CIS 4560 Fall 2022
Housing Prices Tutorial

Authors: Adrian Soriano, Christian Momdjian, Joshua Ng, Louis Banegas, Marc Calvillo

Instructor: Jongwook Woo

Date: 12/18/2022

Lab Tutorial (Team 2)

Christian Momdjian (cmomdji@calstatela.edu)

Adrian Soriano (asoria55@calstatela.edu)

Housing Costs in California

Objectives

In this hands-on lab, you will learn how to:

1. Get data using *wget* and save it to the Oracle cluster
2. Upload the file to HDFS from Oracle
3. Create a database table of all the data
4. Create another table with filtered data to export to a new file
5. Visualize the exported data using Microsoft PowerBI

Step 1: Get data using WGET and save it to the Oracle cluster

This step details how you are going to get the dataset to work with for this tutorial. The original dataset is very large at around 49GB, which makes it hard to download the data straight from the source URL. So, to avoid issues with downloading it, we pulled 2GB worth of the data (which holds all our California records along with some other states) and stored it in a Dropbox location where it can be downloaded using the WGET tool:

1. Connect to the Oracle cluster using your username (Don't use cmomdji):

   ```
   ssh cmomdji@144.24.14.145
   ```

2. Download the data from the Dropbox location we provided:

   ```
   wget https://www.dropbox.com/s/3xt3up4fve78me8/dolthub_us-
   housing-prices_main_sales_REDUCED2.csv?dl=0
   ```

3. Rename the downloaded file:

   ```
   mv dolthub_us-housing-prices_main_sales_REDUCED2.csv?dl=0

   us-housing-prices.csv
   ```

4. Make sure the downloaded file is now in the Oracle cluster:

   ```
   ls -al
   ```

Step 2: Upload the file to HDFS from Oracle

Now that we have the data, we need to upload it to HDFS so we can store it properly and avoid taking up too much space on the Oracle cluster.

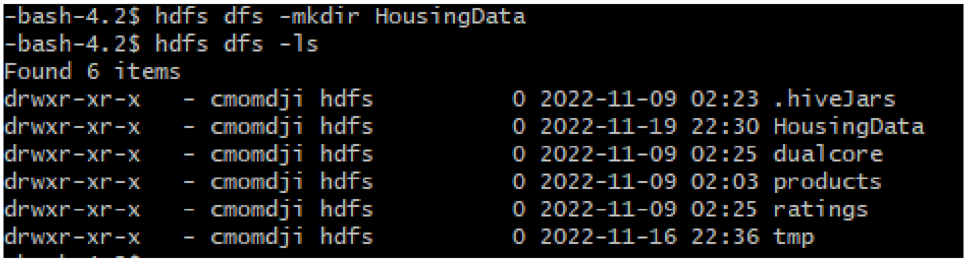
1. First start by creating a new directory to contain the data:

   ```
   hdfs dfs -mkdir HousingData
   ```

2. Check to see that the directory was created in HDFS:

   ```
   hdfs dfs -ls
   ```

   * Ignore the other directories. Just make sure HousingData appears like below



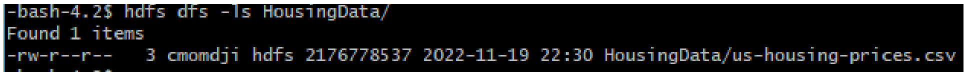
```
-bash-4.2$ hdfs dfs -mkdir HousingData
-bash-4.2$ hdfs dfs -ls
Found 6 items
drwxr-xr-x   - cmomdji hdfs          0 2022-11-09 02:23 .hiveJars
drwxr-xr-x   - cmomdji hdfs          0 2022-11-19 22:30 HousingData
drwxr-xr-x   - cmomdji hdfs          0 2022-11-09 02:25 dualcore
drwxr-xr-x   - cmomdji hdfs          0 2022-11-09 02:03 products
drwxr-xr-x   - cmomdji hdfs          0 2022-11-09 02:25 ratings
drwxr-xr-x   - cmomdji hdfs          0 2022-11-16 22:36 tmp
```

3. Next, we need to upload the data to HDFS:

```
hdfs dfs -put us-housing-prices.csv HousingData/
```

4. Make sure the file is uploaded:

```
hdfs dfs -ls HousingData/
```



```
-bash-4.2$ hdfs dfs -ls HousingData/
Found 1 items
-rw-r--r--   3 cmomdji hdfs 2176778537 2022-11-19 22:30 HousingData/us-housing-prices.csv
```

5. Once the file is uploaded to HDFS, make sure to delete it from Oracle:

```
rm -rf us-housing-prices.csv
```

Step 3: Create a database table using the data with Hive (Beeline)

With our data loaded in HDFS, we can turn to Hive to create the database table and query the data.

1. Connect to Hive using the ‘beeline’ command:

```
beeline
```

2. Remember to use your database. Replace ‘cmomdji’ with your username:

```
use cmomdji;
```

3. Next, we'll create the table with all the fields from the file. Again, make sure to replace cmomdji with your username:

```
DROP TABLE IF EXISTS housing_data;

CREATE EXTERNAL TABLE housing_data(state STRING, property_zip5
STRING, property_street_address STRING, property_city STRING,
property_county STRING, property_id STRING, sale_datetime STRING,
property_type STRING, sale_price double, seller_1_name STRING,
buyer_1_name STRING, building_num_units double,
building_year_built double, source_url STRING, book STRING, page
STRING,transfer_deed_type STRING, property_township STRING,
property_lat STRING, property_lon STRING, sale_id STRING,
deed_date STRING, building_num_stories double, building_num_beds
double, building_num_baths double, building_area_sqft STRING,
building_assessed_value double, building_assessed_date STRING,
land_area_acres STRING, land_area_sqft STRING,
land_assessed_value STRING, seller_2_name STRING, buyer_2_name
STRING, land_assessed_date STRING, seller_1_state STRING,
seller_2_state STRING, buyer_1_state STRING, buyer_2_state
STRING)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

STORED AS TEXTFILE LOCATION '/user/cmomdji/HousingData/'

TBLPROPERTIES("skip.header.line.count"="1");
```

4. Make a test query to see if the table was created with all our data:

```
select state, property_street_address, property_city,
property_county, sale_price from housing_data where state='CA'
and sale_price > 10000 limit 10;
```

*** You should see a result similar to the one below**

```
0: jdbc:hive2://bigdaiwn0.sub02180640120.trai>
```

Step 4: Create a new table with filtered data and save the data to a file

Now that our data is loaded into a database table in Hive, we can filter out the data and save it to a file for us to use later. **Remember to replace cmomdji with your username!**

1. Create a new table that filters our data to only properties in California and sold in 2019 or later. We also want to make sure the sale price is over $100,000 to show realistic housing costs and have only residential properties such as single-family homes and condominiums, rather than, for example, vacant lots or office spaces:

   ```
   DROP TABLE IF EXISTS california_housing_records;

   CREATE TABLE california_housing_records

   ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

   STORED AS TEXTFILE LOCATION
   '/user/cmomdji/HousingData/california_housing_records/' AS SELECT
   state, property_street_address, property_city, property_county,
   sale_price, property_type, sale_datetime FROM housing_data WHERE
   state='CA' AND sale_datetime >= '2019-01-01 00:00:00' AND
   sale_price > 100000 AND (property_type LIKE '%RESIDENCE%' OR
   property_type LIKE '%CONDO%');
   ```

2. Check to see if the file was saved to HDFS (switch to the Oracle terminal):

   ```
   hdfs dfs -ls HousingData/california_housing_records
   ```

3. Download the file from HDFS to Oracle:

   ```
   hdfs dfs –get
   /user/cmomdji/HousingData/california_housing_records/000000_0
   california_housing_records.csv
   ```

4. Verify that the file is there:

```
ls -al
```

5. Finally, open a new terminal and go to a directory where you would like to download the file. Once there, enter the following command to download the file to your computer (note the '.' at the end):

```
scp
cmomdji@144.24.14.145:/home/cmomdji/california_housing_records.csv .
```

6. You should now see the file in your current directory. This will be needed for the next section.

Step 5: Visualization 1 (Average Price per City)

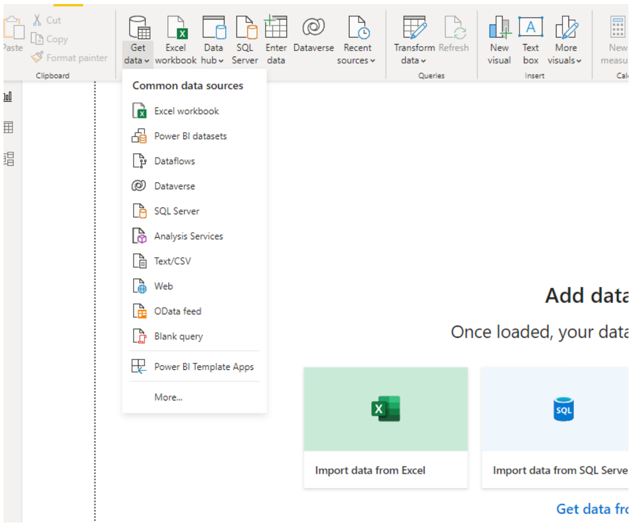
Now that we have our data filtered and downloaded to a new file on our computer, we can start to create useful visualizations to find helpful information about the data we have. The graphs we'll make are going to be simple and straightforward, but they apply useful skills in gathering data, filtering through it and highlighting information about it that can potentially be helpful to us. This section requires Excel and PowerBI.

1. Open the file you downloaded to your computer in Excel. Add a column at the top of the file above the first row and enter the following titles for each column:

   state, **street_address**, **city**, **county**, **sale_price**, **property_type**, **sale_datetime**

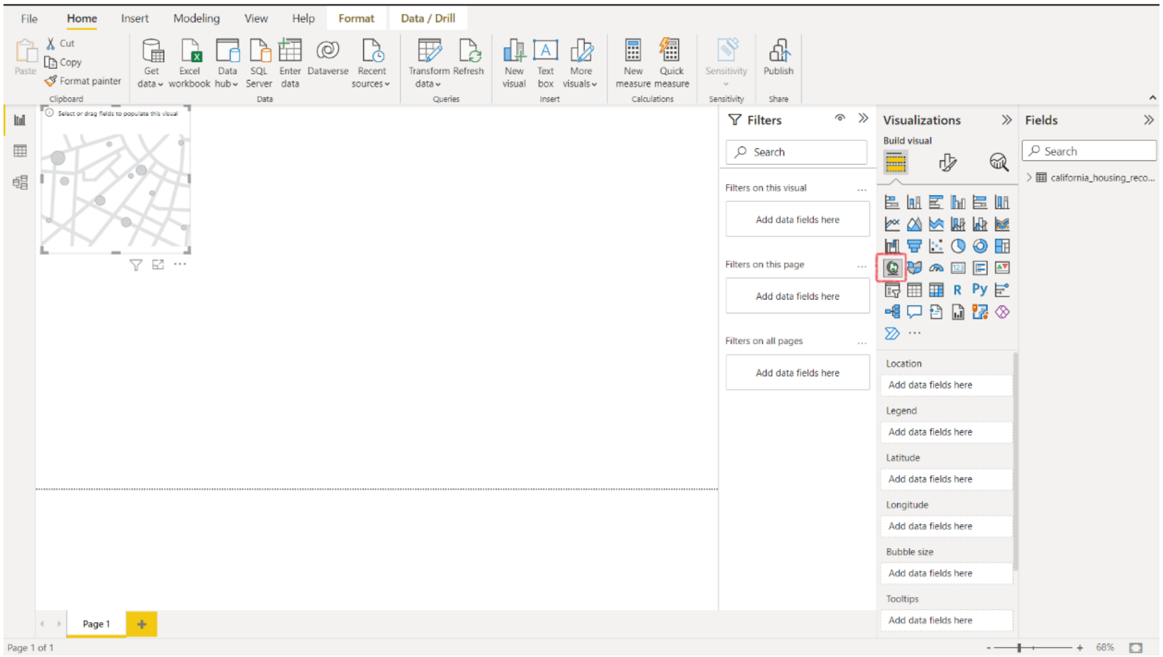
2. Download PowerBI: https://powerbi.microsoft.com/en-us/downloads/

3. Once PowerBI is downloaded on your computer, open it and select the “Get data” dropdown on the top left corner. From the dropdown, select “Text/CSV”.

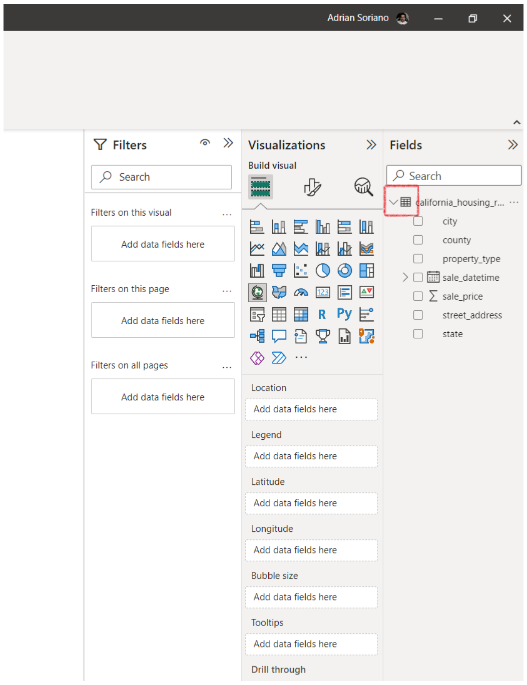


4. Select the **california_housing_records.csv** file we downloaded earlier. Click “Open”, ensure the column names appear above the appropriate data in the preview, then click “Load”.

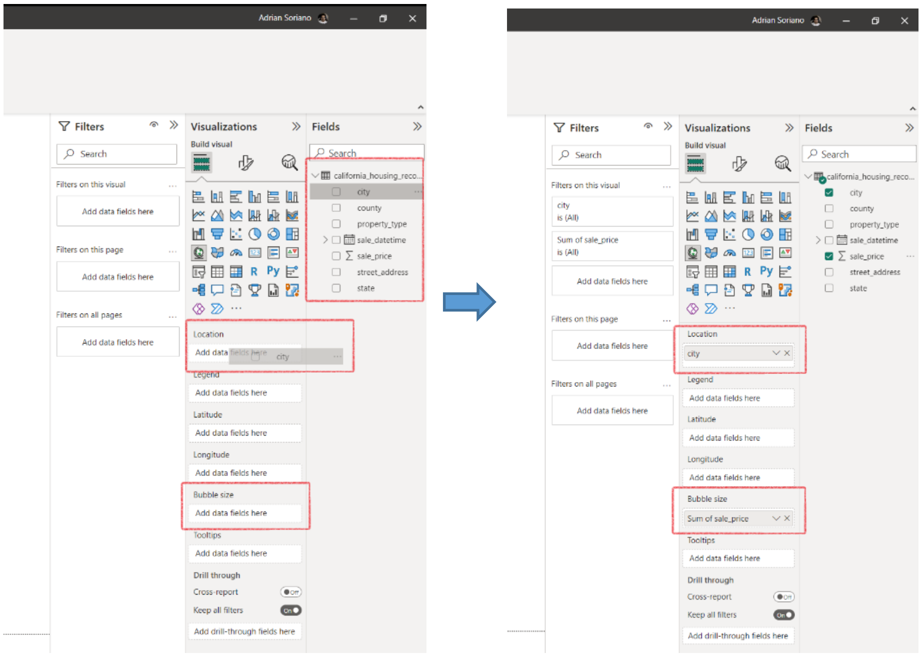
5. Click on the map icon to begin setting up the visualization



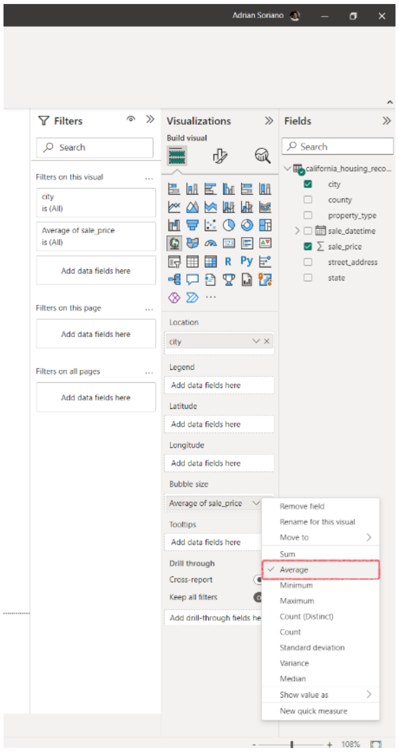
6. Click on the arrow on the left of the california_housing_records dataset to expand the menu and view the fields available



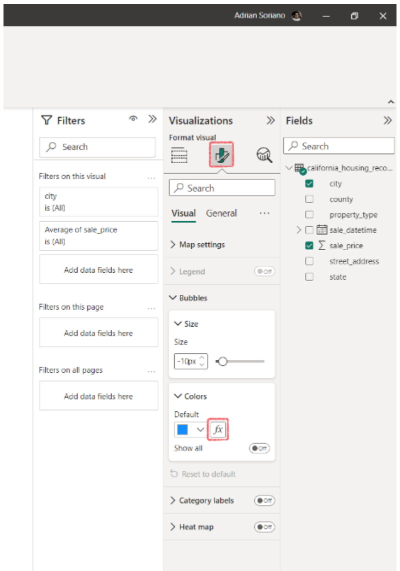
7. Drag the property_city field to the **Location** area and the sale_price field to the **Bubble size** area



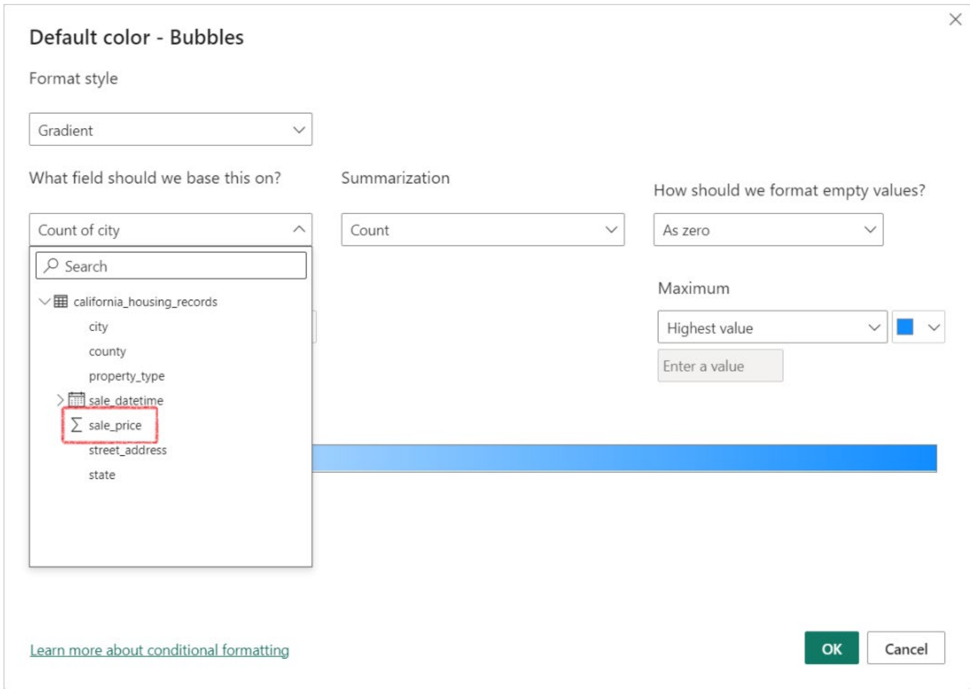
8. Use the drop-down menu under Bubble size to choose the **Average** of sale_price. This will help us show the average property price in each city within the county of Los Angeles.



9. To add gradient color to our bubble visualization, click on the format button under Visualizations, expand the **Bubbles** section, and click the conditional formatting button under **Colors**



10. In the window that pops up, select **sale_price** under the "What field should we base this on?" Field. Next, ensure that the bubbles are based on the **Average of sale_price,** the summarization is the **Average**, and the color set for Minimum to the **lightest shade of red** and set the color for Maximum to the **darkest shade of red**.



Default color - Bubbles

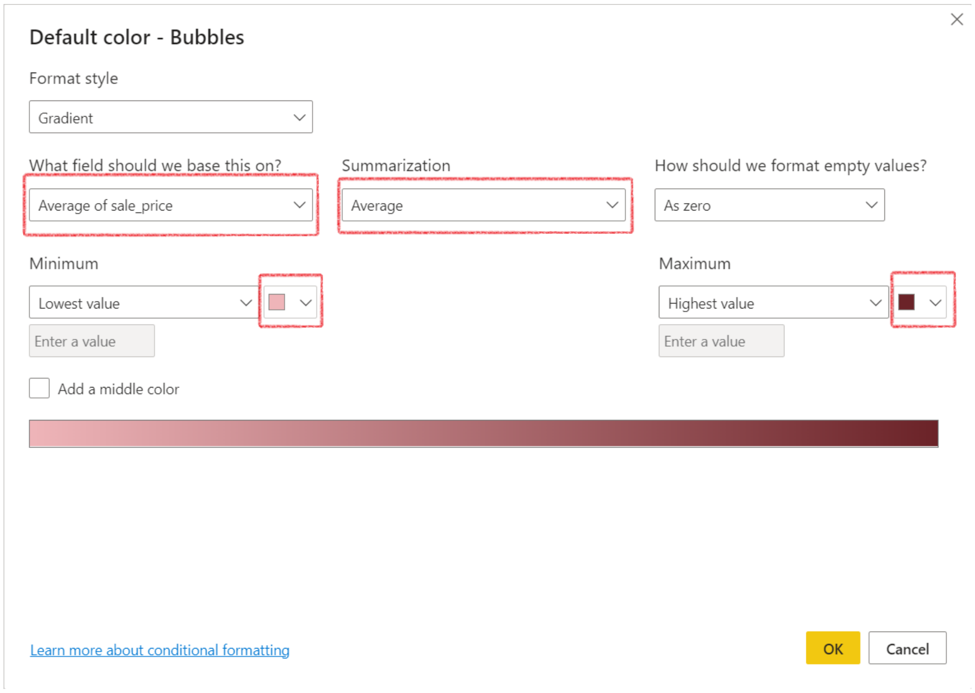
Format style: Gradient

What field should we base this on?: Count of city

Summarization: Count

How should we format empty values?: As zero

Maximum: Highest value



Default color - Bubbles

Format style: Gradient

What field should we base this on?: Average of sale_price

Summarization: Average

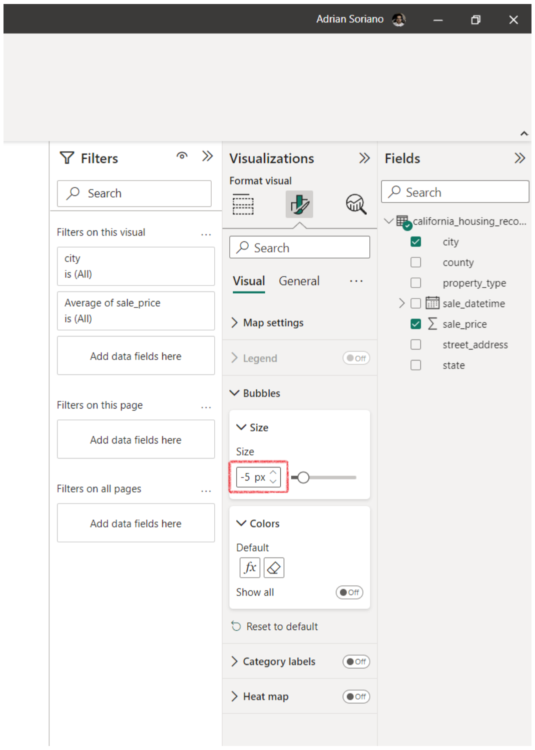
How should we format empty values?: As zero

Minimum: Lowest value

Maximum: Highest value

□ Add a middle color

11. Adjust the Bubble size to **-5px** to enhance the visual presented

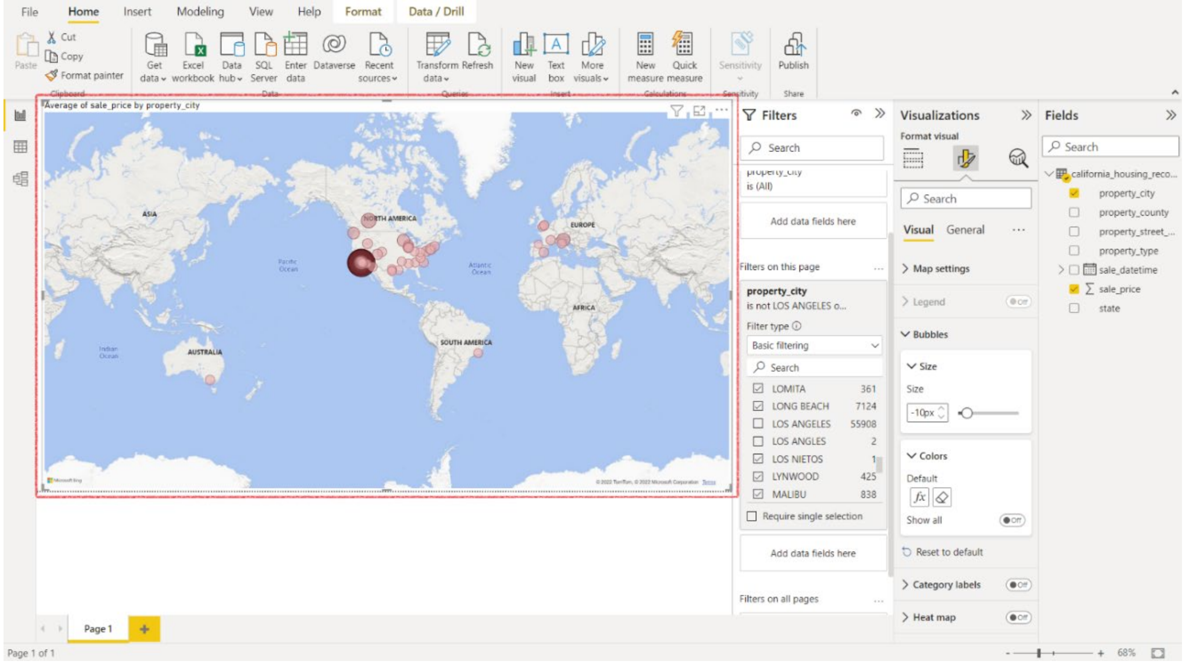


Filters on this visual
city
is (All)
Average of sale_price
is (All)

Size
-5 px

Fields: ☑ city, □ county, □ property_type, □ sale_datetime, ☑ sale_price, □ street_address, □ state

12. Expand the map using its corners and enlarge it until all the space is used to get a better visual. Other cities around the world are marked, but these are simply cities with the same name. We will be focusing on the cities in the Los Angeles County area.



Average of sale_price by property_city

property_city
is (All)

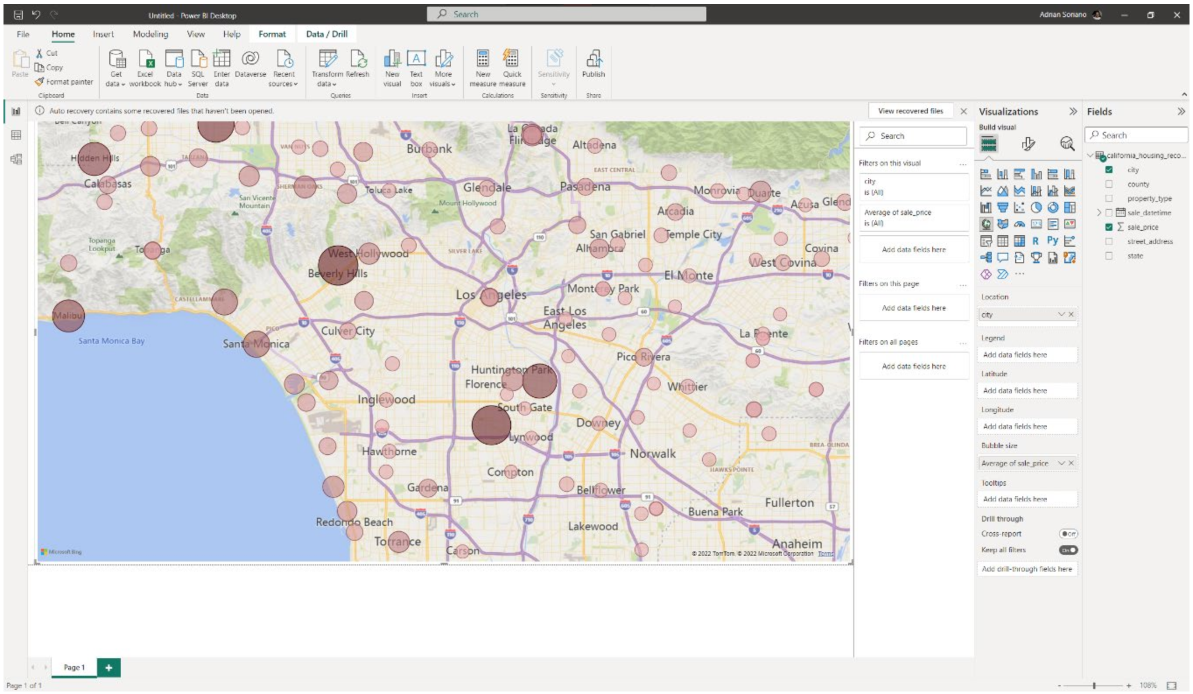
property_city
is not LOS ANGELES o...
Filter type
Basic filtering

☑ LOMITA 361
☑ LONG BEACH 7124
□ LOS ANGELES 55908
□ LOS ANGLES 2
☑ LOS NIETOS 1
☑ LYNWOOD 425
☑ MALIBU 838
□ Require single selection

Size
-10px

Fields: ☑ property_city, □ property_county, □ property_street_..., □ property_type, □ sale_datetime, ☑ sale_price, □ state

13. Zoom in by scrolling the middle click-wheel forward to get a clear view of Los Angeles County. The cities with the largest and darkest colored circles contain the most expensive properties currently on the market.



Filters on this visual

city
is (All)

Average of sale_price
is (All)

Location
city

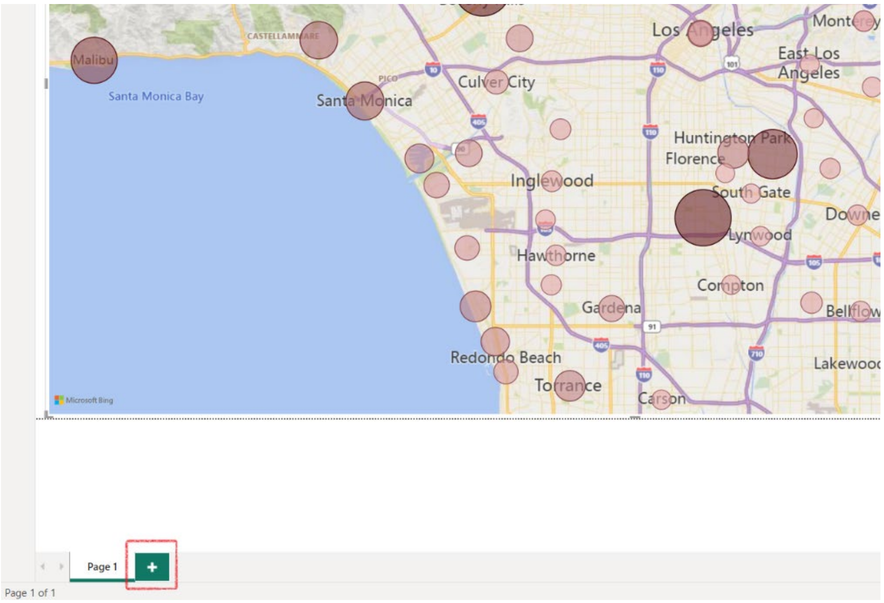
Bubble size
Average of sale_price

14. Click on File-> Export, then **Export to PDF** to save a copy of the data currently being viewed.

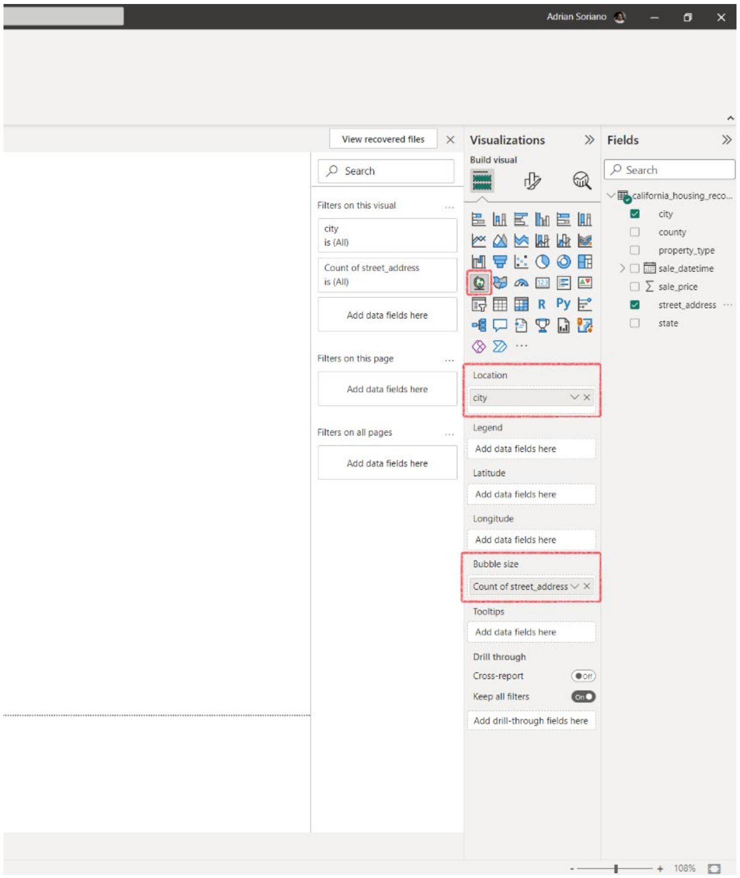
Step 6: Visualization 2 (Number of Properties for Sale per City)

We have created one visualization with the data. Now, using the same file in PowerBI, we will use the data pulled from HDFS to create a visualization that will show us a new perspective.

1. Using the same file from the previous step, we will create a new sheet, a new map, and adjust the settings to prepare the visualization. Begin by clicking on the **+** symbol at the bottom next to the name of the current sheet to create a new blank sheet.



2. In the new sheet, click on the map icon to create a map visualization, drag the **city** field from Fields panel to the Location area, and drag the **street_address** field to the Bubble size area as shown below.



Filters on this visual: city is (All); Count of street_address is (All)

Location: city

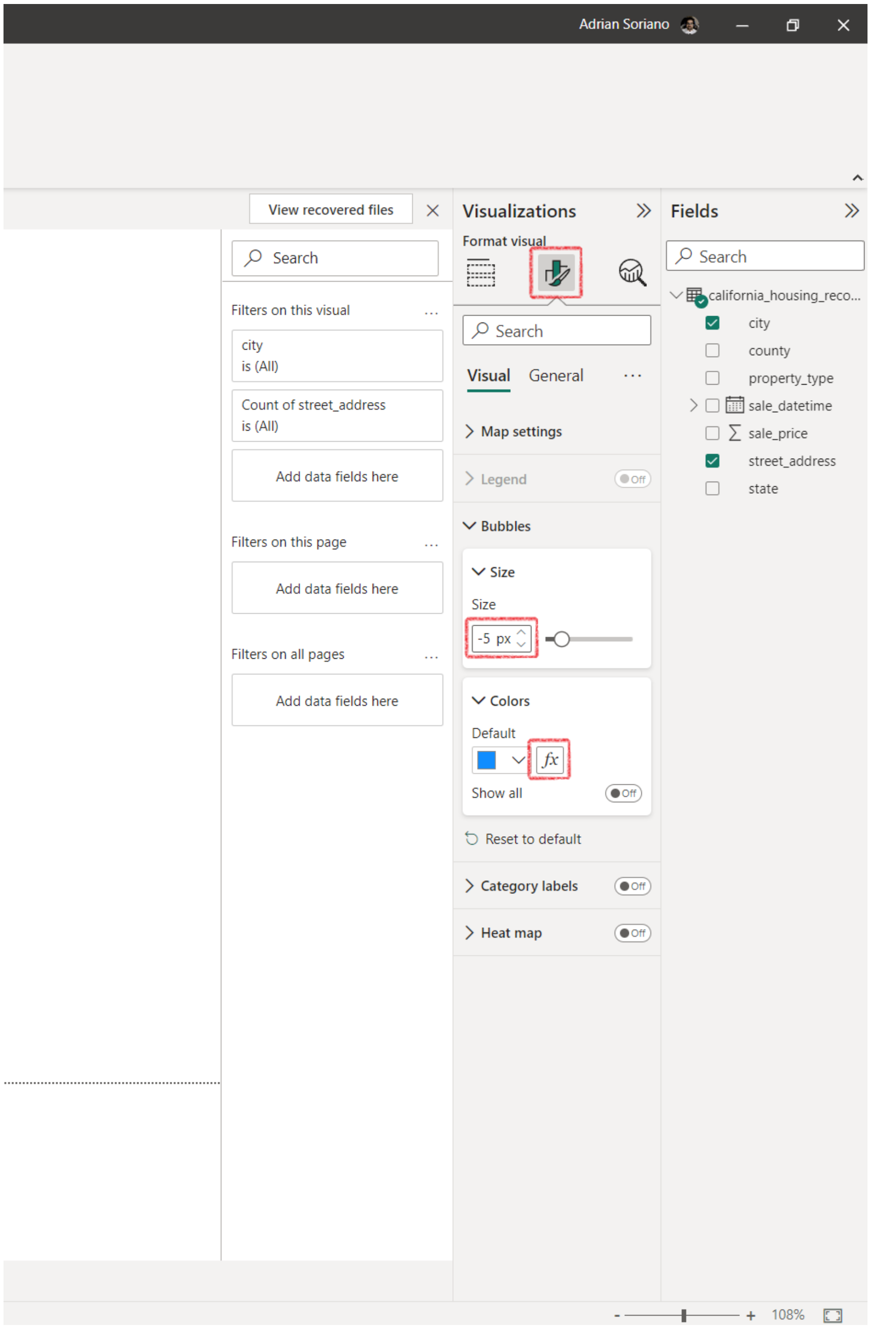
Bubble size: Count of street_address

Cross-report: Off

Keep all filters: On

california_housing_reco...
☑ city
□ county
□ property_type
□ sale_datetime
□ ∑ sale_price
☑ street_address
□ state

3. Like in our previous visualization, we need to adjust the bubble size and color conditions. Click on the Format button under Visualizations (highlighted below), change the bubble size to **-5px**, and lastly, click on the **conditional formatting** button under the Colors section to begin formatting the conditions for the bubble colors on the map.



Filters on this visual

city
is (All)

Count of street_address
is (All)

Size: -5 px

Colors
Default

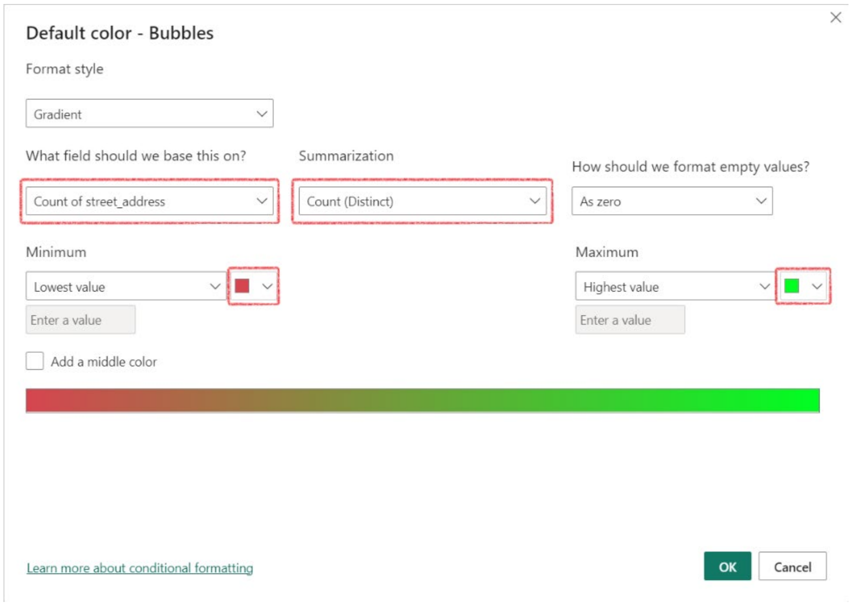
Show all: Off

Legend: Off

Category labels: Off

Heat map: Off

4. In the pop-up window, select **street_address** under “What field should we base this on?”, select **Count(Distinct)** under Summarization (Count should work as well as no address should repeat in the data), and set the Minimum color to **Red** and the Maximum to **Green**



Default color - Bubbles

Format style

Gradient

What field should we base this on?

Count of street_address

Summarization

Count (Distinct)

How should we format empty values?

As zero

Minimum

Lowest value 🟥

Enter a value

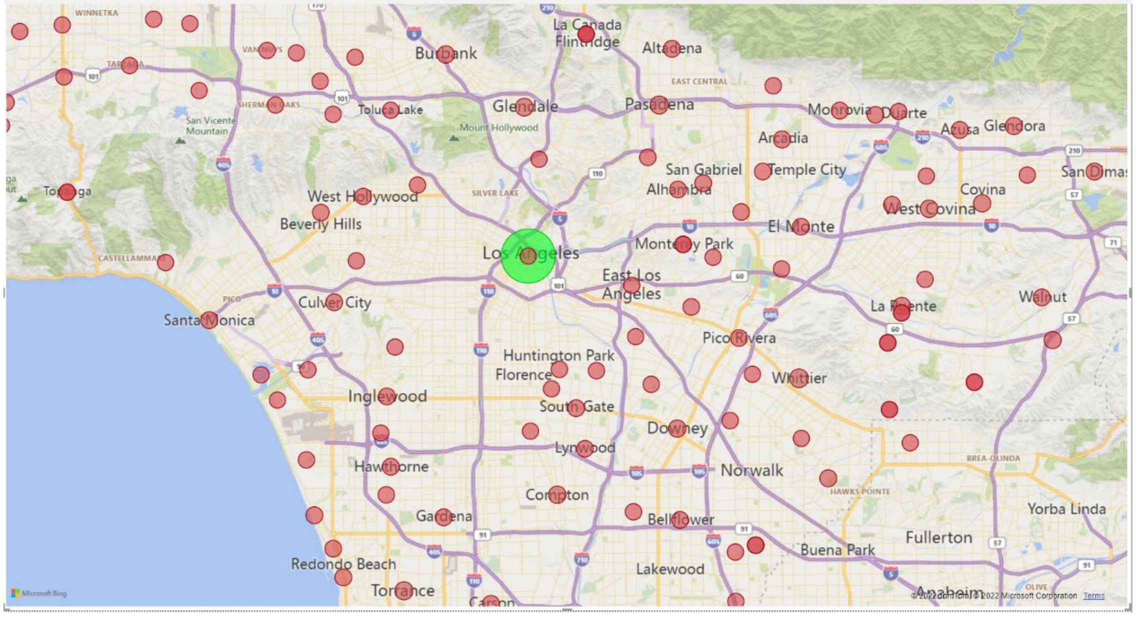
Maximum

Highest value 🟩

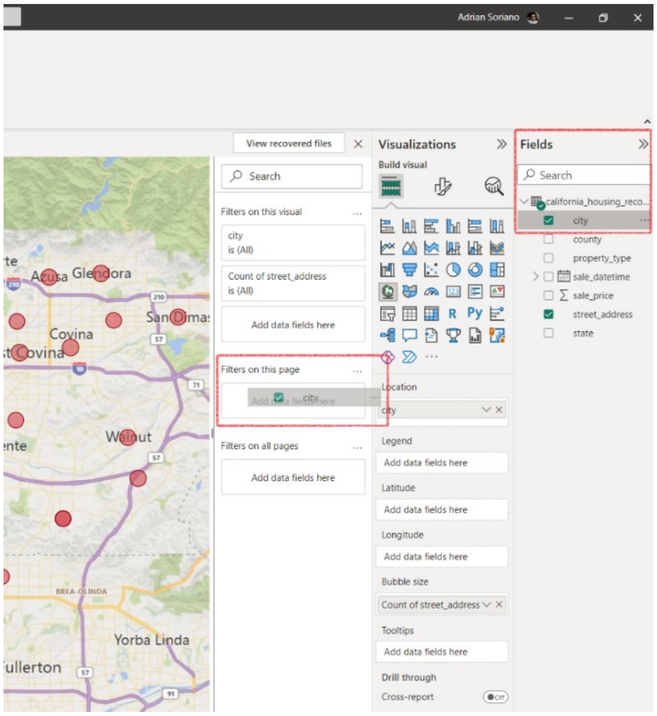
Enter a value

☐ Add a middle color

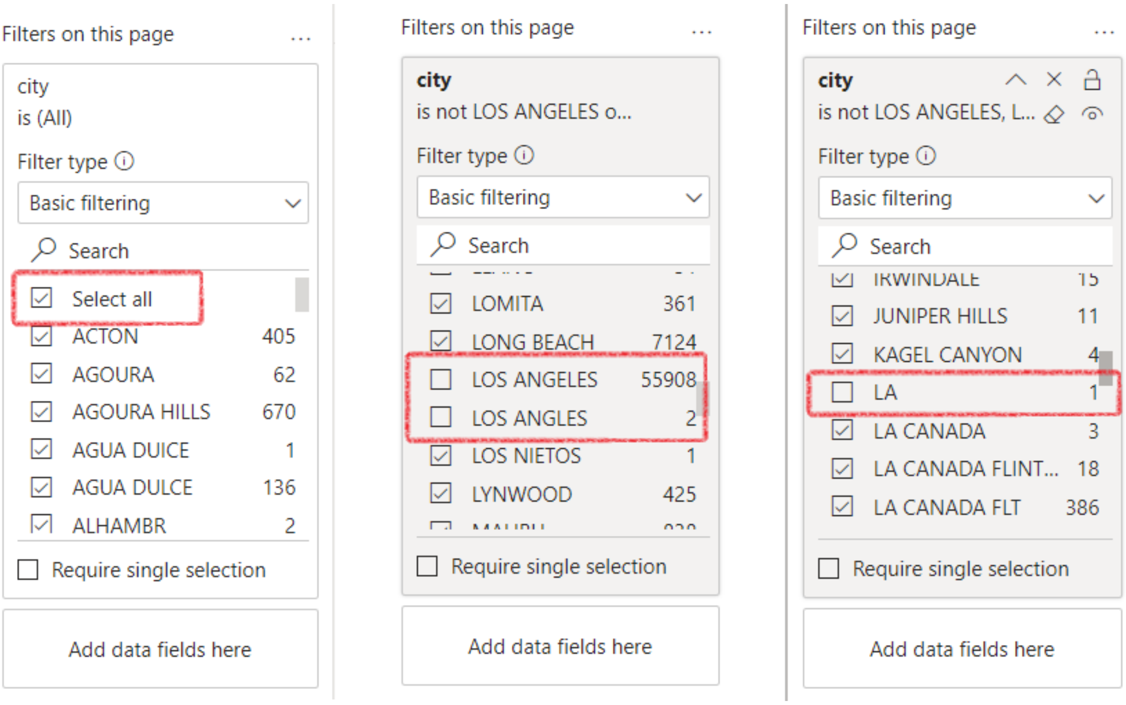
5. Expand the map to see the visual created. We got the desired result, but the data is obscured by a huge number of properties attributed to Los Angeles. To get a better visual of the surrounding areas, let’s filter these numbers out.



6. Drag the **city** field to the Filters on this page section.



7. Select the **Select all** option to highlight all cities. From here, deselect **Los Angeles** as well as incorrectly labeled cities **Los Angles**, and **La**. You may have to type the cities into the search bar to find them.



Filters on this page ...

city
is (All)

Filter type ⓘ

Basic filtering

Search

☑ Select all

☑ ACTON	405
☑ AGOURA	62
☑ AGOURA HILLS	670
☑ AGUA DUICE	1
☑ AGUA DULCE	136
☑ ALHAMBR	2

□ Require single selection

Add data fields here

Filters on this page ...

city
is not LOS ANGELES o...

Filter type ⓘ

Basic filtering

Search

☑ LOMITA	361
☑ LONG BEACH	7124
□ LOS ANGELES	55908
□ LOS ANGLES	2
☑ LOS NIETOS	1
☑ LYNWOOD	425

□ Require single selection

Add data fields here

Filters on this page ...

city
is not LOS ANGELES, L...

Filter type ⓘ

Basic filtering

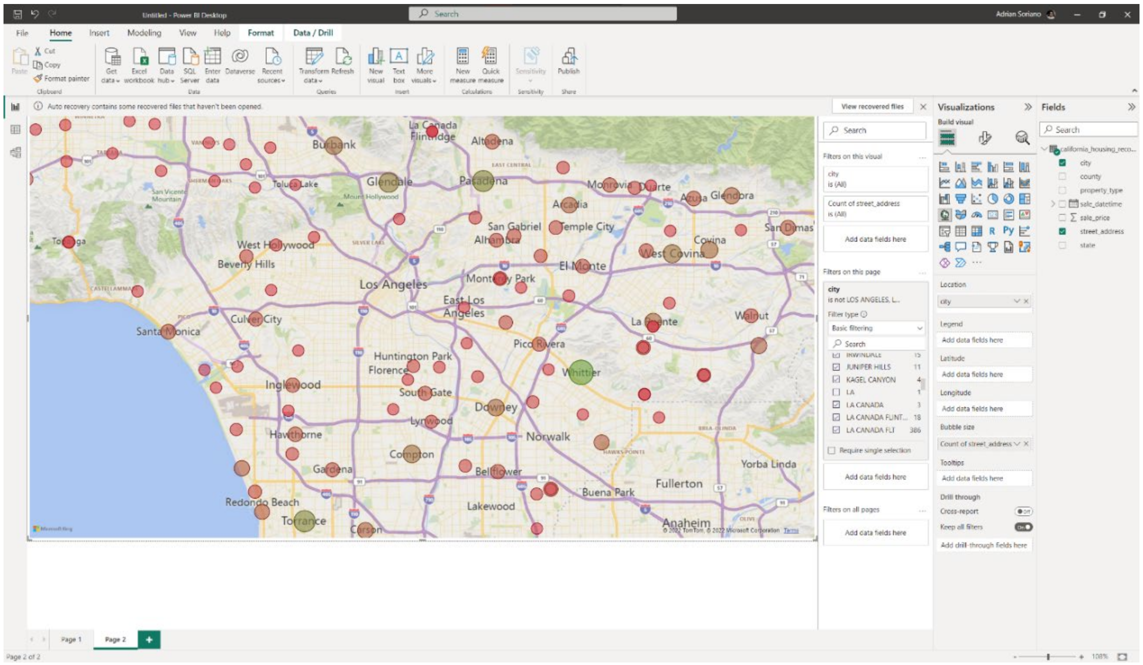
Search

☑ IRWINDALE	15
☑ JUNIPER HILLS	11
☑ KAGEL CANYON	4
□ LA	1
☑ LA CANADA	3
☑ LA CANADA FLINT...	18
☑ LA CANADA FLT	386

□ Require single selection

Add data fields here

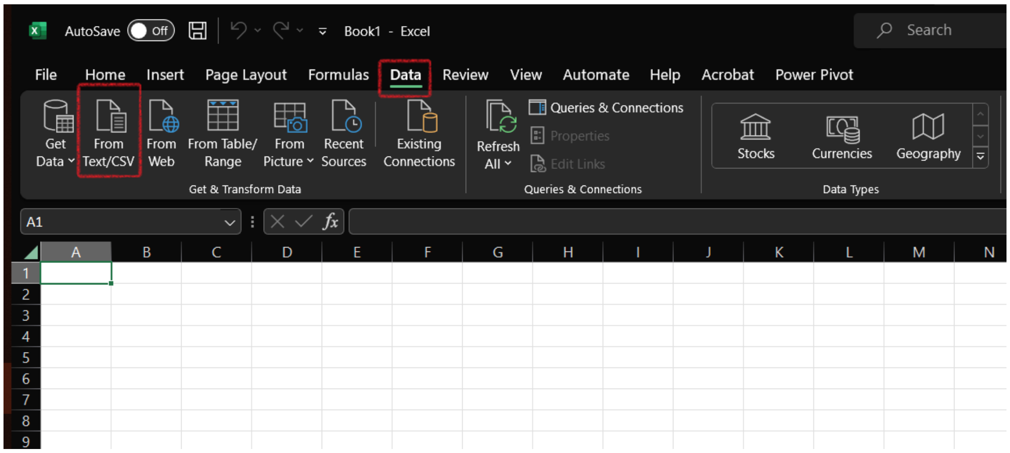
8. The newly filtered data provides a more balanced view.



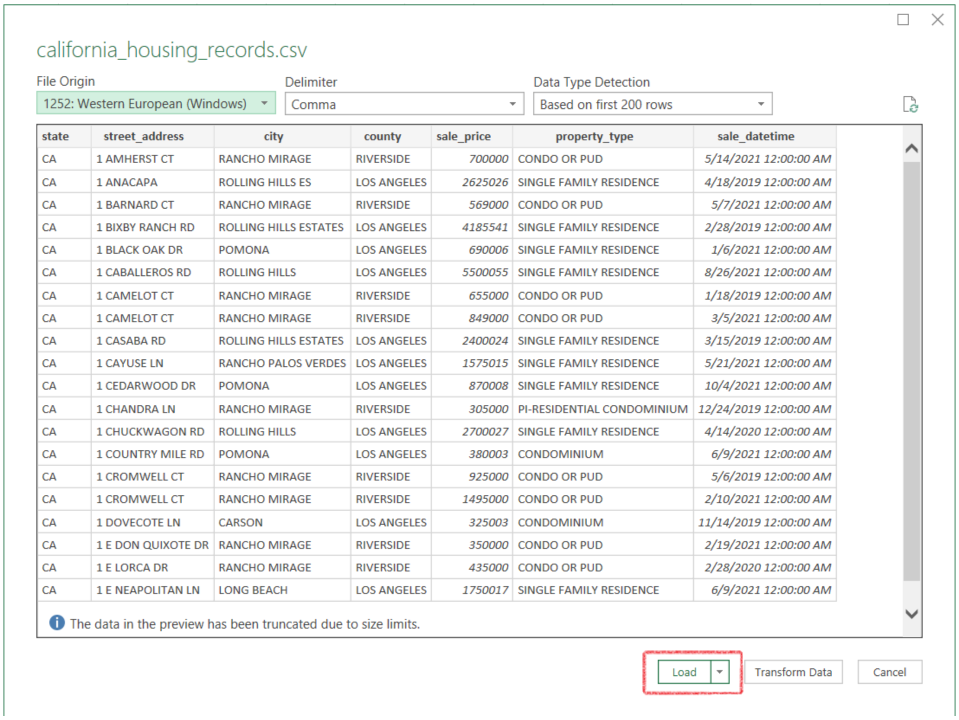
Step 7: Visualization 3 (Temporal Visualization)

We have created two different visuals using PowerBI and a map as a tool. Now we will take our analysis further and use the data we downloaded from HDFS to create a new file in Excel and use their visualization tools to create a Temporal Visualization based on Housing prices from 2019-2021.

1. Open Excel. Under the Data Tab, select **From Text/CSV** to select a file from your computer to use. Select **california_housing_records.csv**



2. In the pop-up window, review the data being uploaded. It should look the same as we are using the same file. Click Load.



california_housing_records.csv

File Origin: 1252: Western European (Windows)

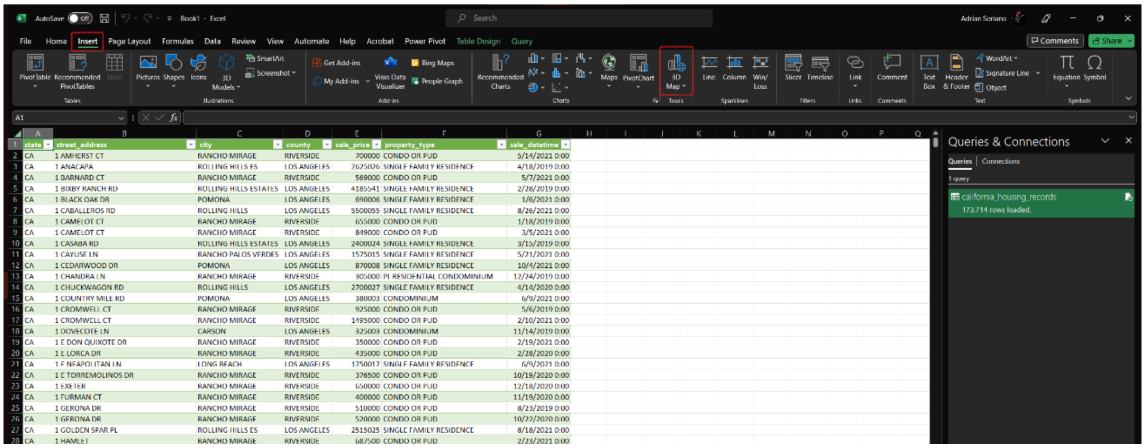
Delimiter: Comma

Data Type Detection: Based on first 200 rows

state	street_address	city	county	sale_price	property_type	sale_datetime
CA	1 AMHERST CT	RANCHO MIRAGE	RIVERSIDE	700000	CONDO OR PUD	5/14/2021 12:00:00 AM
CA	1 ANACAPA	ROLLING HILLS ES	LOS ANGELES	2625026	SINGLE FAMILY RESIDENCE	4/18/2019 12:00:00 AM
CA	1 BARNARD CT	RANCHO MIRAGE	RIVERSIDE	569000	CONDO OR PUD	5/7/2021 12:00:00 AM
CA	1 BIXBY RANCH RD	ROLLING HILLS ESTATES	LOS ANGELES	4185541	SINGLE FAMILY RESIDENCE	2/28/2019 12:00:00 AM
CA	1 BLACK OAK DR	POMONA	LOS ANGELES	690006	SINGLE FAMILY RESIDENCE	1/6/2021 12:00:00 AM
CA	1 CABALLEROS RD	ROLLING HILLS	LOS ANGELES	5500055	SINGLE FAMILY RESIDENCE	8/26/2021 12:00:00 AM
CA	1 CAMELOT CT	RANCHO MIRAGE	RIVERSIDE	655000	CONDO OR PUD	1/18/2019 12:00:00 AM
CA	1 CAMELOT CT	RANCHO MIRAGE	RIVERSIDE	849000	CONDO OR PUD	3/5/2021 12:00:00 AM
CA	1 CASABA RD	ROLLING HILLS ESTATES	LOS ANGELES	2400024	SINGLE FAMILY RESIDENCE	3/15/2019 12:00:00 AM
CA	1 CAYUSE LN	RANCHO PALOS VERDES	LOS ANGELES	1575015	SINGLE FAMILY RESIDENCE	5/21/2021 12:00:00 AM
CA	1 CEDARWOOD DR	POMONA	LOS ANGELES	870008	SINGLE FAMILY RESIDENCE	10/4/2021 12:00:00 AM
CA	1 CHANDRA LN	RANCHO MIRAGE	RIVERSIDE	305000	PI-RESIDENTIAL CONDOMINIUM	12/24/2019 12:00:00 AM
CA	1 CHUCKWAGON RD	ROLLING HILLS	LOS ANGELES	2700027	SINGLE FAMILY RESIDENCE	4/14/2020 12:00:00 AM
CA	1 COUNTRY MILE RD	POMONA	LOS ANGELES	380003	CONDOMINIUM	6/9/2021 12:00:00 AM
CA	1 CROMWELL CT	RANCHO MIRAGE	RIVERSIDE	925000	CONDO OR PUD	5/6/2019 12:00:00 AM
CA	1 CROMWELL CT	RANCHO MIRAGE	RIVERSIDE	1495000	CONDO OR PUD	2/10/2021 12:00:00 AM
CA	1 DOVECOTE LN	CARSON	LOS ANGELES	325003	CONDOMINIUM	11/14/2019 12:00:00 AM
CA	1 E DON QUIXOTE DR	RANCHO MIRAGE	RIVERSIDE	350000	CONDO OR PUD	2/19/2021 12:00:00 AM
CA	1 E LORCA DR	RANCHO MIRAGE	RIVERSIDE	435000	CONDO OR PUD	2/28/2020 12:00:00 AM
CA	1 E NEAPOLITAN LN	LONG BEACH	LOS ANGELES	1750017	SINGLE FAMILY RESIDENCE	6/9/2021 12:00:00 AM

ⓘ The data in the preview has been truncated due to size limits.

3. Under the Insert tab, click on the 3D Map button to access the map visualization tools.



	A	B	C	D	E	F	G
1	state	street_address	city	county	sale_price	property_type	sale_datetime
2	CA	1 AMHERST CT	RANCHO MIRAGE	RIVERSIDE	700000	CONDO OR PUD	5/14/2021 0:00
3	CA	1 ANACAPA	ROLLING HILLS ES	LOS ANGELES	2625026	SINGLE FAMILY RESIDENCE	4/18/2019 0:00
4	CA	1 BARNARD CT	RANCHO MIRAGE	RIVERSIDE	569000	CONDO OR PUD	5/7/2021 0:00
5	CA	1 BIXBY RANCH RD	ROLLING HILLS ESTATES	LOS ANGELES	4185541	SINGLE FAMILY RESIDENCE	2/28/2019 0:00
6	CA	1 BLACK OAK DR	POMONA	LOS ANGELES	690006	SINGLE FAMILY RESIDENCE	1/6/2021 0:00
7	CA	1 CABALLEROS RD	ROLLING HILLS	LOS ANGELES	5500055	SINGLE FAMILY RESIDENCE	8/26/2021 0:00
8	CA	1 CAMELOT CT	RANCHO MIRAGE	RIVERSIDE	655000	CONDO OR PUD	1/18/2019 0:00
9	CA	1 CAMELOT CT	RANCHO MIRAGE	RIVERSIDE	849000	CONDO OR PUD	3/5/2021 0:00
10	CA	1 CASABA RD	ROLLING HILLS ESTATES	LOS ANGELES	2400024	SINGLE FAMILY RESIDENCE	3/15/2019 0:00
11	CA	1 CAYUSE LN	RANCHO PALOS VERDES	LOS ANGELES	1575015	SINGLE FAMILY RESIDENCE	5/21/2021 0:00
12	CA	1 CEDARWOOD DR	POMONA	LOS ANGELES	870008	SINGLE FAMILY RESIDENCE	10/4/2021 0:00
13	CA	1 CHANDRA LN	RANCHO MIRAGE	RIVERSIDE	305000	PI-RESIDENTIAL CONDOMINIUM	12/24/2019 0:00
14	CA	1 CHUCKWAGON RD	ROLLING HILLS	LOS ANGELES	2700027	SINGLE FAMILY RESIDENCE	4/14/2020 0:00
15	CA	1 COUNTRY MILE RD	POMONA	LOS ANGELES	380003	CONDOMINIUM	6/9/2021 0:00
16	CA	1 CROMWELL CT	RANCHO MIRAGE	RIVERSIDE	925000	CONDO OR PUD	5/6/2019 0:00
17	CA	1 CROMWELL CT	RANCHO MIRAGE	RIVERSIDE	1495000	CONDO OR PUD	2/10/2021 0:00
18	CA	1 DOVECOTE LN	CARSON	LOS ANGELES	325003	CONDOMINIUM	11/14/2019 0:00
19	CA	1 E DON QUIXOTE DR	RANCHO MIRAGE	RIVERSIDE	350000	CONDO OR PUD	2/19/2021 0:00
20	CA	1 E LORCA DR	RANCHO MIRAGE	RIVERSIDE	435000	CONDO OR PUD	2/28/2020 0:00
21	CA	1 E NEAPOLITAN LN	LONG BEACH	LOS ANGELES	1750017	SINGLE FAMILY RESIDENCE	6/9/2021 0:00
22	CA	1 E TORREMOLINOS DR	RANCHO MIRAGE	RIVERSIDE	376500	CONDO OR PUD	10/19/2020 0:00
23	CA	1 EXETER	RANCHO MIRAGE	RIVERSIDE	650000	CONDO OR PUD	12/18/2020 0:00
24	CA	1 FURMAN CT	RANCHO MIRAGE	RIVERSIDE	400000	CONDO OR PUD	11/19/2020 0:00
25	CA	1 GERONA DR	RANCHO MIRAGE	RIVERSIDE	510000	CONDO OR PUD	8/23/2019 0:00
26	CA	1 GERONA DR	RANCHO MIRAGE	RIVERSIDE	520000	CONDO OR PUD	10/22/2020 0:00
27	CA	1 GOLDEN SPAR PL	ROLLING HILLS ES	LOS ANGELES	2515025	SINGLE FAMILY RESIDENCE	8/18/2021 0:00
28	CA	1 HAMLET	RANCHO MIRAGE	RIVERSIDE	687500	CONDO OR PUD	2/23/2021 0:00

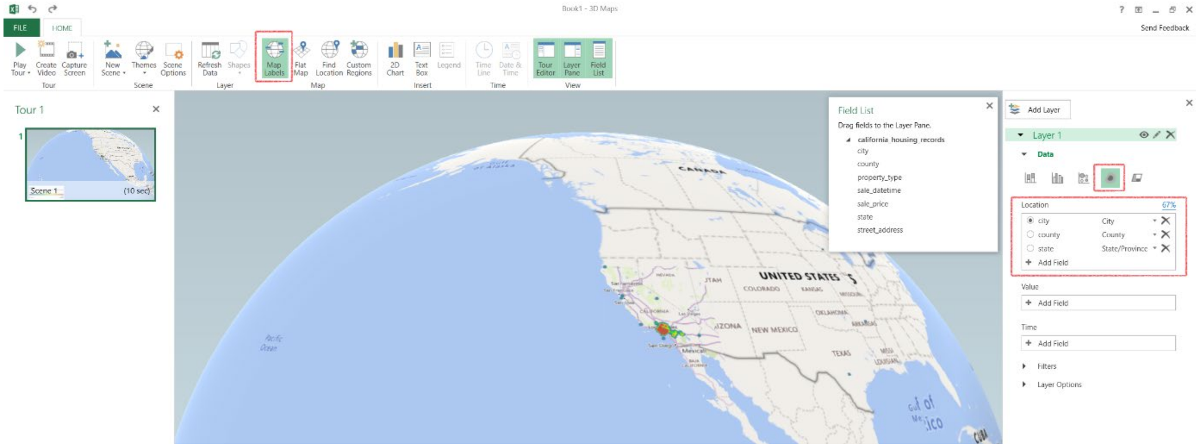
Queries & Connections

Queries | Connections

1 query

california_housing_records
173,714 rows loaded.

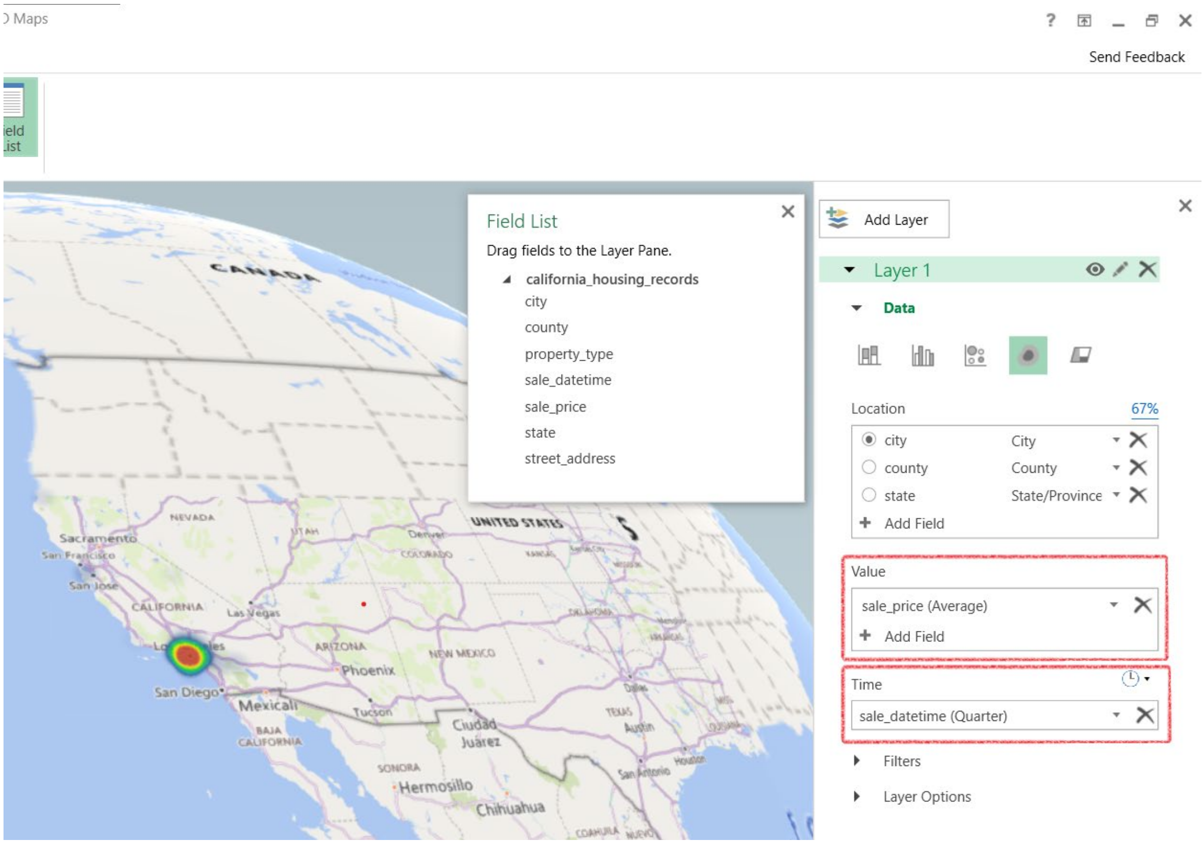
4. In the new window displaying the map, we will make our adjustments to create the visualization. Begin by clicking on **Map Labels** under the Home tab. This will display the names of Cities and States on the map. Then, ensure **city** is chosen as the data option for Location and click on the **Heat Map** icon, as we will be creating a heat map with our data.



Field List
Drag fields to the Layer Pane.
california_housing_records
city
county
property_type
sale_datetime
sale_price
state
street_address

Layer 1
Data
Location 67%
● city City
○ county County
○ state State/Province
Add Field
Value
Add Field
Time
Add Field
Filters
Layer Options

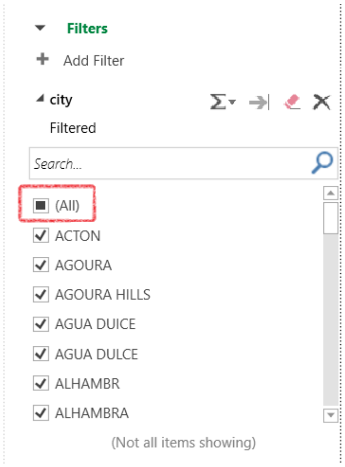
5. Now we need to value to the intensity of the heat. Under the Value section, select the field **sale_price** from our data options. Change the field to **sale_price(Average)** to display the average prices of properties per city. The field may be initially set to sale_price(Sum). Make sure to set it as shown below. Lastly, set the Time section to the only field available, **sale_datetime**. Change the field to **sale_datetime(Quarter)**. This will display quarterly changes in the data.



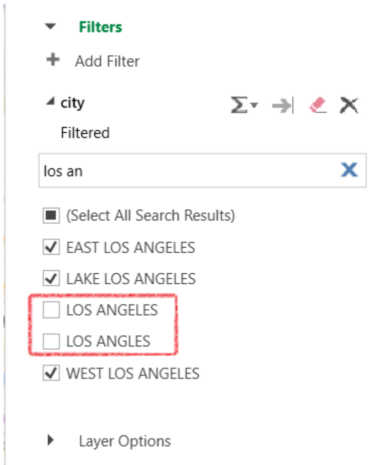
Field List
Drag fields to the Layer Pane.
california_housing_records
city
county
property_type
sale_datetime
sale_price
state
street_address

Layer 1
Data
Location 67%
● city City
○ county County
○ state State/Province
Add Field
Value
sale_price (Average)
Add Field
Time
sale_datetime (Quarter)
Filters
Layer Options

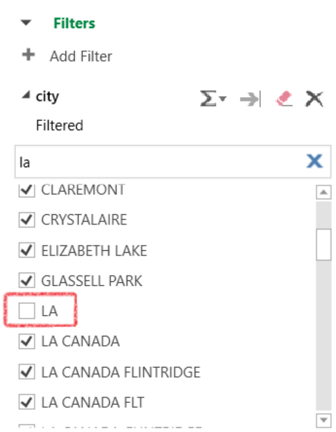
6. To make sure we get similar results to the previous visualizations, we will filter out properties from **Los Angeles** and focus on the surrounding areas. We will also filter out the misspelled **"Los Angles"** and **"LA".** Expand the Filters section and click **+ Add Filter**. This will open the fields available. Select **city,** click **(All)**, and deselect the specified cities as below. You will have to type into the search bar to find the values as there are a lot.



■ (All)
☑ ACTON
☑ AGOURA
☑ AGOURA HILLS
☑ AGUA DUICE
☑ AGUA DULCE
☑ ALHAMBR
☑ ALHAMBRA
(Not all items showing)

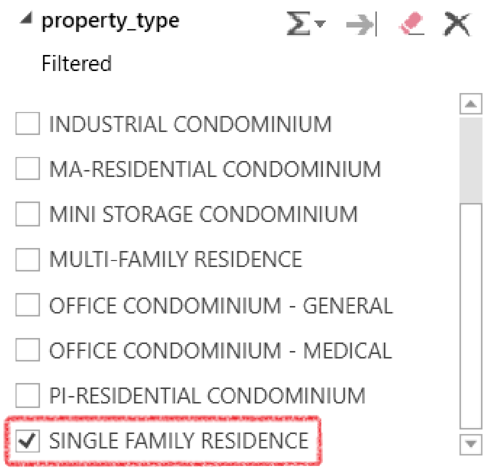


los an
■ (Select All Search Results)
☑ EAST LOS ANGELES
☑ LAKE LOS ANGELES
□ LOS ANGELES
□ LOS ANGLES
☑ WEST LOS ANGELES

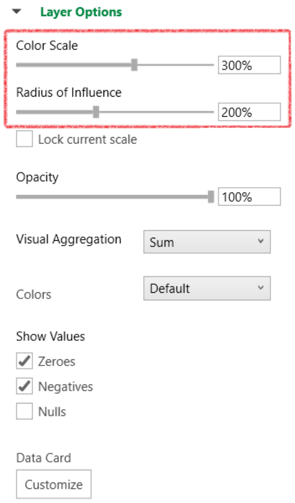


la
☑ CLAREMONT
☑ CRYSTALAIRE
☑ ELIZABETH LAKE
☑ GLASSELL PARK
□ LA
☑ LA CANADA
☑ LA CANADA FLINTRIDGE
☑ LA CANADA FLT

7. For this visual, we will focus only on single-family residences so we will add a filter accordingly. Click **+ Add Filter** once more. Click on property_type and select **SINGLE FAMILY RESIDENCE.** We will also enhance the visual a bit by increasing the color scale and the radius of influence. To do this, expand the Layer options section and increase the Color Scale to **300%** and the Radius of Influence to **200%.** Make these adjustments as shown below.

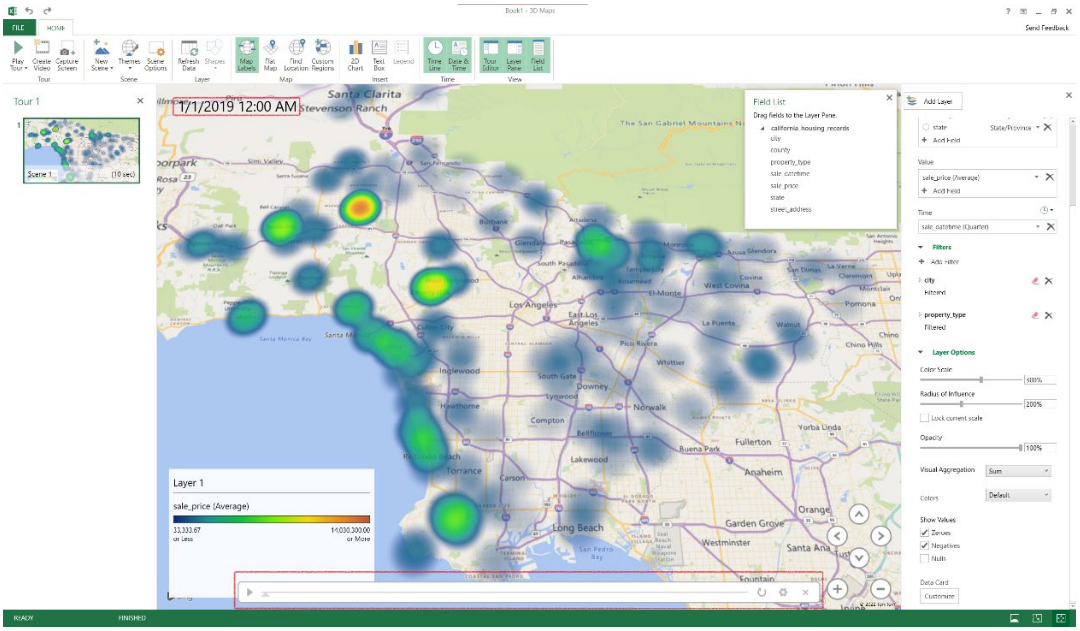


□ INDUSTRIAL CONDOMINIUM
□ MA-RESIDENTIAL CONDOMINIUM
□ MINI STORAGE CONDOMINIUM
□ MULTI-FAMILY RESIDENCE
□ OFFICE CONDOMINIUM - GENERAL
□ OFFICE CONDOMINIUM - MEDICAL
□ PI-RESIDENTIAL CONDOMINIUM
☑ SINGLE FAMILY RESIDENCE



Color Scale 300%
Radius of Influence 200%
□ Lock current scale
Opacity 100%
Visual Aggregation Sum
Colors Default
Show Values
☑ Zeroes
☑ Negatives
□ Nulls
Data Card

8. Zoom in on the map using the middle click-wheel to the Los Angeles area in California. The result will show the price variance across cities using heat intensity. The red areas have the highest price average while the blue (or cold) areas define the lowest prices. Green and yellow areas lie somewhere in between. Use the slider at the bottom to change the date of the map and see the price differences in any quarter between 2019-2021. You should see changes like the ones shown below.



1/1/2019 12:00 AM

Field List

Drag fields to the Layer Pane.

- california_housing_records
  - city
  - county
  - property_type
  - sale_datetime
  - sale_price
  - state
  - street_address

Add Layer

○ state State/Province

+ Add Field

Value

sale_price (Average)

+ Add Field

Time

sale_datetime (Quarter)

Filters

+ Add Filter

city
Filtered

property_type
Filtered

Layer Options

Color Scale 300%

Radius of Influence 200%

☐ Lock current scale

Opacity 100%

Visual Aggregation Sum

Colors Default

Show Values

☑ Zeroes

☑ Negatives

☐ Nulls

Data Card

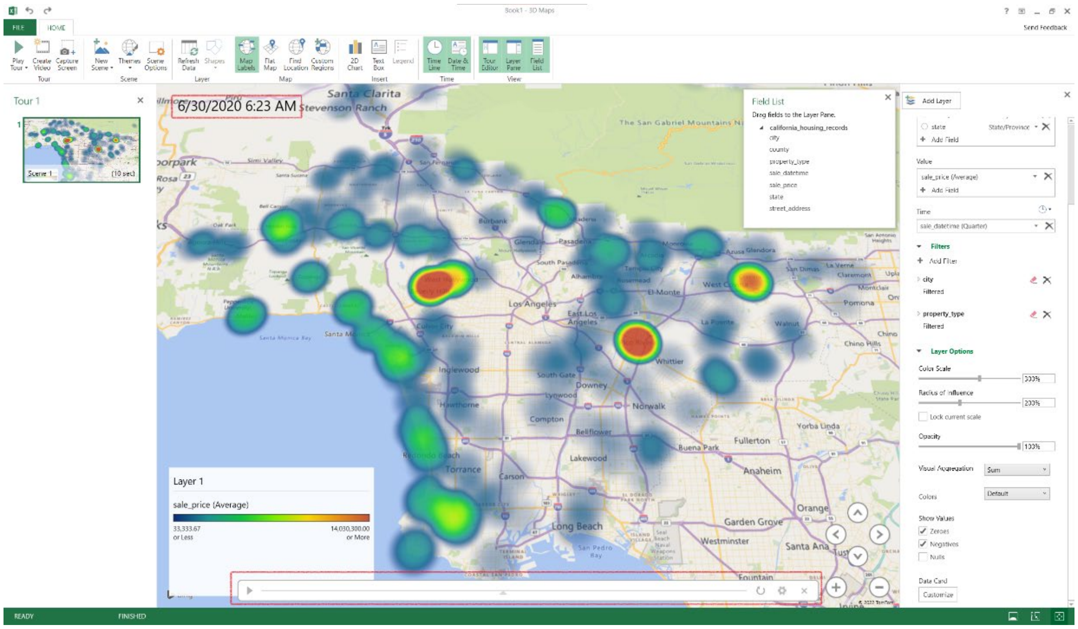
Customize

Layer 1

sale_price (Average)

33,333.67 or Less

14,030,300.00 or More



6/30/2020 6:23 AM

Field List

Drag fields to the Layer Pane.

- california_housing_records
  - city
  - county
  - property_type
  - sale_datetime
  - sale_price
  - state
  - street_address

Add Layer

○ state State/Province

+ Add Field

Value

sale_price (Average)

+ Add Field

Time

sale_datetime (Quarter)

Filters

+ Add Filter

city
Filtered

property_type
Filtered

Layer Options

Color Scale 300%

Radius of Influence 200%

☐ Lock current scale

Opacity 100%

Visual Aggregation Sum

Colors Default

Show Values

☑ Zeroes

☑ Negatives

☐ Nulls

Data Card

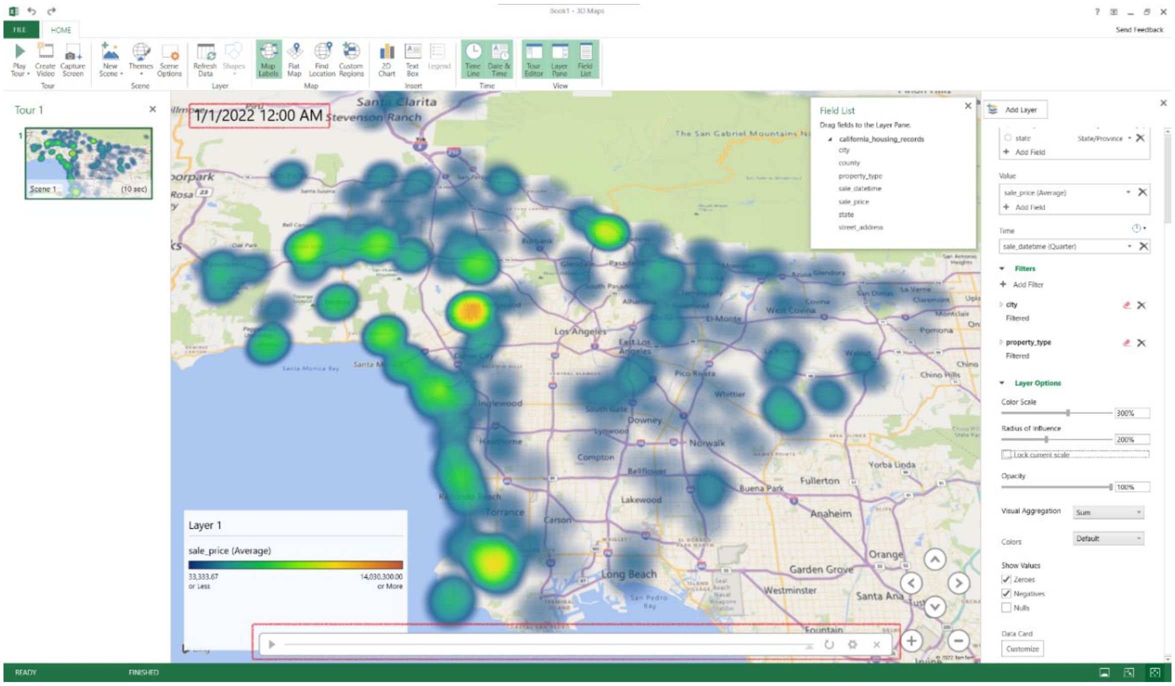
Customize

Layer 1

sale_price (Average)

33,333.67 or Less

14,030,300.00 or More



1/1/2022 12:00 AM

Field List

Drag fields to the Layer Pane.

- california_housing_records
  - city
  - county
  - property_type
  - sale_datetime
  - sale_price
  - state
  - street_address

Add Layer

○ state State/Province

+ Add Field

Value: sale_price (Average)

+ Add Field

Time: sale_datetime (Quarter)

Filters

+ Add Filter

city: Filtered

property_type: Filtered

Layer Options

Color Scale: 300%

Radius of Influence: 200%

□ Lock current scale

Opacity: 100%

Visual Aggregation: Sum

Colors: Default

Show Values

☑ Zeroes

☑ Negatives

□ Nulls

Data Card: Customize

Layer 1

sale_price (Average)

33,333.67 or Less — 14,030,300.00 or More

References

1. Dataset Source: https://www.dolthub.com/repositories/dolthub/us-housing-prices/data/main
2. Reduced Dataset: https://www.dropbox.com/s/3xt3up4fve78me8/dolthub_us-housing-prices_main_sales_REDUCED2.csv?dl=0
3. Github: https://github.com/cmomdji/CIS-4560-Team-2