

## Data Science Independent Project – Home Value Trends

### Objective

“You are asked by a company to help them make more informed decisions on real estate investments. Start by analyzing the data on median estimated values of single family homes by zip codes from the past two decades.”

### Data Collection

The data was collected from <https://www.zillow.com/research/data/> for the years 1997 to 2017. Guidance was taken from answers provided by users "intesers" and "alexis-colvin" on GitHub.

### Disclaimer

I am writing to acknowledge that I do not possess extensive knowledge or professional expertise in real estate value. However, I am utilizing this project as an opportunity to practice and enhance my data analysis skills using SQL. I kindly request your comments or advice to help me improve and grow in this area.

### Queries

#### 1. How many distinct zip codes are in this dataset?

```
1. SELECT COUNT(DISTINCT zip_code) AS 'Number of zip codes'
2. FROM home_value_trends;
```

Number of zip codes
15452

#### 2. How many zip codes are from each state?

```
1. SELECT state, COUNT(DISTINCT zip_code) AS 'Number of zip codes per state'
2. FROM home_value_trends
3. GROUP BY state;
```

state	Number of zip codes per state
AK	28
AL	221
AR	119
AZ	233
CA	1230
CO	261
CT	122
DC	18
DE	39
FL	795
GA	375
HI	62
IA	223

ID	111
IL	571
IN	420
KS	242
KY	163
LA	195
MA	419
MD	322
ME	95
MI	547
MN	416
MO	326
MS	146
MT	72

NC	489
ND	33
NE	89
NH	198
NJ	501
NM	60
NV	101
NY	1081
OH	630
OK	230
OR	224
PA	875

RI	60
SC	215
SD	20
TN	436
TX	940
UT	121
VA	428
VT	62
WA	348
WI	426
WV	82
WY	32

### 3. What range of years are represented in the data?

```

1. SELECT MIN(substr(date,1,4)) 'Start Year' , MAX(substr(date,1,4)) 'End Year'
2. FROM home_value_trends;
3. --substr() returns a substring from a string starting at a specified position with a
predefined length--
-- substr(string, start, length); start is an integer that specifies the starting position of
the returned substring
-- If start is positive, it starts from beginning of the string and if it's negative it starts
at the end of the string (last character has a string of -1)

```

Start Year	End Year
1996	2018

The range of years represented in the data is between 1996 to 2018.

### 4. Using the most recent month of data available, what is the range of estimated home values across the nation?

```

1. SELECT MIN(value) AS 'Minimum value' , MAX(value) AS 'Maximum value'
2. FROM home_value_trends
3. WHERE date = (SELECT MAX(date) FROM home_value_trends);

```

Minimum value	Maximum value
21600.0	17757800.0

With the most recent month available, the range of estimated home values across the US was USD 21600 to 17757800.

### 5. Using the most recent month of data available, which states have the highest average home values? How about the lowest?

```

1. SELECT date, state, ROUND(AVG(value)) AS 'Avg. home values'
2. FROM home_value_trends
3. WHERE date = (SELECT MAX(date) FROM home_value_trends)
4. GROUP BY state
5. ORDER BY ROUND(AVG(value)) DESC;

```

date	state	Avg. home values
2018-11	DC	826572.0
2018-11	CA	750965.0
2018-11	HI	711085.0
2018-11	MA	475927.0
2018-11	CO	442713.0
2018-11	WA	414105.0
2018-11	NJ	403476.0
2018-11	NY	378121.0
2018-11	NV	349835.0
2018-11	UT	343777.0
2018-11	OR	335449.0
2018-11	MD	332744.0
2018-11	RI	323568.0
2018-11	AK	322957.0
2018-11	FL	298899.0
2018-11	CT	293402.0
2018-11	VA	285036.0
2018-11	AZ	280861.0
2018-11	NH	277524.0
2018-11	ME	265288.0
2018-11	ID	261025.0
2018-11	DE	259938.0
2018-11	MT	247658.0
2018-11	WY	238309.0
2018-11	NM	233565.0

2018-11	VT	230911.0
2018-11	MN	226873.0
2018-11	IL	210955.0
2018-11	TX	210330.0
2018-11	GA	207822.0
2018-11	SD	204150.0
2018-11	SC	196124.0
2018-11	WI	193712.0
2018-11	NC	191436.0
2018-11	PA	187597.0
2018-11	ND	186442.0
2018-11	NE	184194.0
2018-11	MI	172428.0
2018-11	MO	169915.0
2018-11	IA	168170.0
2018-11	KY	156956.0
2018-11	TN	154910.0
2018-11	AL	152781.0
2018-11	OH	152599.0
2018-11	LA	148747.0
2018-11	KS	146700.0
2018-11	IN	143075.0
2018-11	AR	139539.0
2018-11	WV	123968.0
2018-11	MS	123058.0
2018-11	OK	121546.0

With the recent month, November of 2018, the state with the highest average home value is DC while the state with the lowest average home value is OK.

#### 6. Which states have the highest/lowest average home values for the year 2017?

What about for the year of 2007? 1997?

```

1. SELECT substr(date, 1, 4) AS 'Year', state, ROUND(AVG(value), 2) AS 'Avg. value'
2. FROM home_value_trends
3. WHERE substr(date, 1, 4) = '2017' --Change the year --
4. GROUP BY state
5. ORDER BY ROUND(AVG(value), 2) DESC;
```

In the year 2017, DC had the highest average home values, whereas OK had the lowest average home values. For the years 2007 and 1997, HI had the highest home values. However, similar to the year 2017, OK had the lowest average home values for 2007, and for the year 1997, it was ND.

Year	state	Avg. value
2017	DC	778756.02
2017	CA	688609.97
2017	HI	653451.75
2017	MA	438244.63
2017	CO	394511.46
2017	NJ	375706.84
2017	WA	369567.15
2017	NY	359342.49
2017	AK	318021.43
2017	MD	314660.25
2017	OR	306799.07
2017	NV	301538.53
2017	UT	294948.76
2017	RI	294574.86
2017	CT	280047.68
2017	FL	272582.98
2017	VA	269285.32
2017	AZ	255581.97
2017	NH	252847.1
2017	DE	248462.39
2017	ME	240862.11
2017	ID	223306.91
2017	WY	222159.9
2017	MT	221810.42
2017	VT	219629.97
2017	NM	212525.42
2017	MN	207513.66
2017	IL	201485.96
2017	TX	192054.76
2017	GA	184896.8
2017	SD	184605.42
2017	ND	183148.99
2017	WI	176782.43
2017	SC	176363.72
2017	PA	174637.41
2017	NC	174062.83
2017	NE	166257.87
2017	IA	156758.26
2017	MI	155352.13
2017	MO	153369.2
2017	KY	143375.72
2017	LA	142874.7
2017	AL	139148.3
2017	OH	138364.51
2017	TN	138124.43
2017	KS	136451.0
2017	AR	128088.31
2017	IN	125044.21
2017	WV	118281.61
2017	MS	115630.65
2017	OK	112451.7

Year	state	Avg. value
2007	HI	611893.93
2007	DC	599893.98
2007	CA	598141.04
2007	NJ	440055.9
2007	MA	377822.66
2007	MD	369100.47
2007	NV	354913.26
2007	CT	341385.11
2007	RI	322582.22
2007	NY	318182.44
2007	WA	316000.91
2007	CO	313130.9
2007	DE	304398.15
2007	FL	299784.12
2007	AZ	295021.39
2007	VA	283551.94
2007	AK	275928.87
2007	OR	275055.33
2007	NH	254791.45
2007	UT	254423.06
2007	NM	232761.79
2007	ME	231709.8
2007	IL	230522.74
2007	VT	224758.48
2007	MT	215423.5
2007	ID	210176.45
2007	WY	202277.08
2007	MN	199287.95
2007	GA	182825.09
2007	SC	180254.11
2007	PA	176372.29
2007	NC	170933.84
2007	WI	169244.15
2007	TX	152363.82
2007	MO	151943.81
2007	AL	148621.43
2007	MI	146659.44
2007	KS	144679.71
2007	SD	143045.61
2007	NE	137831.83
2007	MS	131843.64
2007	OH	131262.94
2007	LA	129153.11
2007	IA	128333.03
2007	KY	126206.01
2007	ND	124564.08
2007	IN	120574.49
2007	AR	120089.75
2007	TN	117068.67
2007	WV	112406.01
2007	OK	91773.39

Year	state	Avg. value
1997	HI	216377.12
1997	CA	207479.51
1997	DC	189769.44
1997	NJ	175576.6
1997	MA	168343.61
1997	UT	162484.05
1997	CT	159184.02
1997	MT	157241.5
1997	NV	157095.79
1997	CO	154106.44
1997	WA	147066.59
1997	MD	144904.11
1997	NM	139284.75
1997	WY	135059.57
1997	AK	134421.13
1997	OR	132921.59
1997	DE	126937.0
1997	RI	126419.31
1997	NY	126348.63
1997	IL	123320.01
1997	VA	122760.53
1997	AZ	122682.65
1997	TX	113221.74
1997	WI	111299.37
1997	NH	109703.39
1997	FL	108010.5
1997	ID	102877.86
1997	MI	99656.75
1997	VT	99300.0
1997	GA	98839.32
1997	MO	97086.51
1997	ME	96941.37
1997	NE	93304.28
1997	OH	92003.07
1997	NC	91369.82
1997	MN	91002.4
1997	PA	89577.67
1997	AL	89567.68
1997	SC	89122.47
1997	KS	88037.43
1997	IN	87203.9
1997	MS	86224.88
1997	KY	82019.24
1997	IA	76168.14
1997	AR	72332.22
1997	WV	69983.47
1997	LA	69803.65
1997	TN	68213.77
1997	SD	59295.0
1997	OK	56325.62
1997	ND	

## 7. What is the percentage change in average home values from 2007 to 2017 by state? How about from 1997 to 2017?

```

1. WITH val_2017 AS(
2.     SELECT     substr(date,1,4) AS year, state, ROUND(AVG(value)) AS average
3.     FROM home_value_trends
4.     WHERE year = '2017'

```

```

5.      GROUP BY 2,1
6. ),
7.      val_2007 AS(
8.      SELECT      substr(date,1,4) AS year, state, ROUND(AVG(value)) AS average
9.      FROM home_value_trends
10.     WHERE year = '2007'
11.     GROUP BY 2,1
12. ),
13.      val_1997 AS(
14.      SELECT      substr(date,1,4) AS year, state, ROUND(AVG(value)) AS average
15.      FROM home_value_trends
16.      WHERE year = '1997'
17.      GROUP BY 2,1
18. )
19.
20. SELECT
21.      val_2017.state, val_1997.average AS '1997 Ave.', val_2007.average AS '2007 Ave.',
22.      val_2017.average AS '2017 Ave.',
23.      ROUND((100.0 * (val_2017.average - val_2007.average) / val_2007.average),2) AS '%
Change 2007-2017',
24.      ROUND((100.0 * (val_2017.average - val_1997.average) / val_1997.average),2) AS '%
Change 1997-2017'
25. FROM val_2017
26. JOIN val_2007
27.     ON val_2017.state = val_2007.state
28. JOIN val_1997
29.     ON val_2007.state = val_1997.state
30. ORDER BY 1 ASC;

```

state	1997 Ave.	2007 Ave.	2017 Ave.	% Change 2007-2017	% Change 1997-2017
AK	134421.0	275929.0	318021.0	15.25	136.59
AL	89568.0	148621.0	139148.0	-6.37	55.35
AR	72332.0	120090.0	128088.0	6.66	77.08
AZ	122683.0	295021.0	255582.0	-13.37	108.33
CA	207480.0	598141.0	688610.0	15.13	231.89
CO	154106.0	313131.0	394511.0	25.99	156.0
CT	159184.0	341385.0	280048.0	-17.97	75.93
DC	189769.0	599894.0	778756.0	29.82	310.37
DE	126937.0	304398.0	248462.0	-18.38	95.74
FL	108011.0	299784.0	272583.0	-9.07	152.37
GA	98839.0	182825.0	184897.0	1.13	87.07
HI	216377.0	611894.0	653452.0	6.79	202.0
IA	76168.0	128333.0	156758.0	22.15	105.81
ID	102878.0	210176.0	223307.0	6.25	117.06
IL	123320.0	230523.0	201486.0	-12.6	63.38
IN	87204.0	120574.0	125044.0	3.71	43.39
KS	88037.0	144680.0	136451.0	-5.69	54.99
KY	82019.0	126206.0	143376.0	13.6	74.81
LA	69804.0	129153.0	142875.0	10.62	104.68
MA	168344.0	377823.0	438245.0	15.99	160.33
MD	144904.0	369100.0	314660.0	-14.75	117.15
ME	96941.0	231710.0	240862.0	3.95	148.46

MI	99657.0	146659.0	155352.0	5.93	55.89
MN	91002.0	199288.0	207514.0	4.13	128.03
MO	97087.0	151944.0	153369.0	0.94	57.97
MS	86225.0	131844.0	115631.0	-12.3	34.1
MT	157242.0	215424.0	221810.0	2.96	41.06
NC	91370.0	170934.0	174063.0	1.83	90.5
ND		124564.0	183149.0	47.03	
NE	93304.0	137832.0	166258.0	20.62	78.19
NH	109703.0	254791.0	252847.0	-0.76	130.48
NJ	175577.0	440056.0	375707.0	-14.62	113.98
NM	139285.0	232762.0	212525.0	-8.69	52.58
NV	157096.0	354913.0	301539.0	-15.04	91.95
NY	126349.0	318182.0	359342.0	12.94	184.4
OH	92003.0	131263.0	138365.0	5.41	50.39
OK	56326.0	91773.0	112452.0	22.53	99.64
OR	132922.0	275055.0	306799.0	11.54	130.81
PA	89578.0	176372.0	174637.0	-0.98	94.96
RI	126419.0	322582.0	294575.0	-8.68	133.01
SC	89122.0	180254.0	176364.0	-2.16	97.89
SD	59295.0	143046.0	184605.0	29.05	211.33
TN	68214.0	117069.0	138124.0	17.99	102.49
TX	113222.0	152364.0	192055.0	26.05	69.63
UT	162484.0	254423.0	294949.0	15.93	81.52
VA	122761.0	283552.0	269285.0	-5.03	119.36
VT	99300.0	224758.0	219630.0	-2.28	121.18
WA	147067.0	316001.0	369567.0	16.95	151.29
WI	111299.0	169244.0	176782.0	4.45	58.84
WV	69983.0	112406.0	118282.0	5.23	69.02
WY	135060.0	202277.0	222160.0	9.83	64.49

**8. How would you describe the trend in home values for each state from 1997 to 2017? How about from 2007 to 2017? Which states would you recommend for making real estate investments?**

```
1. WITH val_2017 AS(
2.     SELECT     substr(date,1,4) AS year, state, ROUND(AVG(value)) AS average
3.     FROM home_value_trends
4.     WHERE year = '2017'
5.     GROUP BY 2,1
6. ),
7.     val_2007 AS(
8.     SELECT     substr(date,1,4) AS year, state, ROUND(AVG(value)) AS average
9.     FROM home_value_trends
10.    WHERE year = '2007'
11.    GROUP BY 2,1
12. ),
13.     val_1997 AS(
14.     SELECT     substr(date,1,4) AS year, state, ROUND(AVG(value)) AS average
15.     FROM home_value_trends
16.     WHERE year = '1997'
17.     GROUP BY 2,1
18. )
19.
20. SELECT
21.     val_2017.state, val_1997.average AS '1997 Ave.', val_2007.average AS '2007 Ave.',
22.     val_2017.average AS '2017 Ave.',
23.     ROUND((100.0 * (val_2017.average - val_2007.average) / val_2007.average),2) AS '%
Change 2007-2017',
24.     ROUND((100.0 * (val_2017.average - val_1997.average) / val_1997.average),2) AS '%
Change 1997-2017'
25. FROM val_2017
26. JOIN val_2007
27.     ON val_2017.state = val_2007.state
28. JOIN val_1997
29.     ON val_2007.state = val_1997.state
30. ORDER BY 6 DESC;--Change to 5 to see trend from 2000 to 2017--
```

From 1997 to 2017, the home values in most states experienced a significant appreciation with DC, CA, SD, HI, and NY having the highest percentage increase in home values with percentage changes ranging from approximately 184% to over 310%. Other states have also experienced growth in home values including MA, CO, WA, ME, and AK, with percentage changes ranging from approximately 120% to 156%. On the other hand, states like MS, MT, IN ,OH, and NM exhibited relatively lower percentage increases in home values ranging from 34% to 52%.

From 2007 to 2017, DC, SD, and ID experienced significant increases in home values with percentage changes ranging from 117% to 211%. However, a number of states have reported declines or minimal growth. States like DE, NJ, MD, FL, and NV reported negative changes, indicating a decrease in home values.

Based on the data, DC, SD, and ID exhibited strong growth in home values and may be attractive options for real estate investments, as they demonstrated a substantial increase in home values. Investors should be cautious of states such as NJ, MD, FL, and NV which have reported negative or minimal growth in home values.

state	1997 Ave.	2007 Ave.	2017 Ave.	% Change 2007-2017	% Change 1997-2017
DC	189769.0	599894.0	778756.0	29.82	310.37
CA	207480.0	598141.0	688610.0	15.13	231.89
SD	59295.0	143046.0	184605.0	29.05	211.33
HI	216377.0	611894.0	653452.0	6.79	202.0
NY	126349.0	318182.0	359342.0	12.94	184.4
MA	168344.0	377823.0	438245.0	15.99	160.33
CO	154106.0	313131.0	394511.0	25.99	156.0
FL	108011.0	299784.0	272583.0	-9.07	152.37
WA	147067.0	316001.0	369567.0	16.95	151.29
ME	96941.0	231710.0	240862.0	3.95	148.46
AK	134421.0	275929.0	318021.0	15.25	136.59
RI	126419.0	322582.0	294575.0	-8.68	133.01
OR	132922.0	275055.0	306799.0	11.54	130.81
NH	109703.0	254791.0	252847.0	-0.76	130.48
MN	91002.0	199288.0	207514.0	4.13	128.03
VT	99300.0	224758.0	219630.0	-2.28	121.18
VA	122761.0	283552.0	269285.0	-5.03	119.36
MD	144904.0	369100.0	314660.0	-14.75	117.15
ID	102878.0	210176.0	223307.0	6.25	117.06
NJ	175577.0	440056.0	375707.0	-14.62	113.98
AZ	122683.0	295021.0	255582.0	-13.37	108.33
IA	76168.0	128333.0	156758.0	22.15	105.81
LA	69804.0	129153.0	142875.0	10.62	104.68
TN	68214.0	117069.0	138124.0	17.99	102.49
OK	56326.0	91773.0	112452.0	22.53	99.64
SC	89122.0	180254.0	176364.0	-2.16	97.89
DE	126937.0	304398.0	248462.0	-18.38	95.74
PA	89578.0	176372.0	174637.0	-0.98	94.96
NV	157096.0	354913.0	301539.0	-15.04	91.95
NC	91370.0	170934.0	174063.0	1.83	90.5
GA	98839.0	182825.0	184897.0	1.13	87.07
UT	162484.0	254423.0	294949.0	15.93	81.52
NE	93304.0	137832.0	166258.0	20.62	78.19
AR	72332.0	120090.0	128088.0	6.66	77.08
CT	159184.0	341385.0	280048.0	-17.97	75.93
KY	82019.0	126206.0	143376.0	13.6	74.81
TX	113222.0	152364.0	192055.0	26.05	69.63
WV	69983.0	112406.0	118282.0	5.23	69.02
WY	135060.0	202277.0	222160.0	9.83	64.49
IL	123320.0	230523.0	201486.0	-12.6	63.38
WI	111299.0	169244.0	176782.0	4.45	58.84
MO	97087.0	151944.0	153369.0	0.94	57.97
MI	99657.0	146659.0	155352.0	5.93	55.89
AL	89568.0	148621.0	139148.0	-6.37	55.35
KS	88037.0	144680.0	136451.0	-5.69	54.99
NM	139285.0	232762.0	212525.0	-8.69	52.58
OH	92003.0	131263.0	138365.0	5.41	50.39
IN	87204.0	120574.0	125044.0	3.71	43.39
MT	157242.0	215424.0	221810.0	2.96	41.06
MS	86225.0	131844.0	115631.0	-12.3	34.1
ND		124564.0	183149.0	47.03	

state	1997 Ave.	2007 Ave.	2017 Ave.	% Change 2007-2017	% Change 1997-2017
ND		124564.0	183149.0	47.03	
DC	189769.0	599894.0	778756.0	29.82	310.37
SD	59295.0	143046.0	184605.0	29.05	211.33
TX	113222.0	152364.0	192055.0	26.05	69.63
CO	154106.0	313131.0	394511.0	25.99	156.0
OK	56326.0	91773.0	112452.0	22.53	99.64
IA	76168.0	128333.0	156758.0	22.15	105.81
NE	93304.0	137832.0	166258.0	20.62	78.19
TN	68214.0	117069.0	138124.0	17.99	102.49
WA	147067.0	316001.0	369567.0	16.95	151.29
MA	168344.0	377823.0	438245.0	15.99	160.33
UT	162484.0	254423.0	294949.0	15.93	81.52
AK	134421.0	275929.0	318021.0	15.25	136.59
CA	207480.0	598141.0	688610.0	15.13	231.89
KY	82019.0	126206.0	143376.0	13.6	74.81
NY	126349.0	318182.0	359342.0	12.94	184.4
OR	132922.0	275055.0	306799.0	11.54	130.81
LA	69804.0	129153.0	142875.0	10.62	104.68
WY	135060.0	202277.0	222160.0	9.83	64.49
HI	216377.0	611894.0	653452.0	6.79	202.0
AR	72332.0	120090.0	128088.0	6.66	77.08
ID	102878.0	210176.0	223307.0	6.25	117.06
MI	99657.0	146659.0	155352.0	5.93	55.89
OH	92003.0	131263.0	138365.0	5.41	50.39
WV	69983.0	112406.0	118282.0	5.23	69.02
WI	111299.0	169244.0	176782.0	4.45	58.84
MN	91002.0	199288.0	207514.0	4.13	128.03
ME	96941.0	231710.0	240862.0	3.95	148.46
IN	87204.0	120574.0	125044.0	3.71	43.39
MT	157242.0	215424.0	221810.0	2.96	41.06
NC	91370.0	170934.0	174063.0	1.83	90.5
GA	98839.0	182825.0	184897.0	1.13	87.07
MO	97087.0	151944.0	153369.0	0.94	57.97
NH	109703.0	254791.0	252847.0	-0.76	130.48
PA	89578.0	176372.0	174637.0	-0.98	94.96
SC	89122.0	180254.0	176364.0	-2.16	97.89
VT	99300.0	224758.0	219630.0	-2.28	121.18
VA	122761.0	283552.0	269285.0	-5.03	119.36
KS	88037.0	144680.0	136451.0	-5.69	54.99
AL	89568.0	148621.0	139148.0	-6.37	55.35
RI	126419.0	322582.0	294575.0	-8.68	133.01
NM	139285.0	232762.0	212525.0	-8.69	52.58
FL	108011.0	299784.0	272583.0	-9.07	152.37
MS	86225.0	131844.0	115631.0	-12.3	34.1
IL	123320.0	230523.0	201486.0	-12.6	63.38
AZ	122683.0	295021.0	255582.0	-13.37	108.33
NJ	175577.0	440056.0	375707.0	-14.62	113.98
MD	144904.0	369100.0	314660.0	-14.75	117.15
NV	157096.0	354913.0	301539.0	-15.04	91.95
CT	159184.0	341385.0	280048.0	-17.97	75.93
DE	126937.0	304398.0	248462.0	-18.38	95.74