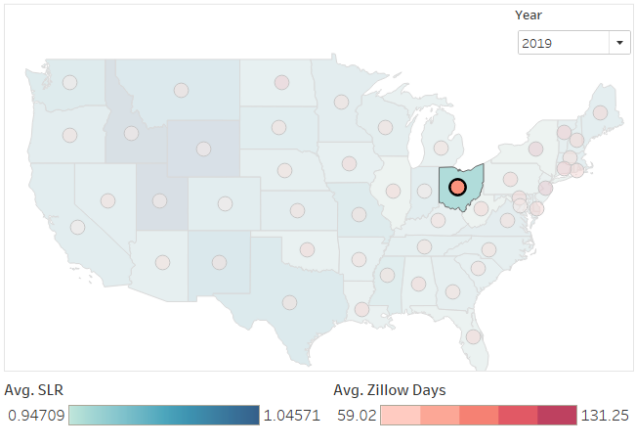
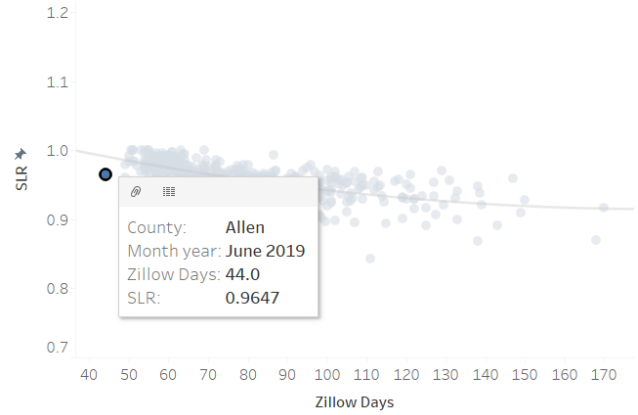
Analyzing key sales metrics for Zillow homes, USA market - Sale price to List price ratio (SLR) and No of Days the home was on Zillow platform before sale (Zillow Days)

Data source - https://www.zillow.com/research/data/

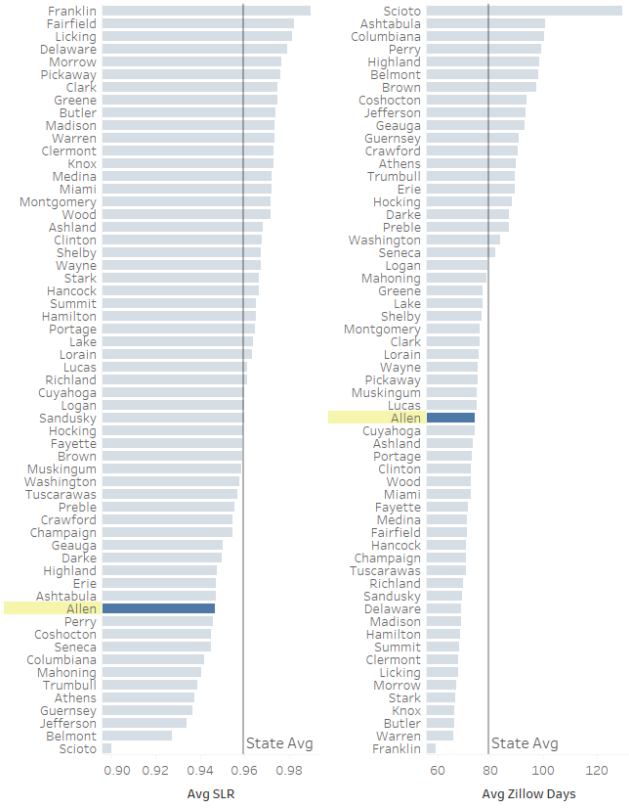
State wise overview



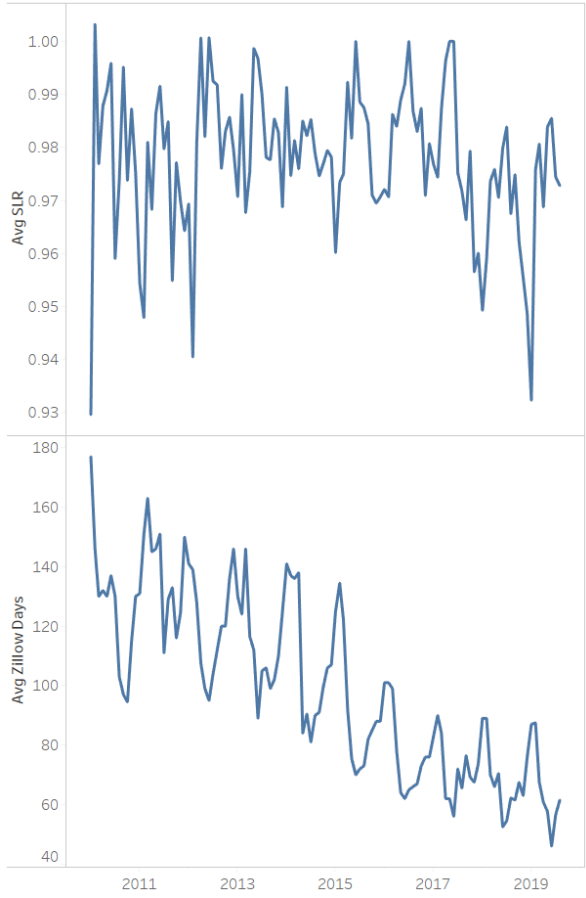
SLR vs Zillow days - Ohio, 2019



Ohio, 2019 - County wise averages



Allen County - trends over time



Business use case: The VP of sales, Zillow Group, Inc., has just returned from Las Vegas after being amazed by the power of Tableau at the Tableau Conference 2019. He now wants his sales team to port his current sales reports from excel to Tableau. As a start, he has instructed his sales team to provide a dashboard to analyze key sales metrics for USA – SLR and Zillow days for the Zillow platform. This dashboard would be used to evaluate what states and counties are seeing higher SLRs and longer time for homes to sell on Zillow and use these insights to identify geographies with potential for improvement.

Data source: https://www.zillow.com/research/data/

https://public.tableau.com/profile/adil.kumar#!/vizhome/ZillowHomesinUS/Dashboard?publish=yes

SLR - Sale price to List price ratio

Zillow Days - No of Days the home was on Zillow platform before sale

Dashboard actions guide

Select any year - select any state on map chart – Look at the relationship between SLR and Zillow days for that state in scatter chart – Hover on any county in scatter chart to see same highlighted county in bar chart for county averages – Click on county in bar chart or click on any data point in scatter chart to filter line chart – look at that county trends over time for both metrics.

Description:

I have used the Sale to List price ratio and Days on Zillow excel sheets at County level from the above data source. The datasets are 2 files, one for each metric – each file having county level data with time period on columns. Initially I decided to use Tableau data blending to import both files as separate data sources and use pivot option to pivot the time period columns into rows since that would be the right format for visualizing time series. However, the performance of the dashboard was a bit slow so I decided to just use R to prep the data in the right format. So I imported both data files into R, pivoted both datasets to make sure time periods moved from columns to rows, and then joined the 2 datasets to have 1 file with both measures. I then imported this file into Tableau.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
SizeRank	RegionID	RegionNar	RegionTyp	StateNam	2010-01	2010-02	2010-03	2010-04	2010-05	2010-06	2010-07	2010-08	2010-09	2010-10	2010-11	2010-12	2011-01	2011-02	2011-03	2011-04	2011-05	2011-06	20
1	3101	Los Angele	County	CA	0.966574	0.958159	0.965846	0.969931	0.973445	0.970285	0.969231	0.964912	0.960452	0.955686	0.952608	0.946009	0.945652	0.93985	0.945946	0.951557	0.954592	0.952987	0.

The first graph on top left is a Map that shows overview of SLR and Zillow days by state in the US. We can filter the dashboard by year to see the specific years data. The states are colored by average SLR.I have used circles to represent average Zillow days. I initially sized the circles by Avg Zillow days, but the circles looked very similar because of the way the data was. So, I kept all circle sizes same and used color by instead on the circles on each state. I felt the map chart would be best to get a state wise overview of both sales metrics. I also intend to use the map chart as a starting point to drill down by state in the following charts. Clicking on any state in the map would filter the scatter plot (bottom right) and the bar charts (right side of map) to show only that particular state data. For both colors I used a gradient to show darker colors for higher values.

I then wanted to use a scatter plot to see if there was any relationship between SLR and Zillow days. I have edited the X axis to not include 0 since there are no data points less than 40. I have also edited the Y axis to show values only above 0.7 since that is the lowest value in the data. A hypothesis could be that, the longer the house is on the Zillow platform, the lesser Sale price it might actually fetch in comparison to the list price. It might also be that this relationship could vary by state. I also added in a trend line to depict the overall relationship between the

measures. Since the relationship does not exactly look linear, I used the polynomial function of order 2, as it seemed to depict the relationship better. We can then see what data points were not exactly following the relationship or looked like outliers. Highlighting this data point would, highlight the same county in the bar charts to see its particular county average in comparison with this data point. Selecting a data point in the scatter plot also filters by that specific county in the line chart in the far right, to see trends across time for that county for both measures.

The bar plots in the center serve the purpose of giving the country wise averages for both metrics for any selected state. I have edited the axes of the bar plots to not start from 0 since the differences among data points is not very evident when scale starts from 0. I have also added an average line to show the average for a particular state, so we can see how the country average was in comparison to state average. Hovering on any county bar on either bar chart would highlight that particular county's data points in the scatter plot and also that particular county in the other bar chart to see the average metric. Selecting and county would filter the line charts on the right to show that particular county's trends over time for both metrics.

The Line charts on the extreme right serve the purpose of looking at trends over time for both measures at the lowest level of granularity in the data – county. There are no further actions from here on. This would be the last view before you go back to the state map to look at any other state. I have specified independent axis ranges for both line charts so that the trends are more evident.