"TXHOUSING" DATA SET RECOMMENDATION

As the process of choosing an approach to the problem and coding has been documented in the R Script file, I will occasionally refer to the findings in this recommendation without stating the complete motivation that led me to such outcomes.

Firstly, the customers want to know about the growth rate of different cities. The visualization below shows the top 10 cities with the highest growth rate are Houston, Dallas, Austin, San Antonio, Collin County, Fort Bend, Montgomery County, Fort Worth, Denton County, and NE Tarrant County.
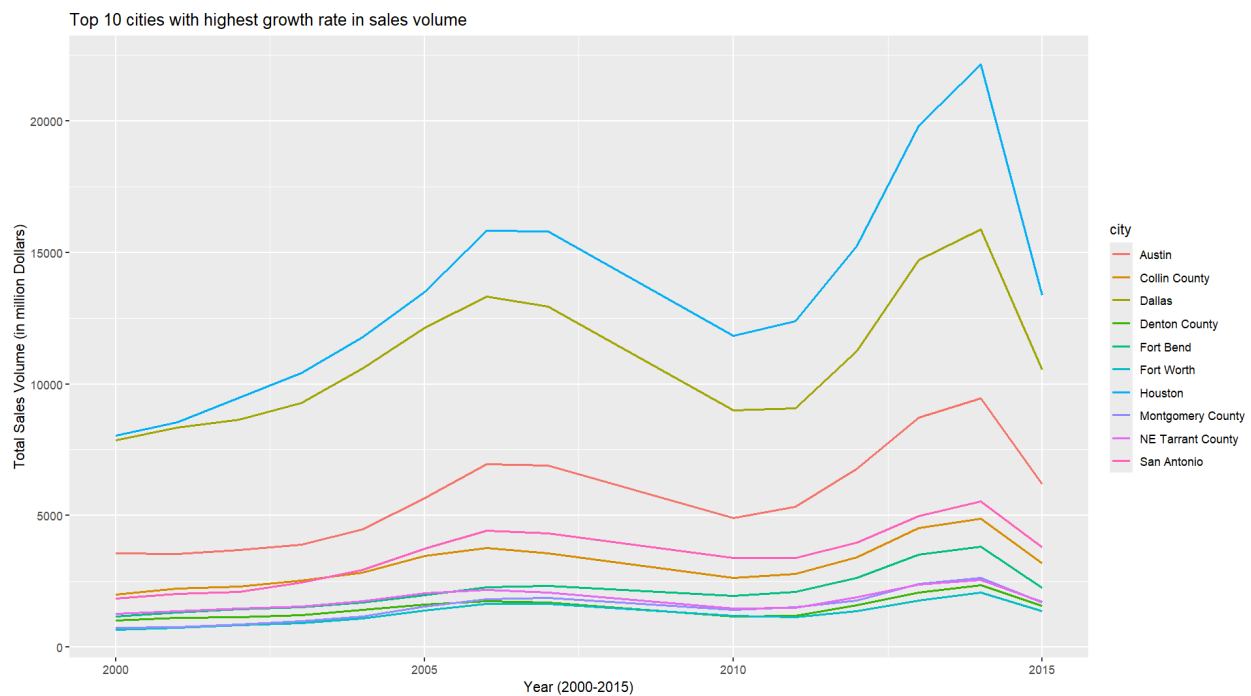
In finding out whether or not these growth rates are affected by outliers, I have taken the average house price and the average median house price of 2000 and 2015 by city and compared the difference between the two measurements. The result (figure on the right) shows that there is usually a gap between the average house price and the average median house price. The gap ranges from $20,000 to more than $70,000. However, when comparing these numbers to the average median house

```
##    city              year mean_price avg_median_price    diff
##    <chr>            <int>      <dbl>            <dbl>   <dbl>
## 1  Austin            2000    191238.           143925  47313.
## 2  Austin            2015    327829.           259000  68829.
## 3  Collin County     2000    199498.          163692. 35806.
## 4  Collin County     2015    332479.           289400  43079.
## 5  Dallas            2000    173338.          133758. 39579.
## 6  Dallas            2015    286742.          227271. 59470.
## 7  Denton County     2000    167655.          143958. 23696.
## 8  Denton County     2015    264182.          228143. 36039.
## 9  Fort Bend         2000    161490.          133308. 28181.
## 10 Fort Bend         2015    307532.           272400  35132.
## 11 Fort Worth        2000    105405.           83733. 21671.
## 12 Fort Worth        2015    189657.          154671. 34985.
## 13 Houston           2000    153285.          113283. 40001.
## 14 Houston           2015    278123.          208471. 69651.
## 15 Montgomery County 2000    164757.           131025 33732.
## 16 Montgomery County 2015    311985.          239029. 72956.
## 17 NE Tarrant County 2000    178596.          138058. 40538.
## 18 NE Tarrant County 2015    298328.          223614. 74714.
## 19 San Antonio       2000    119498.           93208. 26289.
## 20 San Antonio       2015    230703.          190543. 40160.
```

price of a city in a particular year, the difference is not that big. For example, we can look at San Antonio. In 2000, the difference between the two measurements was around $27,000. In 2015, the difference rose to approximately $40,000. When putting these two numbers side by side with San Antonio's average median house price, we observe that $27,000 is not that significant if an average home is priced around $93,000 and $40,000 is not that noticeable when the median is about $190,000. The same can be said with NE Tarrant County, whose gap and average median are roughly $70,000 and $220,000, respectively.

A note when looking at the findings is that we need to also constantly pay attention to how much average house prices and average median house prices have grown alongside the difference between the two. For instance, if the mean price rises by a lot, while the median barely moves, we should be aware of the reason that leads to such disparity. However, in our case, since the difference is not that big of a deal, a possible explanation for the price gaps is that the income of the people increased throughout the timeframe, or people grew to prefer residence in these cities. Thus, there may have been more high-valued houses built with a few being exceptionally valuable, pushing the mean price and median price up.

As the growth rates for these cities are not swayed by outliers, if our customers enjoy living in upper-middle-class neighborhoods, I would recommend the top cities. However, if they prefer otherwise, I would not.

Moreover, to know when is the best time to sell or buy a house, we will need to look at the median house prices throughout the timeframe for each city. The reason I prefer the median house price as an indicator over the mean house price is due to the resistance of the median value. If there is any outlier, the numbers customers receive will be, to some extent, representative of the whole city.

The result (on the right) shows the best time and cities to invest in. The notable thing about this finding is that the time of these prices is not recent and the cities presented are not on our fastest-growing list. The reason may be due to inflation and changes in income or purchasing power of people throughout the years.

The slight exceptions are Paris in 2010 and 2012. Although Paris is not on our list, its median house prices for these two years are low considering they are on par with those of different cities almost a decade ago. If our customers are interested, we can also do some more research about house prices and

| city	time	median
<chr>	<chr>	<dbl>
1 Harlingen	2001-3	50000
2 Nacogdoches	2005-2	54000
3 Nacogdoches	2005-3	55000
4 Paris	2002-12	55000
5 Paris	2010-11	55000
6 Port Arthur	2002-3	55000
7 Brownsville	2000-4	55800
8 Nacogdoches	2005-4	55800
9 Paris	2001-1	56000
10 Port Arthur	2000-7	56200
11 Paris	2000-11	57000
12 Paris	2012-10	57500
13 Abilene	2000-3	58100
14 Paris	2002-11	58300
15 Paris	2000-5	58600
16 Abilene	2000-2	58700
17 Harlingen	2001-8	58800
18 Port Arthur	2000-10	58800
19 Abilene	2000-10	59300
20 Abilene	2002-11	59400

house quality in Paris in recent years with the hope that there are houses today with low prices as in the past.

The outcome of the best time and cities to sell houses, unsurprisingly, consists of more median prices of recent months and years, and those prices also mostly come from cities with a rapid growth rate.

```
   city            date median
   <chr>          <dbl>  <dbl>
 1 Collin County 2015. 304200
 2 Collin County 2015. 300400
 3 Collin County 2016. 292600
 4 Collin County 2015. 291400
 5 Collin County 2015. 285800
 6 Fort Bend     2015. 284200
 7 Collin County 2015. 283400
 8 Midland       2014. 283100
 9 Fort Bend     2014. 282300
10 Fort Bend     2016. 280400
11 Fort Bend     2015. 279700
12 Fort Bend     2015. 277300
13 Fort Bend     2015. 276100
14 Fort Bend     2015. 273500
15 Collin County 2015. 272200
16 Collin County 2014. 271600
17 Collin County 2014. 271500
18 Austin        2015. 271200
19 Fort Bend     2015. 270800
20 Austin        2015. 270300
```

Finally, we will discuss the relationship between the active listings and the median and mean house prices. The following two graphs show the relationship between listings and median house prices and between listings and average house prices, respectively. Both graphs display an unusual spike at around $150,000 to $200,000. This, subsequently, spike conveys there tend to be more listings for houses with mean and median house prices clustering around this area on the graphs. The reason for the similarity may be due to the mean and median for the houses in this data set do not differ that much, which has already been foreshadowed by earlier findings in this report.

One possible factor that causes such an unusual spike is the similar demographics in these cities. Since most people in our society come from the middle class, including the lower and upper middle class, there should be more houses listed in the market that are aimed towards this class rather than the lower or upper class. Since the spike is also within the price range that is considered affordable to middle-class people, we can hypothesize that demographics is one of the main factors that cause such a surge in listings across different prices.
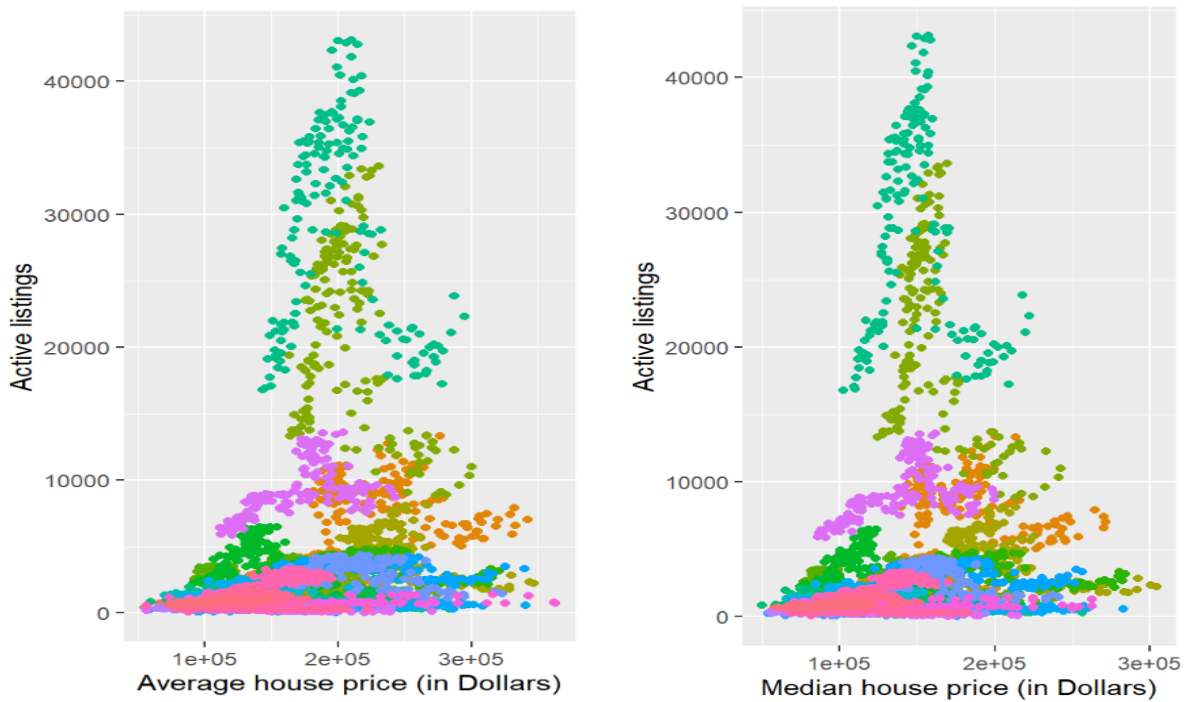
Ultimately, the final recommendation is to invest in the top cities if our customers' financial conditions allow. The perks of these houses revolve around growing neighborhoods and values. People can invest in these cities regardless of their purposes: from doing business to seeking a permanent settlement. However, the downside of these types of houses is the availability. Compared to the listings in the lower range, there are fewer active houses for sale in these cities. Still, if the limited availability appears to be a problem for customers, we can always go down the price range a little bit in exchange for more active listings. Nevertheless, if our customers' finances do not allow for such purchases, depending on the actual budget, we can advise them to have a look into houses lying within their desired price range. Since there are significantly more active listings in the lower to middle price range, customers will have more potential options.