	97258	22276185	0.03	0.67	0.64
316	Scranton--Wilkes-Barre, PA	117535	29808582	0.1	0.73	0.63
269	Peoria, IL	113059	23669529	0.08	0.69	0.61
351	Tulsa, OK	145255	57818637	0.25	0.85	0.6
322	Shreveport-Bossier City, LA	112524	22012841	0.07	0.66	0.59
222	McAllen-Edinburg-Mission, TX	115335	22341504	0.08	0.67	0.59
107	El Paso, TX	133721	33235319	0.17	0.77	0.59
224	Memphis, TN-MS-AR	150881	78866609	0.28	0.87	0.59

MSAs with best and worst GDP changes 2018-19







Education and Information have been contributing more to increase in GDP in the last four years.

Real Estate is a contributor in MSAs like Barnstable Town MA, while Manufacturing led to GDP decline in MSAs like Rocky Mount NC.

Troubled MSAs

```
SELECT
    x.msa_id,
    l.name,
    ((y.gdp - x.gdp) / x.gdp) * 100 AS PCT_Change_GDP,
    ((y.employment - x.employment) / x.employment) * 100 AS PCT_Change_Employment,
    ((y.personal_income - x.personal_income) / x.personal_income) * 100 AS PCT_Change_PersInc
FROM
    fact_msa x JOIN
    fact_msa y ON y.msa_id = x.msa_id JOIN
    dim_location l ON x.msa_id = l.msa_id AND
    y.date_id = '24'
WHERE
    x.date_id = '23'
HAVING PCT_Change_GDP <0
OR PCT_Change_Employment<0
OR PCT_Change_PersInc<0;
```

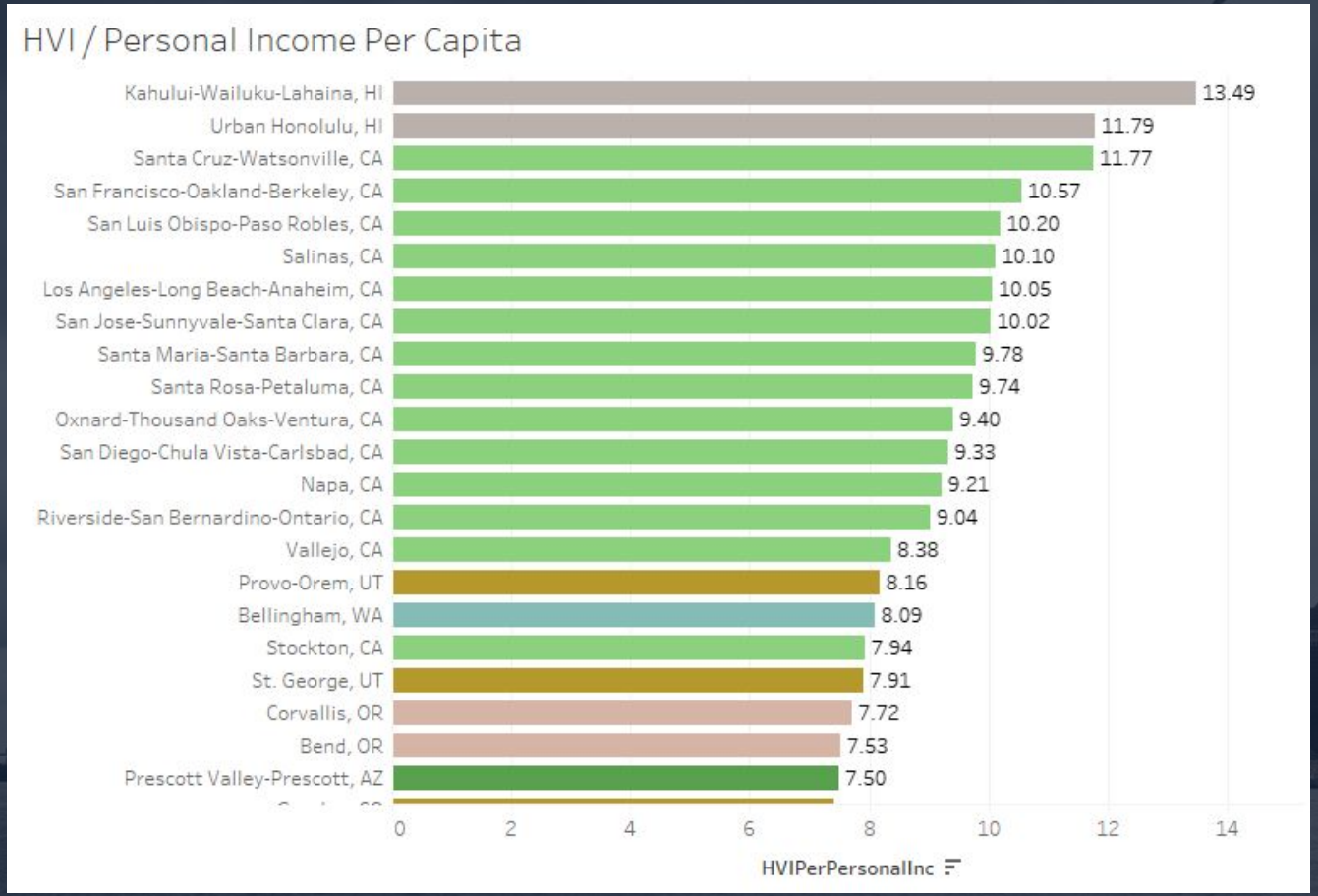
msa_id	name	PCT_Change_GDP	PCT_Change_Employment	PCT_Change_PersInc
28	Battle Creek, MI	1.3019	-0.5893	2.8847
30	Beaumont-Port Arthur, TX	-3.5521	0.9108	2.1234
39	Bloomington, IL	0.2253	-0.0860	0.2214
63	Charleston, WV	-2.4976	-3.1224	1.1788
70	Chico, CA	8.0748	-1.0928	2.1683
88	Dalton, GA	1.7021	-1.3297	2.4113
94	Decatur, IL	1.1505	-1.6755	0.6778
109	Elkhart-Goshen, IN	-1.1953	-2.3589	1.6964
110	Elmira, NY	4.5850	-0.1236	3.6977
115	Fairbanks, AK	0.3340	-0.6288	2.0685
117	Farmington, NM	0.4678	-0.8896	0.7888
120	Flagstaff, AZ	-0.8151	0.0310	1.3241
129	Gadsden, AL	2.6463	-0.8733	2.5571

HVI/ Personal Income per Capita

```
SELECT
    s.msa_id,
    s.name,
    ROUND(CAST( f.zillow_home_value_index as FLOAT)/((f.personal_income/f.population)*1000),2) AS
HVI_Per_PersonalInc_per_Capita
FROM
    fact_msa AS f
        INNER JOIN
    dim_location AS s ON f.msa_id = s.msa_id
WHERE f.date_id = 24
HAVING HVI_Per_PersonalInc_per_Capita <>0
ORDER BY HVI_Per_PersonalInc_per_Capita DESC;
```

msa_id	name	HVI_Per_PersonalInc_per_Capita
178	Kahului-Wailuku-Lahaina, HI	13.49
356	Urban Honolulu, HI	11.79
311	Santa Cruz-Watsonville, CA	11.77
308	San Francisco-Oakland-Berkeley, CA	10.57
310	San Luis Obispo-Paso Robles, CA	10.2
302	Salinas, CA	10.1
211	Los Angeles-Long Beach-Anaheim, CA	10.05
309	San Jose-Sunnyvale-Santa Clara, CA	10.02
313	Santa Maria-Santa Barbara, CA	9.78
314	Santa Rosa-Petaluma, CA	9.74
264	Oxnard-Thousand Oaks-Ventura, CA	9.4

HVI/ Personal Income per Capita



The average home in Hawaii is worth 11-13x the average annual income of its population.

13 out of the top 15 most expensive home relative to income belong to the state of California

MSAs with top home values by home type in 2019

Top Condo HVI, 2019

Name	
Los Angeles-Long Beach-A..	404,404
Napa, CA	381,412
Ocean City, NJ	381,142
Salinas, CA	376,887
Crestview-Fort Walton Be..	321,526
Urban Honolulu, HI	294,866
Midland, MI	288,308
Barnstable Town, MA	271,517
Santa Fe, NM	250,255
Boulder, CO	239,808
Panama City, FL	239,076

Top One Bedroom HVI

Name	
Salinas, CA	408,929
Los Angeles-Long Beach-A..	315,809
Urban Honolulu, HI	256,751
Santa Fe, NM	237,523
Barnstable Town, MA	215,089
Boulder, CO	212,284
Crestview-Fort Walton Be..	203,485
Missoula, MT	186,950
Salt Lake City, UT	178,673
Greeley, CO	173,498
Panama City, FL	172,517

Top Two Bedroom HVI

Name	
Salinas, CA	438,083
Los Angeles-Long Beach-A..	393,096
Urban Honolulu, HI	343,749
Barnstable Town, MA	295,502
Santa Fe, NM	276,070
Boulder, CO	267,318
Ocean City, NJ	235,879
Bend, OR	234,407
Greeley, CO	191,893
Salt Lake City, UT	191,825
Missoula, MT	190,976

Top Three Bedroom HVI

Name	
Napa, CA	478,460
Los Angeles-Long Beach-A..	472,329
Salinas, CA	448,178
Barnstable Town, MA	381,674
Boulder, CO	362,291
Ocean City, NJ	306,705
Santa Fe, NM	303,671
Bend, OR	271,538
Missoula, MT	241,275
North Port-Sarasota-Brad..	235,511
Greeley, CO	235,370

Top Four Bedroom HVI

Name	
Los Angeles-Long Beach-A..	597,491
Napa, CA	549,017
Barnstable Town, MA	525,625
Ocean City, NJ	510,825
Salinas, CA	504,721
Boulder, CO	463,352
Santa Fe, NM	400,011
Bend, OR	367,439
North Port-Sarasota-Brad..	347,286
Crestview-Fort Walton Be..	295,123
Missoula, MT	289,210

Top Five Bedroom HVI

Name	
Ocean City, NJ	902,676
Los Angeles-Long Beach-A..	893,343
Urban Honolulu, HI	767,516
Crestview-Fort Walton Be..	655,616
Santa Fe, NM	627,033
Napa, CA	588,451
Boulder, CO	579,421
Salinas, CA	552,599
Bend, OR	516,544
North Port-Sarasota-Brad..	493,733
Panama City, FL	351,105

One bedroom and two bedroom house HVIs drop off faster than condos, but five bedroom houses are more consistent in HVI across top MSAs.

Coastal and mountainous MSAs feature prominently in top 11 by HVI.

North Port - Sarasota FL shows up in top 10 only in 3+ bedroom HVIs.

Recommendations

Individuals

- Large MSAs that have shown fast growth in per capita GDP and Income, but HVI has not caught up: Boston, Seattle, Denver and Minneapolis.
- Low-cost MSAs with good opportunities: Homosassa Springs FL, Lake Charles LA, Sebring-Avon Park FL

Businesses

- Midland, TX showed one of the highest increases in GDP YoY (17.45% increase)
- Midland has a HVI per Personal Income lower than the national average (2.09x vs 4x)

Government

- Rocky Mount, NC experienced one of the highest declines in manufacturing GDP (-21.4%)
- Retraining manufacturing workers for other industries could help revitalize this area

Our Predictions

- Professional Services and Information Technology have shown to be the best industry contributors on average to long term growth in GDP, and remain the best bets for individual or business investments.
- MSA level opportunities will abound with many variations caused by spikes and dips in real estate, natural resources and manufacturing. As an example, Midland TX was one of America's hottest labor markets in 2019, fueled by fracking, but in 2020 it became a depression zone.
- Coastal cities like San Francisco, Honolulu and San Diego have seen a high increase in HVI over the last decade, but COVID-19 has altered the economics with work-from-home, and it will likely drive people to lower cost MSAs in 2021 without the need for a commute.
- Low-cost MSAs on coasts or in mountains are likely to benefit in 2021, going by people's past preferences on where to live.

Corrective Measures

- More measures can be added to provide more meaningful recommendations on where to move to. For example weather, crime, education data etc.
- Current analysis is based on descriptive findings. We could build predictive models to identify features that are statistically significant. We could also use macro features to predict future GDP, home value or rental prices.
- Industry-specific data was only pulled for the last four years to limit the size of the files. The industry data can be expanded to look at longer term trends.


Thank You