Hands-on Activity – Data Cleaning

In this hands-on activity, you will explore several Microsoft Excel functions that will help you better understand the software and its applications. Specifically, in this hands-on activity, you will learn some basic Excel functions and tricks on performing data cleaning. In real life, the Excel dataset given to you may not necessarily be "perfect" and contain no errors or "dirty" data. Here are some techniques on how to ensure that your data is clean and ready for use.

About the Dataset

The dataset that will be cleaning contains information about Boba Tea Shops in San Francisco. Unfortunately, the data is not in the greatest shape.

What you'll do

You will clean the data and try to make it as perfect as possible. Through this process, you will get to explore several Microsoft Excel functions that are useful not just for data cleaning but potentially for other purposes as well.

By the end of this exercise, you should be able to

- identify if and why a dataset is dirty,
- remove duplicate data,
- use the COUNTIF and TRIM functions to clean data,
- use the Convert Text to Columns Wizard and CONCATENATE functions to clean data,
- use LEFT, RIGHT, MID, and LEN to retrieve parts of a cell within a dataset.

Before proceeding, make sure you have downloaded and saved the "San Francisco Boba Tea Shops Information.csv" file on your computer.

Activity 1: Identify dirty elements in the data.

In this activity, you will not be using any functions. You will explore and understand the data and then identify the weaknesses in the data that require cleaning.

First of all, launch the file "San Francisco Boba Tea Shops Information.csv." You should see data in 6 columns: id (column A), name (column B), rating (column C), and location (Columns D-F).

Very quickly, you can identify one major problem with this data in Column B. The names of the Boba shops are confusing because they contain hyphens. Furthermore, the names of these shops should be capitalized.

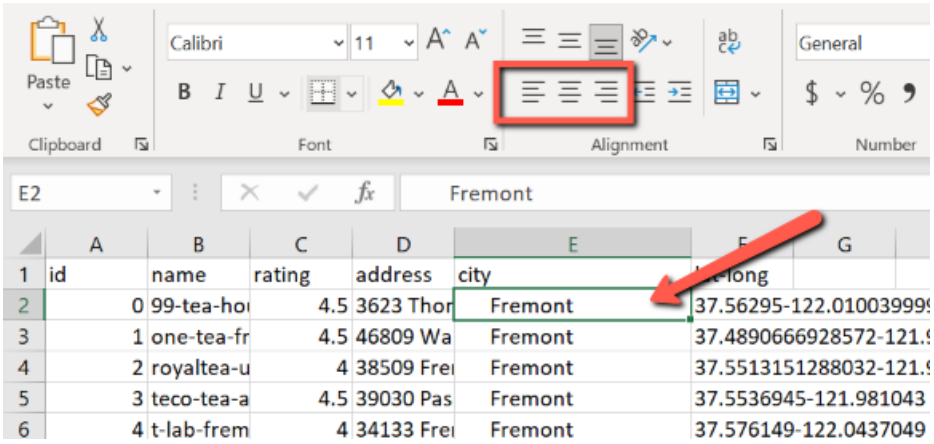
Secondly, the rating should fall between 0 and 5. However, at least one rating falls outside this range, which is in row 8.

	A	B	C	D	E	F	G	H	I	J
1	id	name	rating	address	city	lat-long				
2	0	99-tea-hou	4.5	3623 Thor	Fremo	37.56295-122.010039999999				
3	1	one-tea-fr	4.5	46809 Wa	Fremo	37.4890666928572-121.929413750767				
4	2	royaltea-u	4	38509 Fre	Fremo	37.5513151288032-121.993849799037				
5	3	teco-tea-a	4.5	39030 Pas	Fremo	37.5536945-121.981043				
6	4	t-lab-frem	4	34133 Fre	Fremo	37.576149-122.0437049				
7	5	q-tea-mor	4	39181 Ced	Newar	37.5229604101756-122.005785632481				
8	6	gong-cha-	6.7	46827 Wa	Fremo	37.4885682635695-121.929191268869				
9	6	gong-cha-	4	46827 Wa	Fremo	37.4885682635695-121.929191268869				
10	7	happy-lem	4.5	46873 Wa	Fremo	37.4884429093476-121.930383669657				
11	8	factory-te	3.5	46461 Mis	Fremo	37.4922976027806-121.927918713539				
12	9	super-cue	3.5	43743 Bos	Fremo	37.5007782876492-121.973167955875				
13	10	milk-and-h	3.5	34265 Fre	Fremo	37.575448103287-122.042586920966				

Third, there is at least one duplicate line, as seen in rows 8 and 9:

7	5	q-tea-mor	4	39181 Ced	Newar	37.5229604101756-122.005785632481
8	6	gong-cha-	6.7	46827 Wa	Fremo	37.4885682635695-121.929191268869
9	6	gong-cha-	4	46827 Wa	Fremo	37.4885682635695-121.929191268869
10	7	happy-lem	4.5	46873 Wa	Fremo	37.4884429093476-121.930383669657
11	8	factory-te	3.5	46461 Mis	Fremo	37.4922976027806-121.927918713539
12	9	super-cue	3.5	43743 Bos	Fremo	37.5007782876492-121.973167955875
13	10	milk-and-h	3.5	34265 Fre	Fremo	37.575448103287-122.042586920966
14	11	tea-island	4	46196 Wa	Fremo	37.4935414360442-121.929889351584

Moreover, there seems to be a spacing issue with the city column. Notice that the column is not formatted as **Center align**. Therefore, about five extra spaces have been added to the left of the city name. The data is not technically wrong in this format, but it is not as visually appealing as it can be.



	A	B	C	D	E	F	G
1	id	name	rating	address	city	lat-long	
2	0	99-tea-hou	4.5	3623 Thor	Fremont	37.56295-122.010039999	
3	1	one-tea-fr	4.5	46809 Wa	Fremont	37.4890666928572-121.9	
4	2	royaltea-u	4	38509 Fre	Fremont	37.5513151288032-121.9	
5	3	teco-tea-a	4.5	39030 Pas	Fremont	37.5536945-121.981043	
6	4	t-lab-frem	4	34133 Fre	Fremont	37.576149-122.0437049	

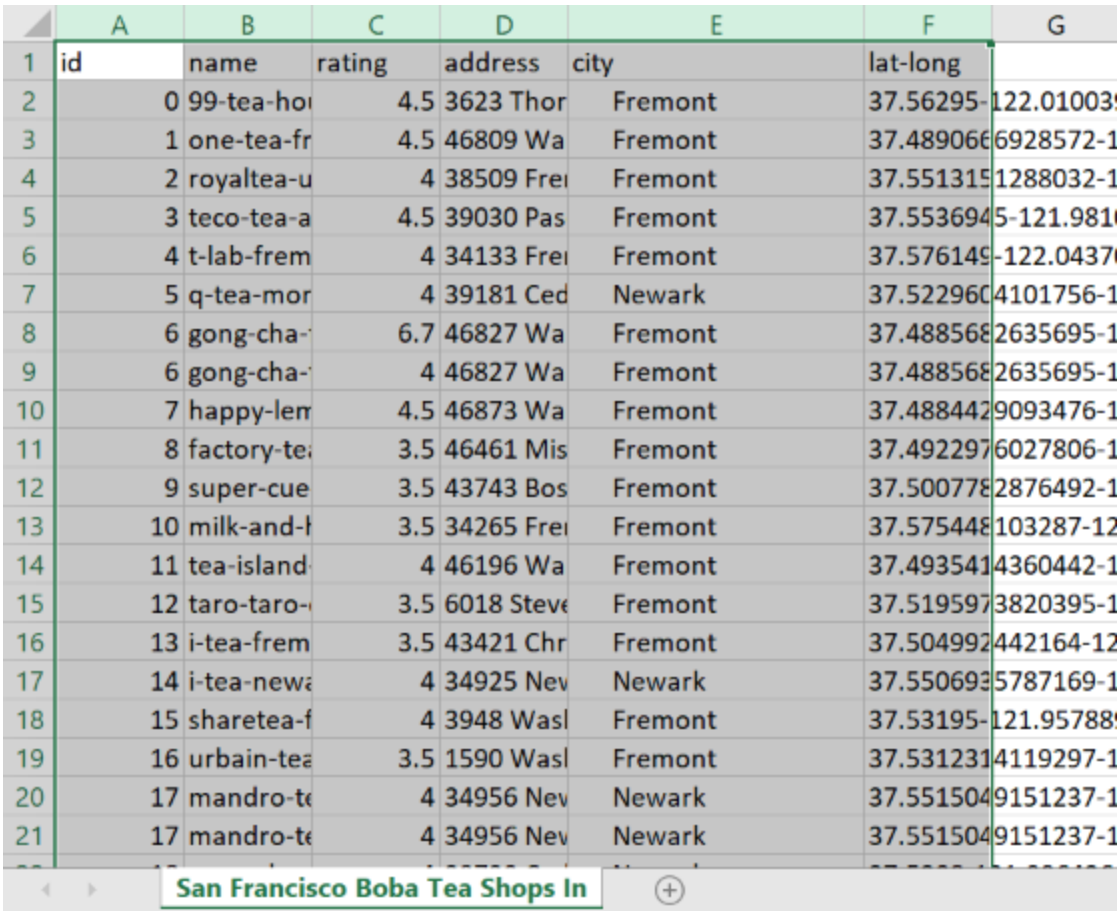
Finally, for someone to use this data, the latitude and longitude should be in separate columns.

city	lat-long
Fremont	37.56295-122.010039999999
Fremont	37.4890666928572-121.929413750767
Fremont	37.5513151288032-121.993849799037
Fremont	37.5536945-121.981043
Fremont	37.576149-122.0437049
Newark	37.5229604101756-122.005785632481

The goal of the following activities is to fix these errors to create a clean dataset.

Activity 2: Removing Duplicates

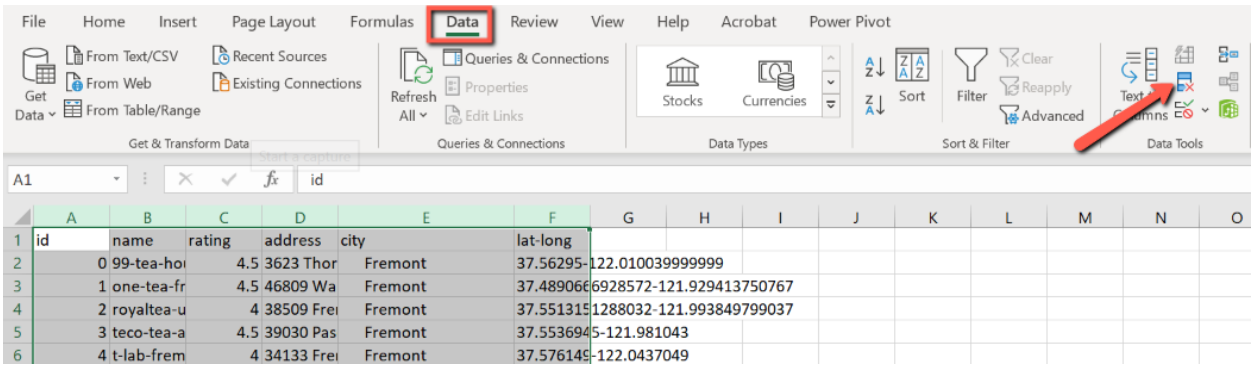
1) Highlight Columns A through F



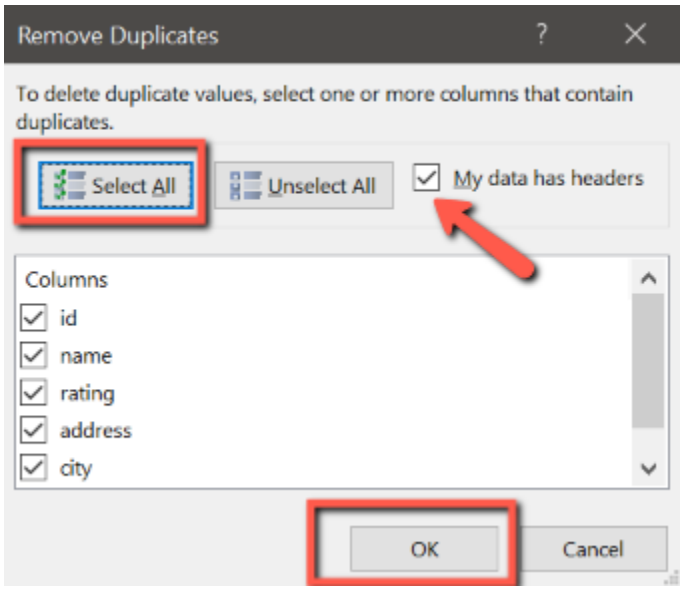
	A	B	C	D	E	F	G
1	id	name	rating	address	city	lat-long	
2	0	99-tea-hou	4.5	3623 Thor	Fremont	37.56295-	122.01003
3	1	one-tea-fr	4.5	46809 Wa	Fremont	37.489066	6928572-1
4	2	royaltea-u	4	38509 Frei	Fremont	37.551315	1288032-1
5	3	teco-tea-a	4.5	39030 Pas	Fremont	37.553694	5-121.981
6	4	t-lab-frem	4	34133 Frei	Fremont	37.576149	-122.0437
7	5	q-tea-mor	4	39181 Ced	Newark	37.522960	4101756-1
8	6	gong-cha-	6.7	46827 Wa	Fremont	37.488568	2635695-1
9	6	gong-cha-	4	46827 Wa	Fremont	37.488568	2635695-1
10	7	happy-lem	4.5	46873 Wa	Fremont	37.488442	9093476-1
11	8	factory-tea	3.5	46461 Mis	Fremont	37.492297	6027806-1
12	9	super-cue	3.5	43743 Bos	Fremont	37.500778	2876492-1
13	10	milk-and-h	3.5	34265 Frei	Fremont	37.575448	103287-12
14	11	tea-island	4	46196 Wa	Fremont	37.493541	4360442-1
15	12	taro-taro-	3.5	6018 Steve	Fremont	37.519597	3820395-1
16	13	i-tea-frem	3.5	43421 Chr	Fremont	37.504992	442164-12
17	14	i-tea-newa	4	34925 Nev	Newark	37.550693	5787169-1
18	15	sharetea-f	4	3948 Wasl	Fremont	37.53195-	121.95788
19	16	urbain-tea	3.5	1590 Wasl	Fremont	37.531231	4119297-1
20	17	mandro-te	4	34956 Nev	Newark	37.551504	9151237-1
21	17	mandro-te	4	34956 Nev	Newark	37.551504	9151237-1

San Francisco Boba Tea Shops In

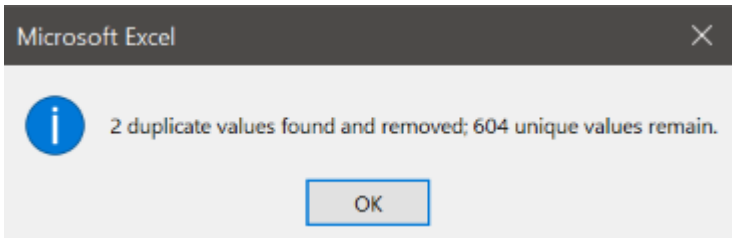
2) Next, go to Data tab -> select Remove Duplicates



3) Then, make sure My data has headers and all columns have been checked (selected). Click OK.



4) If done correctly, the following prompt will be shown. Click OK to complete this activity.



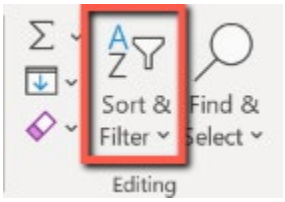
Activity 3: Use the COUNTIF and TRIM functions to clean data

Next, we need to identify out-of-range entries in the rating column and replace them with a reasonable value within the specified range.

1) In cell I2, type in the formula =COUNTIF(C:C,">5"). The first entry, C:C, refers to the range where you are counting the data. In this case, you are referring to Column C, which contains the ratings. Instead

of typing C:C, you can also select the entire column. The second entry, >5, tells the function to count numbers greater than 5. Press ENTER or RETURN. You will see that the function has returned a value of 9. Therefore, you have nine rows that have a value greater than 5.

2) For this activity, replace all the ratings with a value greater than 5 with the value 5. Select cell C1 -> Sort & Filter -> select Sort Z to A.



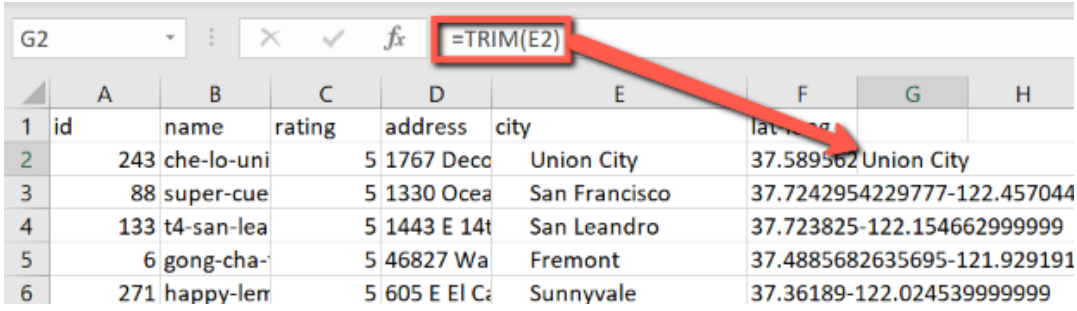
3) Replace all the following values with 5.

	A	B	C	D	E	F	G	H	I	J
1	id	name	rating	address	city	lat-long				
2	243	che-lo-uni	9.2	1767 Decc	Union City	37.5895628278523-122.022492			9	
3	88	super-cue	8.9	1330 Ocea	San Francisco	37.7242954229777-122.457044541931				
4	133	t4-san-lea	7.4	1443 E 14t	San Leandro	37.723825-122.154662999999				
5	6	gong-cha-	6.7	46827 Wa	Fremont	37.4885682635695-121.929191268869				
6	271	happy-lem	6.2	605 E El Ca	Sunnyvale	37.36189-122.024539999999				
7	218	ohana-hav	5.7	5410 Sunc	Pleasanton	37.6522999999999-121.8786				
8	65	infinitea-s	5.6	5351 Gear	San Francisco	37.780295679705-122.477084781597				
9	160	amor-cafe	5.4	110 E San	San Jose	37.3354549999999-121.886596				
10	23	boba-que	5.2	34420 Frei	Fremont	37.5757-122.039769999999				
11	89	puppy-bol	5	1142 Gran	San Francisco	37.7975399525428-122.406789958477				
12	128	qteabar-o	5	478 Lake F	Oakland	37.8110686341717-122.24723573774				
13	147	bobateani	5	75 E Santa	San Jose	37.33709-121.88941				

4) After replacing the erroneous data (rating greater than 5) with the value 5, you will see that the value in cell I2 is 0. This confirms that we no longer have values above 5. Since we don't need this information anymore, you can now delete the formula from cell I2.

5) Next, clean up the city names. There is some extra spacing in front of each city name in Column E. Use the TRIM to remove the additional space in those entries.

In cell G2, type =TRIM(E2), and then press Enter.



6) Double-click on the green square in the bottom right corner of cell G2 to copy the formula down the entire column. Next, you will replace Column E with these entries.

   Select the G column, right-click, and select Copy.

   Then select Column E, right-click, under Paste Options -> select Values (V).

   Rename Column E as city, then select column G, right-click, and select Delete column. Your first ten rows should look like this:

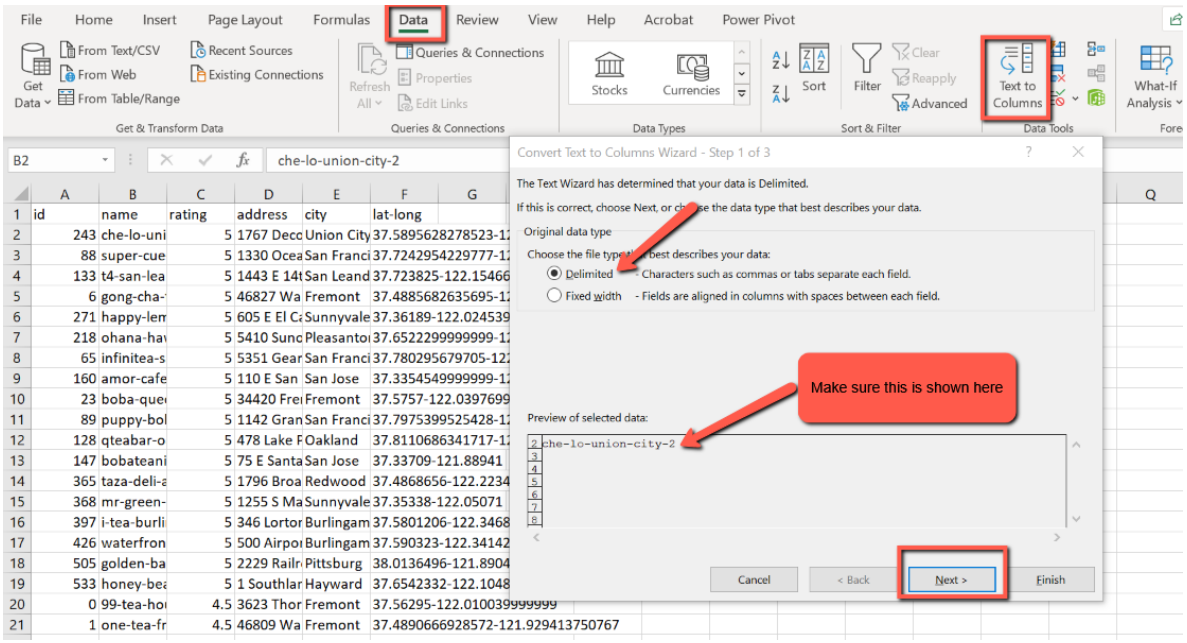
	A	B	C	D	E	F	G	H	I
1	id	name	rating	address	city	lat-long			
2	243	che-lo-uni	5	1767 Decc	Union City	37.5895628278523-122.022492714298			
3	88	super-cue	5	1330 Ocea	San Francisco	37.7242954229777-122.457044541931			
4	133	t4-san-lea	5	1443 E 14t	San Leandro	37.723825-122.154662999999			
5	6	gong-cha-	5	46827 Wa	Fremont	37.4885682635695-121.929191268869			
6	271	happy-lem	5	605 E El Ca	Sunnyvale	37.36189-122.024539999999			
7	218	ohana-hav	5	5410 Sunc	Pleasanton	37.6522299999999-121.8786			
8	65	infinitea-s	5	5351 Gear	San Francisco	37.780295679705-122.477084781597			
9	160	amor-cafe	5	110 E San	San Jose	37.3354549999999-121.886596			

Activity 4: Use the Convert Text to Columns Wizard and CONCATENATE functions to clean data.

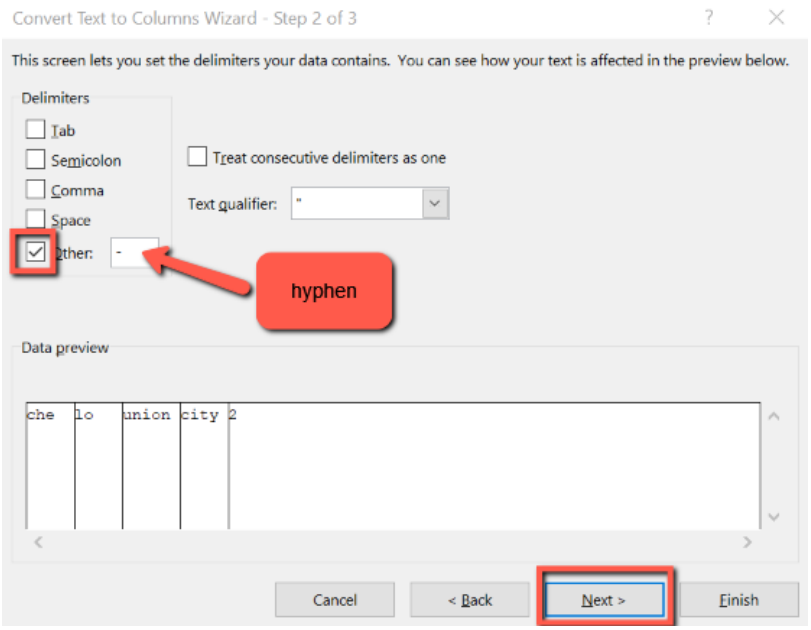
Next, clean up the names of the boba shops by removing the hyphens and capitalizing all of the words.

1) Select column B. Make sure the entire column is selected.

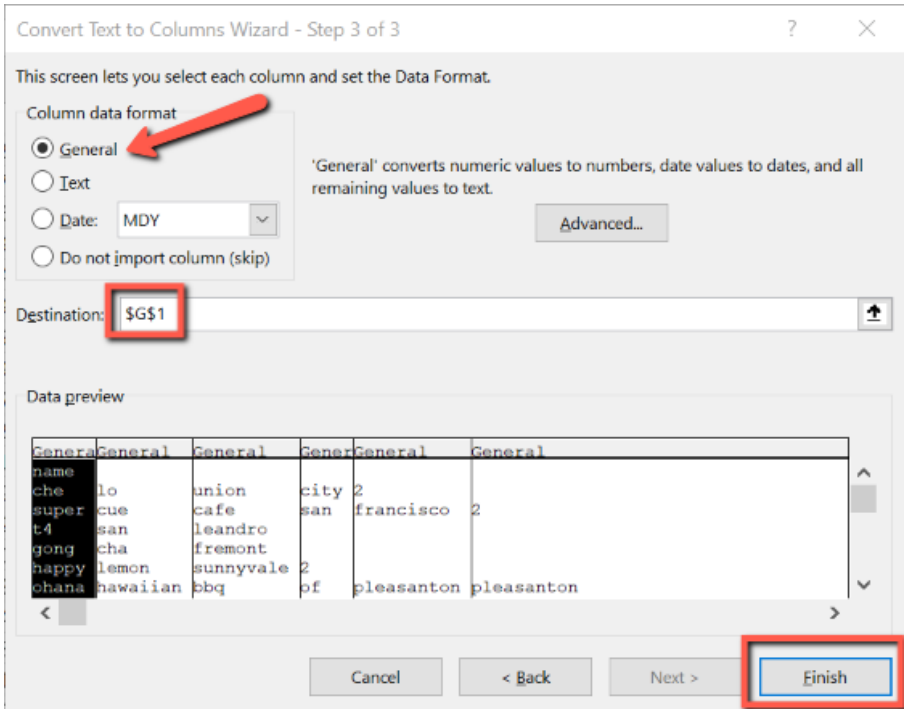
2) Go to Data tab -> select Text to Columns -> select Delimited -> click Next >



3) Check Other: in the Delimiters and insert a – into the box next to it. Notice that in the Data preview, the name of the Boba shop has been split into several words. Finally, click Next >.



4) Select General in the Column data format. The Destination should be G1. Finally, click Finish.



5) After clicking on the Finish button, the result shows each fragment of the cell surrounding a hyphen in a different cell:

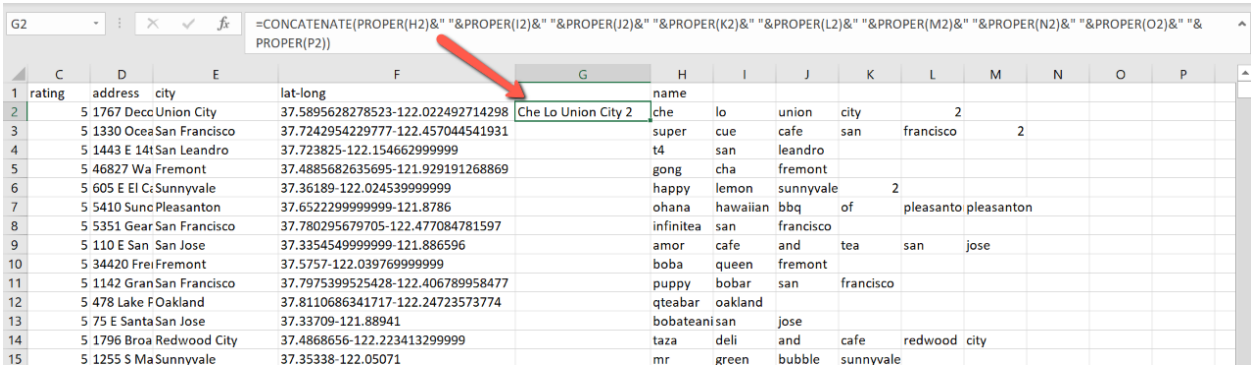
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	name	rating	address	city	lat-long	name						
2	243	che-lo-uni	5	1767 Decc	Union City	37.58956	che	lo	union	city	2		
3	88	super-cue	5	1330 Ocea	San Franci	37.72429	super	cue	cafe	san	francisco	2	
4	133	t4-san-lea	5	1443 E 14t	San Leand	37.72382	t4	san	leandro				
5	6	gong-cha-	5	46827 Wa	Fremont	37.48856	gong	cha	fremont				
6	271	happy-lem	5	605 E El Ca	Sunnyvale	37.36189	happy	lemon	sunnyvale	2			
7	218	ohana-hav	5	5410 Sunc	Pleasanto	37.65222	ohana	hawaiian	bbq	of	pleasanto	pleasanton	
8	65	infinitea-s	5	5351 Gear	San Franci	37.78029	infinitea	san	francisco				
9	160	amor-cafe	5	110 E San	San Jose	37.33545	amor	cafe	and	tea	san	jose	
10	23	boba-que	5	34420 Frei	Fremont	37.5757-1	boba	queen	fremont				
11	89	puppy-bol	5	1142 Gran	San Franci	37.79753	puppy	bobar	san	francisco			
12	128	qteabar-o	5	478 Lake F	Oakland	37.81106	qteabar	oakland					
13	147	bobateani	5	75 E Santa	San Jose	37.33709	bobateani	san	jose				
14	365	taza-deli-a	5	1796 Broa	Redwood	37.48686	taza	deli	and	cafe	redwood	city	
15	368	mr-green-	5	1255 S Ma	Sunnyvale	37.35338	mr	green	bubble	sunnyvale			

6) Now, insert a new column to the left of the G column. You may do this by right-clicking on the G column -> Insert. The new column added is now named G.

7) Next, capitalize each part of the name, and join the cells back together. Use the PROPER function to capitalize words and the CONCATENATE function to rejoin the names.

8) The longest entries from after splitting the name column are 9 cells long and have entries from Columns H to O. Type the function in cell G2: `=CONCATENATE(PROPER(H2)&" "&PROPER(I2)&"`

"&PROPER(J2)&" "&PROPER(K2)&" "&PROPER(L2)&" "&PROPER(M2)&" "&PROPER(N2)&" "&PROPER(O2)&" "&PROPER(P2)), and press Enter.



	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	rating	address	city	lat-long		name								
2	5	1767 Decc	Union City	37.5895628278523-122.022492714298	Che Lo Union City 2	che	lo	union	city	2				
3	5	1330 Ocea	San Francisco	37.7242954229777-122.457044541931		super	cue	cafe	san	francisco	2			
4	5	1443 E 14t	San Leandro	37.723825-122.154662999999		t4	san	leandro						
5	5	46827 Wa	Fremont	37.4885682635695-121.929191268869		gong	cha	fremont						
6	5	605 E El Ca	Sunnyvale	37.36189-122.024539999999		happy	lemon	sunnyvale	2					
7	5	5410 Sunc	Pleasanton	37.6522299999999-121.8786		ohana	hawaiian	bbq	of	pleasanto	pleasanton			
8	5	5351 Gear	San Francisco	37.780295679705-122.477084781597		infinitea	san	francisco						
9	5	110 E San	San Jose	37.3354549999999-121.886596		amor	cafe	and	tea	san	jose			
10	5	34420 Frei	Fremont	37.5757-122.039769999999		boba	queen	fremont						
11	5	1142 Gran	San Francisco	37.7975399525428-122.406789958477		puppy	bobar	san	francisco					
12	5	478 Lake F	Oakland	37.8110686341717-122.24723573774		qteabar	oakland							
13	5	75 E Santa	San Jose	37.33709-121.88941		bobateani	san	jose						
14	5	1796 Broa	Redwood City	37.4868656-122.223413299999		taza	deli	and	cafe	redwood	city			
15	5	1255 S Ma	Sunnyvale	37.35338-122.05071		mr	green	bubble	sunnyvale					

9) Apply the formula in step 8 to all other cells from G3 onwards. You may do this by double-clicking on the bottom right corner of cell G2. Note: You will see a + sign when you move your mouse cursor to the bottom right corner of G2. Name the cell G1 as name.

	C	D	E	F	G	H	I	J	K
1	rating	address	city	lat-long	name	name			
2	5	1767 Decc	Union City	37.5895628278523-122.022492714298	Che Lo Union City 2	che	lo	union	city
3	5	1330 Ocea	San Francisco	37.7242954229777-122.457044541931	Super Cue Cafe San Francisco 2	super	cue	cafe	san
4	5	1443 E 14t	San Leandro	37.723825-122.154662999999	T4 San Leandro	t4	san	leandro	
5	5	46827 Wa	Fremont	37.4885682635695-121.929191268869	Gong Cha Fremont	gong	cha	fremont	
6	5	605 E El Ca	Sunnyvale	37.36189-122.024539999999	Happy Lemon Sunnyvale 2	happy	lemon	sunnyvale	2
7	5	5410 Sunc	Pleasanton	37.6522299999999-121.8786	Ohana Hawaiian Bbq Of Pleasanton Pleasanton	ohana	hawaiian	bbq	of
8	5	5351 Gear	San Francisco	37.780295679705-122.477084781597	Infinitea San Francisco	infinitea	san	francisco	
9	5	110 E San	San Jose	37.3354549999999-121.886596	Amor Cafe And Tea San Jose	amor	cafe	and	tea
10	5	34420 Frei	Fremont	37.5757-122.039769999999	Boba Queen Fremont	boba	queen	fremont	
11	5	1142 Gran	San Francisco	37.7975399525428-122.406789958477	Puppy Bobar San Francisco	puppy	bobar	san	francisco
12	5	478 Lake F	Oakland	37.8110686341717-122.24723573774	Qteabar Oakland	qteabar	oakland		

10) Right-click on column G -> Copy. Then, right-click on column B -> under Paste Options: select Values (V).

11) Now that column B, called name, has all the correct Boba shop names, it's time to remove all unwanted columns. Specifically, delete columns G to P.

	A	B	C	D	E	F	G	H	I
1	id	name	rating	address	city	lat-long			
2	243	Che Lo Union City 2	5	1767 Decc	Union City	37.5895628278523-122.022492714298			
3	88	Super Cue Cafe San Francisco 2	5	1330 Ocea	San Francisco	37.7242954229777-122.457044541931			
4	133	T4 San Leandro	5	1443 E 14t	San Leandro	37.723825-122.154662999999			
5	6	Gong Cha Fremont	5	46827 Wa	Fremont	37.4885682635695-121.929191268869			
6	271	Happy Lemon Sunnyvale 2	5	605 E El Ca	Sunnyvale	37.36189-122.024539999999			
7	218	Ohana Hawaiian Bbq Of Pleasanton Pleasanton	5	5410 Sunc	Pleasanton	37.6522299999999-121.8786			
8	65	Infinitea San Francisco	5	5351 Gear	San Francisco	37.780295679705-122.477084781597			
9	160	Amor Cafe And Tea San Jose	5	110 E San	San Jose	37.3354549999999-121.886596			
10	23	Boba Queen Fremont	5	34420 Frei	Fremont	37.5757-122.039769999999			
11	89	Puppy Bobar San Francisco	5	1142 Gran	San Francisco	37.7975399525428-122.406789958477			
12	128	Qteabar Oakland	5	478 Lake F	Oakland	37.8110686341717-122.24723573774			
13	147	Bobateani San Jose	5	75 E Santa	San Jose	37.33709-121.88941			
14	365	Taza Deli And Cafe Redwood City	5	1796 Broa	Redwood City	37.4868656-122.223413299999			
15	368	Mr Green Bubble Sunnyvale	5	1255 S Ma	Sunnyvale	37.35338-122.05071			
16	397	I Tea Burlingame 2	5	346 Lortor	Burlingame	37.5801206-122.346889099999			

Activity 5: Use LEFT, RIGHT, MID, and LEN to retrieve parts of a cell within a dataset.

To complete the cleanup, split the latitude and longitude (Column F) into two cells, and make them shorter. To do this, use the LEFT, RIGHT, and LEN functions.

1) To enter only the latitude into a cell, use the LEFT function to insert all values to the left of the hyphen in cell G2. In cell G2, type `=LEFT(F2, FIND("-", F2) - 1)`. The FIND function is used to determine the position of a hyphen in the string, while the LEFT function returns all the values to its left. Inserting the "-1" ensures that the hyphen is not returned because the second input in the function is equal to the Number of characters desired to be extracted.

2) After pressing ENTER or RETURN, the following should be in cell G2:

	A	B	C	D	E	F	G	H
1	id	name	rating	address	city	lat-long		
2	243	Che Lo Un	5	1767 Decc	Union City	37.5895628278523-122.022492714298	37.5895628278523	
3	88	Super Cue	5	1330 Ocea	San Franci	37.7242954229777-122.457044541931		
4	133	T4 San Lea	5	1443 E 14t	San Leand	37.723825-122.154662999999		
5	6	Gong Cha	5	46827 Wa	Fremont	37.4885682635695-121.929191268869		
6	271	Happy Ler	5	605 E El Ca	Sunnyvale	37.36189-122.024539999999		
7	218	Ohana Ha	5	5410 Sunc	Pleasantor	37.6522299999999-121.8786		
8	65	Infinitea S	5	5351 Gear	San Franci	37.780295679705-122.477084781597		
9	160	Amor Cafe	5	110 E San	San Jose	37.3354549999999-121.886596		

3) Left-click on the bottom right corner of cell G2 to copy the formula down and populate the entire row in the same manner.

4) Give this new column a name. In cell G1, type in latitude.

5) Next, complete a similar process with longitude, but use the RIGHT function. Using the RIGHT function in cell H2 will return all values to the right of the hyphen. In cell H2, type `=RIGHT(F2, LEN(F2) - FIND("-", F2))`. The LEN function measures the length of the cell, and the FIND function subtracts the length of the values to the left of the cell. This makes the Number of characters removed equal to those to the right of the hyphen. This entire function returns all values to the right of the hyphen.

6) After pressing ENTER or RETURN, the following should be in cell H2:

F	G	H
lat-long	latitude	
37.5895628278523-122.022492714298	37.5895628278523	122.022492714298
37.7242954229777-122.457044541931	37.7242954229777	
37.723825-122.154662999999	37.723825	
37.4885682635695-121.929191268869	37.4885682635695	
37.36189-122.024539999999	37.36189	
37.6522299999999-121.8786	37.6522299999999	
37.780295679705-122.477084781597	37.780295679705	
37.3354549999999-121.886596	37.3354549999999	

7) Now, complete the same steps for longitude. Left-click the bottom right corner of the cell H2 to copy the formula down the entire column.

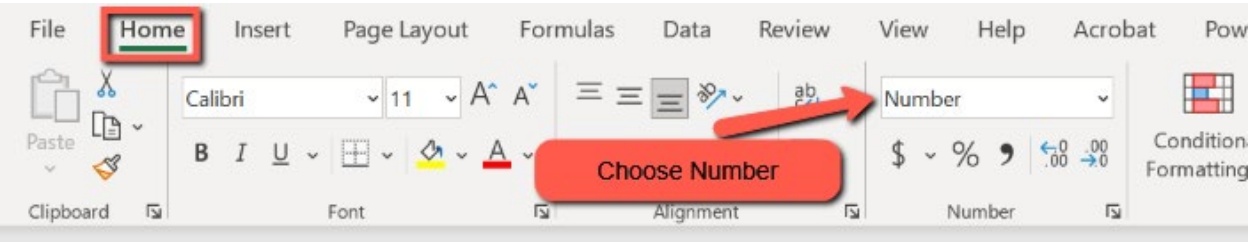
8) Give this new column a name. In cell H1, type in longitude.

9) Select both columns G and H -> right-click and select Copy -> paste these two columns in column I as Values (V).

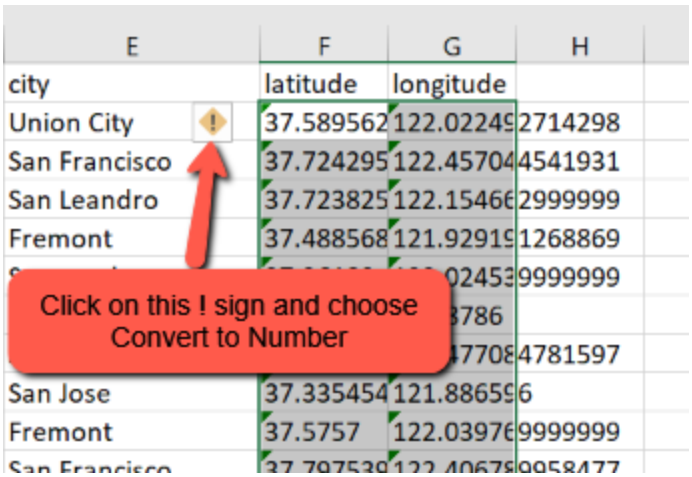
10) Now, columns F, G, and H are no longer needed. Delete these columns. You should get the following result:

	A	B	C	D	E	F	G
1	id	name	rating	address	city	latitude	longitude
2	243	Che Lo Un	5	1767 Decoto Blvd	Union City	37.5895628278523	122.022492714298
3	88	Super Cue	5	1330 Ocean Ave	San Francisco	37.7242954229777	122.457044541931
4	133	T4 San Lea	5	1443 E 14th St	San Leandro	37.723825	122.154662999999
5	6	Gong Cha	5	46827 Warm Springs Blvd	Fremont	37.4885682635695	121.929191268869
6	271	Happy Ler	5	605 E El Camino Real	Sunnyvale	37.36189	122.024539999999
7	218	Ohana Ha	5	5410 Sunol Blvd	Pleasanton	37.6522299999999	121.8786
8	65	Infinitea S	5	5351 Geary Blvd	San Francisco	37.780295679705	122.477084781597
9	160	Amor Cafe	5	110 E San Fernando St	San Jose	37.3354549999999	121.886596

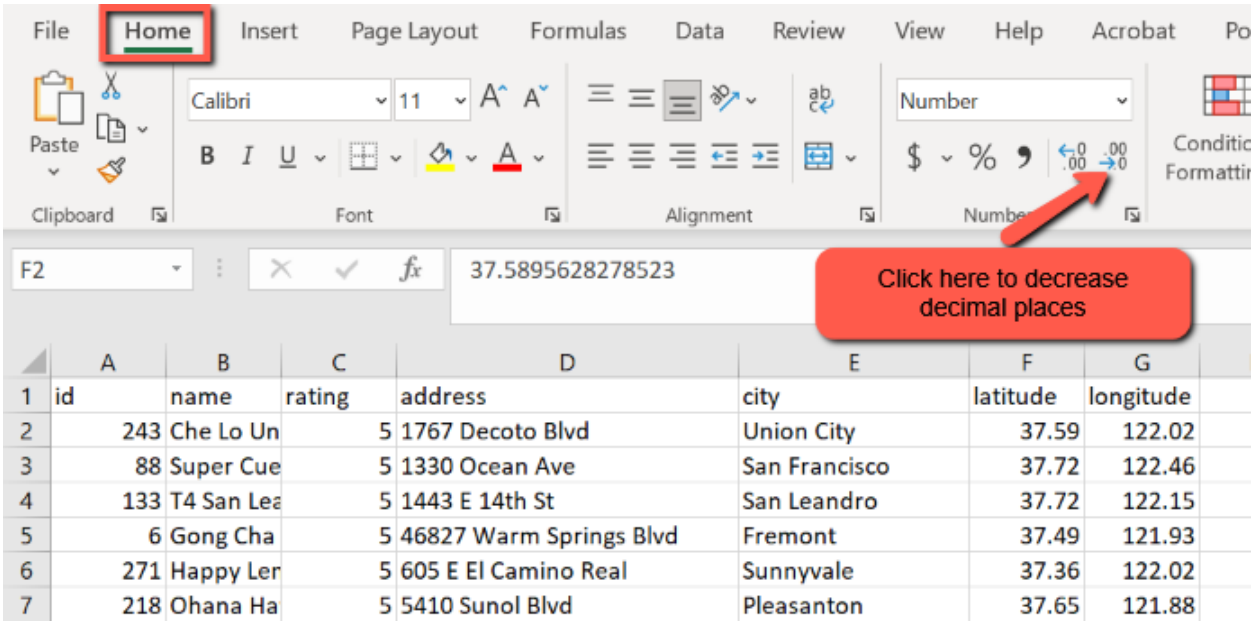
11) Notice that a green arrow is at the top left of the cells in columns F and G. These values (latitudes and longitudes) are currently formatted as General. We need to convert them into Number. Select both columns F and G -> select Home tab -> in the Number section drop-down list, choose Number.



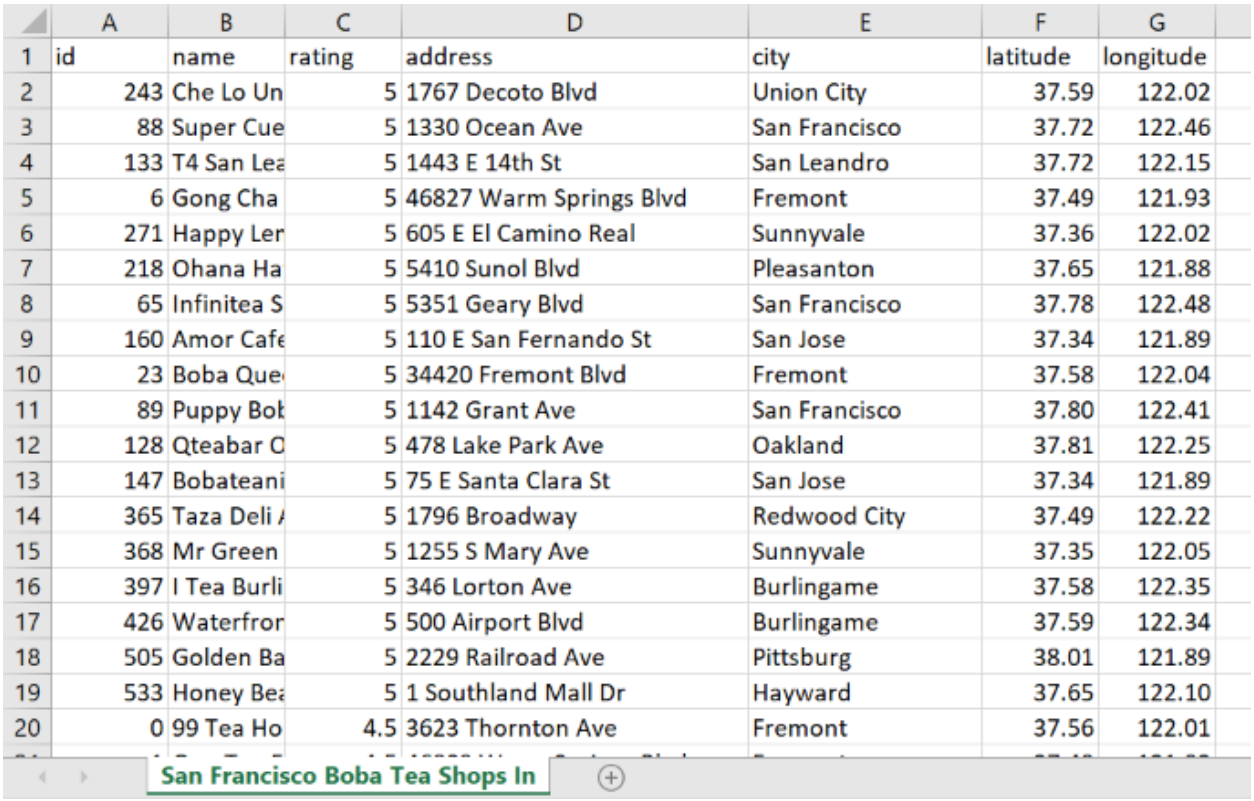
12) Format columns F and G to take in the latitude and longitude values in 2 decimal places. Select all the cells from F2:G605. Next to cell F2, there is an exclamation sign (!). Click ! sign and choose Convert to Number.



13) Finally, decrease the number of decimal places for these two columns. Go to the Home tab -> click on Decrease Decimal until you get two decimal places.



Here is your cleaned dataset:



	A	B	C	D	E	F	G
1	id	name	rating	address	city	latitude	longitude
2	243	Che Lo Un	5	1767 Decoto Blvd	Union City	37.59	122.02
3	88	Super Cue	5	1330 Ocean Ave	San Francisco	37.72	122.46
4	133	T4 San Lea	5	1443 E 14th St	San Leandro	37.72	122.15
5	6	Gong Cha	5	46827 Warm Springs Blvd	Fremont	37.49	121.93
6	271	Happy Ler	5	605 E El Camino Real	Sunnyvale	37.36	122.02
7	218	Ohana Ha	5	5410 Sunol Blvd	Pleasanton	37.65	121.88
8	65	Infinitea S	5	5351 Geary Blvd	San Francisco	37.78	122.48
9	160	Amor Cafe	5	110 E San Fernando St	San Jose	37.34	121.89
10	23	Boba Que	5	34420 Fremont Blvd	Fremont	37.58	122.04
11	89	Puppy Bob	5	1142 Grant Ave	San Francisco	37.80	122.41
12	128	Qteabar C	5	478 Lake Park Ave	Oakland	37.81	122.25
13	147	Bobateani	5	75 E Santa Clara St	San Jose	37.34	121.89
14	365	Taza Deli A	5	1796 Broadway	Redwood City	37.49	122.22
15	368	Mr Green	5	1255 S Mary Ave	Sunnyvale	37.35	122.05
16	397	I Tea Burli	5	346 Lorton Ave	Burlingame	37.58	122.35
17	426	Waterfror	5	500 Airport Blvd	Burlingame	37.59	122.34
18	505	Golden Ba	5	2229 Railroad Ave	Pittsburg	38.01	121.89
19	533	Honey Bea	5	1 Southland Mall Dr	Hayward	37.65	122.10
20	0	99 Tea Ho	4.5	3623 Thornton Ave	Fremont	37.56	122.01

San Francisco Boba Tea Shops In

Congratulations!

Well done! Your data is now in much better shape than before. You may save your file and close it now.

Reference

The above hands-on activity sheet was adopted (with modifications) from the online course "Google Data Analytics Certificate."