codecademy

Trends Estimated Home Value

Analyze Data with SQL

Petru Apachitei

Date: 09-12-2020

Example Table of Contents

1. Introduction

1.1 Get familiar with dataset

1. How many zip codes are in dataset?
   - 15453 distinct zip codes
2. How many zip codes are in each state?
   - Top 5 states
3. What range of years are represented in the data?
   - The range of years is between 1996 to 2018
4. Using the most recent month of data available, what is the average of estimated home values across the nation?

The SQL queries for these question are attached at the Appendix 1.

states	Zip_codes
CA	334560
NY	294032
TX	255680
PA	238000
FL	216240
CA	334560

Months 2018	MIN	MAX	Range
01	20800	18329300	18308500
02	21600	18177700	18156100
03	21700	17729100	17707400
04	21700	17436800	17415100
05	22100	17475300	17453200
06	22200	17731900	17709700
07	22000	18024900	18002900
08	21800	18296000	18274200
09	21700	18357900	18336200
10	21500	18114800	18093300
11	21600	17757800	17736200

2. Analysis

2.1 Analysis 2018

1. Using the most recent month of data available. Which states have the highest average home and lowest value ?
   - Column Average 2018 and Lowest 2018, displays the top 10 states.

```
-- create a query for top 10 states with highest average value of the
last months
SElECT state,
            ROUND(avg(value), 1) as average_value_2018
FROM home_value_data
WHERE date like '2018%'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;

-- create a query for top 10 states with lowest value of the last months

SElECT state,
            ROUND(min(value), 1) as average_value_2018
FROM home_value_data
WHERE date like '2018%'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;
```

State	Average value 2018	State	Lowest value 2018
DC	810175.8	DC	303200.0
CA	740019.7	AK	204600.0
HI	696020.2	HI	175300.0
MA	465974.6	UT	164100.0
CO	426110.6	RI	161800.0
WA	404123.3	WY	141400.0
NJ	397220.3	NV	140900.0
NY	374716.2	SD	128500.0
NV	333730.7	MA	126000.0
MD	328104.0	MT	119500.0

2.2 Analysis 2017

1. Using the data available. Which states have the highest and lowest value ?
   - Column Highest 2017 and Lowest 2017, displays the top 10 states.

```
-- create a query for top 10 states with highest average value for 2017

SElECT state,
            ROUND(max(value), 1) as average_value_2018
FROM home_value_data
WHERE date like '2017%'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;
-- create a query for top 10 states with lowest average value for 2017

SElECT state,
            ROUND(min(value), 1) as average_value_2017
FROM home_value_data
WHERE date like '2017%'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;
```

State	Highest value 2017	State	Lowest value 2017
NY	19342000.0	DC	259600.0
CA	6507000.0	AK	203100.0
FL	4596000.0	HI	157400.0
CO	4259900.0	UT	155800.0
WA	2699300.0	RI	142100.0
GA	2434800.0	WY	134300.0
NV	2148300.0	SD	120300.0
MA	2034100.0	MT	113800.0
NJ	1890400.0	NV	113200.0
SC	1707800.0	MA	108800.0

2.3 Analysis 2007

1. Using the data available. Which states have the highest and lowest value ?
   - Column Average 2007 and Lowest 2007, displays the top 10 states.

```
-- create a query for top 10 states with highest average value for 2017

SElECT state,
            ROUND(max(value), 1) as average_value_2007
FROM home_value_data
WHERE date like '2007%'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;
-- create a query for top 10 states with lowest average value for 2017

SElECT state,
            ROUND(min(value), 1) as average_value_2007
FROM home_value_data
WHERE date like '2007%'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;
```

State	Highest value 2007	State	Lowest value 2007
NY	10058000.0	DC	245700.0
CO	4754700.0	RI	194700.0
CA	3729100.0	AK	184000.0
FL	3435300.0	NV	157100.0
GA	2936200.0	CT	151000.0
NV	2054000.0	DE	140100.0
NJ	2017800.0	UT	129900.0
CT	1808000.0	MA	129300.0
WA	1805300.0	CA	127600.0
IL	1718800.0	WY	114000.0

2.4 Analysis 1997

1. Using the data available. Which states have the highest and lowest value ?
   - Column Average 1997 and Lowest 1997, displays the top 10 states.

```
-- create a query for top 10 states with highest average value for 2017

SElECT state,
            ROUND(max(value), 1) as average_value_1997
FROM home_value_data
WHERE date like '1997%'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;
-- create a query for top 10 states with lowest average value for 1997

SElECT state,
            ROUND(min(value), 1) as average_value_1997
FROM home_value_data
WHERE date like '1997%'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;
```

State	Highest value 1997	State	Lowest value 1997
NY	4165900.0	MT	109800.0
CO	1684300.0	NV	93600.0
CA	1379200.0	AK	93400.0
FL	1029900.0	WY	91200.0
TX	754100.0	UT	88700.0
CT	750100.0	DC	82500.0
IL	700100.0	NM	78400.0
WA	689800.0	RI	70100.0
MA	673600.0	CT	70000.0
NJ	655600.0	MA	58900.0

2.5 Combine tables

State	Lowest value 2018	State	Lowest value 2017	State	Lowest value 2007	State	Lowest value 1997
DC	303200.0	DC	259600.0	DC	245700.0	MT	109800.0
AK	204600.0	AK	203100.0	RI	194700.0	NV	93600.0
HI	175300.0	HI	157400.0	AK	184000.0	AK	93400.0
UT	164100.0	UT	155800.0	NV	157100.0	WY	91200.0
RI	161800.0	RI	142100.0	CT	151000.0	UT	88700.0
WY	141400.0	WY	134300.0	DE	140100.0	DC	82500.0
NV	140900.0	SD	120300.0	UT	129900.0	NM	78400.0
SD	128500.0	MT	113800.0	MA	129300.0	RI	70100.0
MA	126000.0	NV	113200.0	CA	127600.0	CT	70000.0
MT	119500.0	MA	108800.0	WY	114000.0	MA	58900.0

State	Highest value 2017	State	Highest value 2007	State	Highest value 1997
NY	19342000.0	NY	10058000.0	NY	4165900.0
CA	6507000.0	CO	4754700.0	CO	1684300.0
FL	4596000.0	CA	3729100.0	CA	1379200.0
CO	4259900.0	FL	3435300.0	FL	1029900.0
WA	2699300.0	GA	2936200.0	TX	754100.0
GA	2434800.0	NV	2054000.0	CT	750100.0
NV	2148300.0	NJ	2017800.0	IL	700100.0
MA	2034100.0	CT	1808000.0	WA	689800.0
NJ	1890400.0	WA	1805300.0	MA	673600.0
SC	1707800.0	IL	1718800.0	NJ	655600.0

3.1 Show growth for home values in 10 years 2007 to 2017.

Calculate the percent change for average home value by state, then order the top ten state with higher percent change.

The related query is attached under Appendix 2.

State	Avg_value 2007	Avg_value 2017	Percent Change %
ND	124564.08	183148.99	47.0
DC	599893.98	778756.02	29.8
SD	143045.61	184605.42	29.1
TX	152363.82	192054.76	26.1
CO	313130.9	394511.46	26.0
OK	91773.39	112451.7	22.5
IA	128333.03	156758.26	22.1
NE	137831.83	166257.87	20.6
TN	117068.67	138124.43	18.0
WA	316000.91	369567.15	17.0

3.2 Show growth for home values in 20 years 1997 to 2017.

Calculate the percent change for average home value by state, then order the top ten state with higher percent change.

The related query is attached under Appendix 3.

State	Avg_value 1997	Avg_value 2017	Percent Change %
DC	189769.44	778756.02	310.4
CA	207479.51	688609.97	231.9
SD	59295.0	184605.42	211.3
HI	216377.12	653451.75	202.0
NY	126348.63	359342.49	184.4
MA	168343.61	438244.63	160.3
CO	154106.44	394511.46	156.0
FL	108010.5	272582.98	152.4
WA	147066.59	369567.15	151.3
ME	96941.37	240862.11	148.5

3.3 Insights

How would you describe the trend in home values from 1997 to 2017?

- According to data collected the average price for 2007 was 239040 $ and then 249666 $ in 2017, that shows a growth of more than 5% in 10 years.
- The average price for 1997 was 166366 $ and then 249666 $ in 2017, that shows a growth of more than 100% in 20 years.

Which states would you recommend for making real estate investments for 2018?
Calculate the percent change for average home value from 2016 to 2017.
Top five states for investments property are:

WA	12.5%
OR	9.5%
UT	9.3%
ID	8.8%
NV	8.8%

Appendix 1

```
-- create a query to filter unique zip codes
SELECT Count(Distinct zip_code)
FROM home_value_data;

-- create a query to group zip codes on states
                        SELECT state, Count(*)
FROM home_value_data
GROUP BY 1
ORDER BY 2 DESC;

-- create a query to filter range year from date column
SELECT Distinct(substr(date, 0, 5)) as range_year
FROM home_value_data;

-- create a query to filter recents months and group them by
estimated home values
SElECT substr(date, 6, 8) as month,
             MIN(value)as min_value, MAX(value)as max_value,
             (MAX(value) - MIN(value)) as range_value
FROM home_value_data
WHERE date like '2018%'
GROUP BY 1;
```

Appendix 2

```
WITH year_2007 as
             (SElECT state,
                          ROUND(avg(value), 2) as avg_value_1997
             FROM home_value_data
             WHERE date like '2007%'
                                                          GROUP BY 1),

year_2017 as
             (SELECT state,
                          ROUND(avg(value), 2) as avg_value_2017
             FROM home_value_data
             WHERE date like '2017%'
             GROUP BY 1)
SELECT year_2007.state,
                          year_2007.avg_value_2007,
                          year_2017.avg_value_2017,
                          ROUND(((year_2017.avg_value_2017 -
year_1997.avg_value_1997) / year_2007.avg_value_2007) * 100.0, 1) as
percent_change
FROM year_2007
JOIN year_2017
ON year_2007.state == year_2017.state
ORDER BY 4 DESC;
```

Appendix 3

```
WITH year_1997 as
             (SElECT state,
                          ROUND(avg(value), 2) as avg_value_1997
             FROM home_value_data
             WHERE date like '1997%'
             GROUP BY 1),

year_2017 as
             (SELECT state,
                          ROUND(avg(value), 2) as avg_value_2017
             FROM home_value_data
             WHERE date like '2017%'
             GROUP BY 1)
SELECT year_1997.state,
                          year_1997.avg_value_1997,
                          year_2017.avg_value_2017,
                          ROUND(((year_2017.avg_value_2017 -
year_1997.avg_value_1997) / year_1997.avg_value_1997) * 100.0, 1) as
percent_change
FROM year_1997
JOIN year_2017
ON year_1997.state == year_2017.state
ORDER BY 4 DESC;
```