



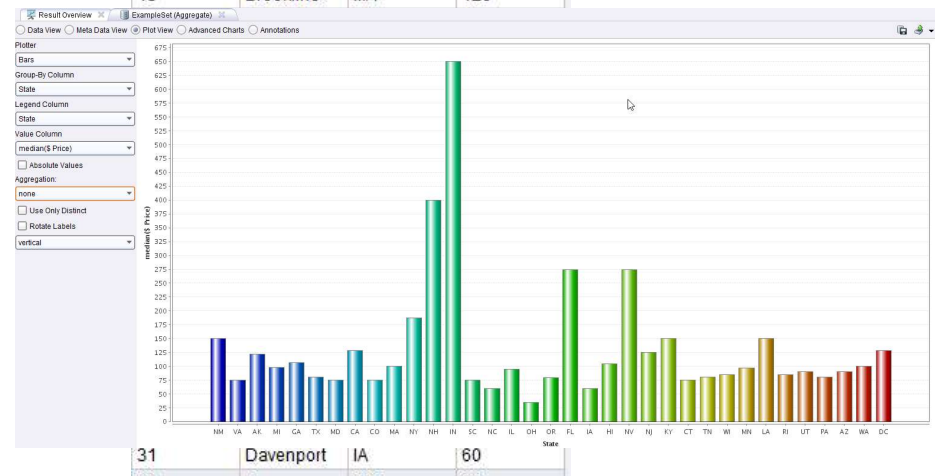
Mark D. Adams

DATA STUDIO ANALYST

Exercise

- ▶ I do a lot of my analysis in RapidMiner
- ▶ Took 30 seconds
- ▶ Can plot on a graph if needed
- ▶ Can even group by state in graph

Row No.	City	State	median(\$ P...
1	Albuquerque	NM	75
2	Alexandria	VA	100
3	Anchorage	AK	122
4	Ann Arbor	MI	98
5	Arlington	VA	108
6	Atlanta	GA	89
7	Austin	TX	199
8	Baltimore	MD	80
9	Berkeley	CA	100
10	Berkeley	CO	65
11	Beverly Hills	CA	160
12	Boston	MA	160
13	Boulder	CO	100
14	Bronx	NY	75
15	Brookline	MA	120



Power BI

- ▶ I'm starting to use Power BI more for exploring data sets, and it is able to handle this data set
- ▶ Again, took about 30 seconds
- ▶ It also has Maps

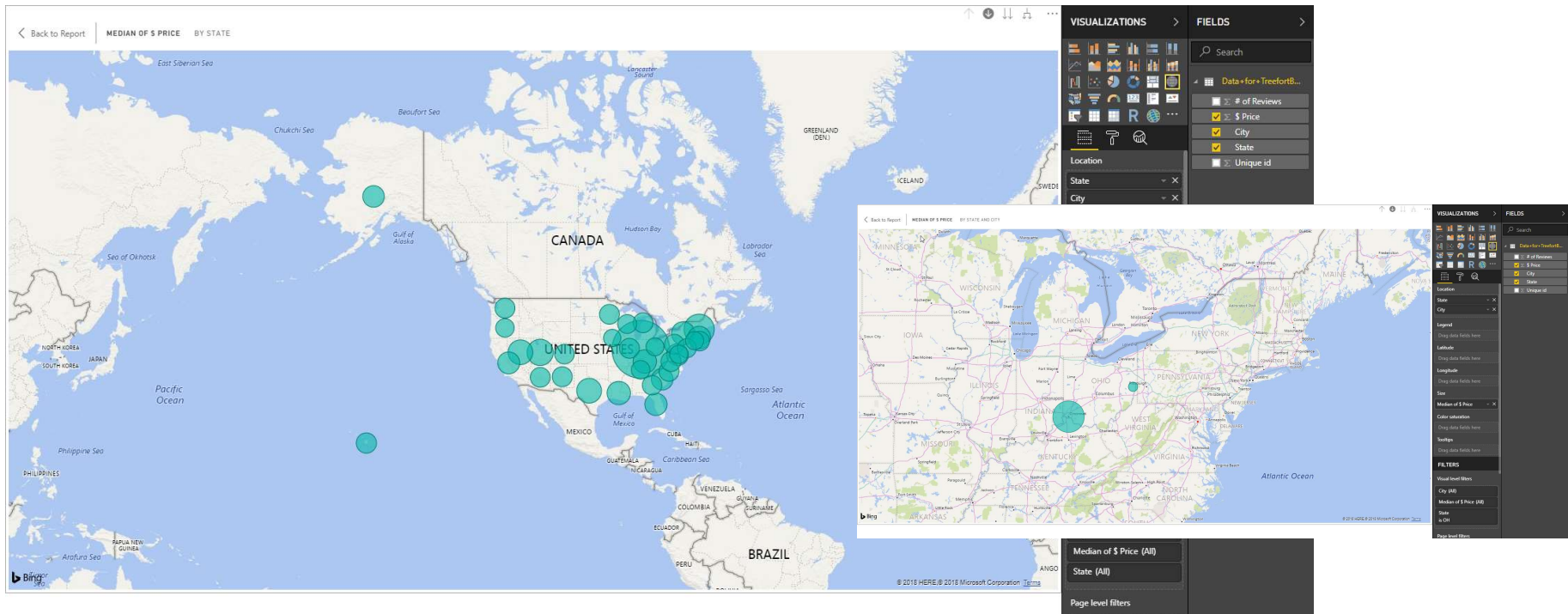
Data+for+TreefortBnB+Puzzle.csv

File Origin: 1252: Western European (Windows) | Delimiter: Comma | Data Type Detection: Based on entire dataset

Unique id	City	State	\$ Price	# of Reviews
1	Portland	OR	75	.5
2	San Diego	CA	95	.3
3	New York	NY	149	37
4	Los Angeles	CA	199	45
5	Denver	CO	56	99
6	Denver	CO	65	83
7	Denver	CO	80	0
8	New York	NY	125	26
9	New York	NY	85	32
10	New York	NY	110	29
11	New York	NY	98	36
12	San Francisco	CA	165	13
13	Minneapolis	MN	250	0
14	Phoenix	AZ	400	0
15	Denver	CO	95	1
16	Denver	CO	95	0
17	Minneapolis	MN	230	0
18	Denver	CO	145	0
19	Sausalito	CA	120	11
20	Miami	FL	99	47
21	New York	NY	150	57
22	Austin	TX	160	8

The screenshot shows the Power BI Desktop interface. The main area displays a table of data with columns for City, State, Price, and Reviews. A context menu is open over the 'Median' aggregation function in the 'Visualizations' pane. The menu options include: Remove field, Rename, Conditional formatting, Remove conditional formatting, Don't summarize, Sum (checked), Average, Minimum, Maximum, Count (Distinct), Count, Standard deviation, Variance, Median (highlighted), Show values, and New quick measure.

Power BI map with drilldown



Problem: Add aggregations to query

Unique id	City	State	\$ Price	
1	Portland	OR		
2	San Diego	CA		
3	New York	NY		
4	Los Angeles	CA		
5	Denver	CO		
6	Denver	CO		
7	Denver	CO		
8	New York	NY		
9	New York	NY		
10	New York	NY		
11	New York	NY		
12	San Francisco	CA		
13	Minneapolis	MN		
14	Phoenix	AZ		
15	Denver	CO		
16	Denver	CO		
17	Minneapolis	MN		
18	Denver	CO		
19	Sausalito	CA		
20	Miami	FL		
21	New York	NY		
22	Austin	TX		
23	New York	NY	110	99
24	Minneapolis	MN	95	49
25	Portland	OR	47	31
26	New York	NY	70	3
27	New York	NY	99	79
28	Denver	CO	70	34
29	New Haven	CT	64	37
30	New York	NY	119	6

Group By

Basic Advanced

Specify the columns to group by and one or more outputs.

Group by

City

State

Add grouping

New column name	Operation	Column
CountDistinct	Count Distinct Rows	
MedianPrice	Median	\$ Price

Add aggregation

Cancel

Create rank and filter

Ranking = RANKX (ALLSELECTED ('Data+for+Treefort8nB+Puzzle'), CALCULATE (SUM ('Data+for+Treefort8nB+Puzzle' [CountDistinct])))

City	State	MedianPrice
Tucson	AZ	90.00
Salt Lake City	UT	89.50
Atlanta	GA	89.00
Dallas	TX	87.00
Denver	CO	86.00

FILTERS

Visual level filters

City (All)

MedianPrice (All)

Ranking ^ x

is less than or equal t... ◇

Show items when the value:

is less than or equal to ▾

100

And Or

▾

▾

Apply filter

VISUALIZATIONS > **FIELDS** >

Search

Data+for+Treefort8nB+Puzzle

- City
- CountDistinct
- MedianPrice
- Ranking
- State

Values

Sort and export

City	State	MedianPrice
Indianapolis	IN	650.00
Malibu	CA	304.00
Park City	UT	299.00
Healdsburg	CA	275.00
Truckee	NV	275.00
Laguna Beach	CA	268.50
Incline Village	NV	259.00
Manhattan Beach	CA	209.00
Charlotte	NC	200.00

City	State	MedianPrice
Indianapolis	IN	650.00
Malibu	CA	304.00
Park City	UT	299.00
Healdsburg	CA	275.00
Truckee	NV	275.00
Laguna Beach	CA	268.50
Incline Village	NV	259.00
Manhattan Beach	CA	209.00
Charlotte	NC	200.00

- Export data
- Show Data
- Remove
- Automatically find clusters
- Spotlight
- Sort Descending
- Sort Ascending
- Sort by

Questions?

- ▶ Mark D. Adams
- ▶ Cell: 707-205-0982
- ▶ mark@bigdatamark.com
- ▶ <https://www.bigdatamark.com/>