Portfolio Project: Analysis of Cabinet Sales in Tucson Market

Brett Amione

Colorado State University – Global Campus

Cabinet Project

In 2017, the top 10 home building companies reporting closings on 168,060 homes in the United States (Builder, 2019). The process for buying a new home from a production builder typically begins with buyers selecting a plan type and structural options with an on-site sales associate and then selecting interior upgrades with a design team. Homebuilders have a great deal of control over what customers can select and home to offer those products to the buyers.

One of the largest production builders, Richmond American Homes, offers thousands of upgrades to their buyers on each home. Although this provides the company with many revenue opportunities, it also presents some challenges. The variety of upgrades can be overwhelming to buyers and to interior designers. It can also be difficult to track from a bidding and systems setup perspective.

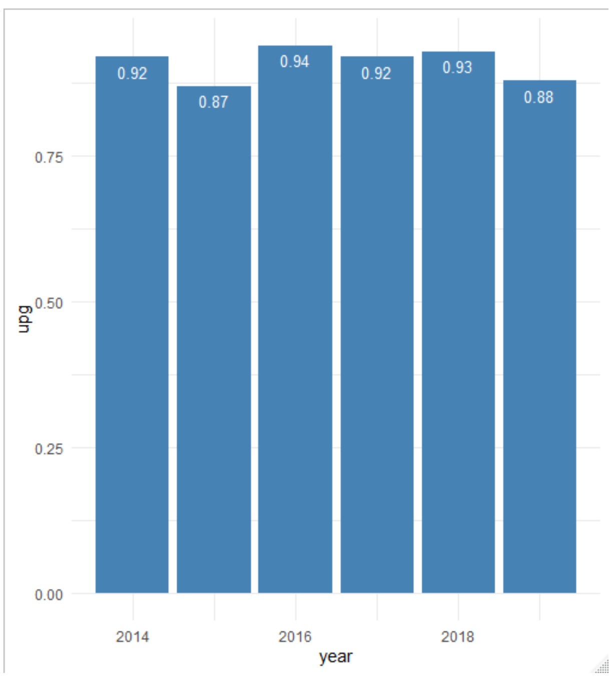
As part of their 2018 revenue tracking, the company decided to investigate one of their top revenue categories, cabinets in the Tucson market. Cabinetry is the third largest contributor to annual revenue and cabinet accessories are the seventh largest contributor. The primary question that the company wants to investigate is whether or not the designers have had continued success in selling cabinet upgrades and if there was a significant change in sales between 2018 and 2019.

To investigate this, an analyst has pulled all the cabinet sales from 2018 and 2019. This output was returned as a .CSV with detailed information that goes beyond the scope of this investigation. Because the amount of lines returned is only in the hundreds, the analyst can load this .CSV file in Excel, create a copy, and start to clean the data. Using their subject matter knowledge, they know that some cabinets which have been sold were not meant to be sold and so these lines are removed from the dataset. After removing columns that holds information not relevant to the business question, they are left with just a column that indicates whether the cabinet sale was an upgrade from the base cabinets and what year date it was sold on. This information is then sorted into a count by year and converted to a proportion. The analyst finds that in 2018, 92.56% of buyers upgraded their cabinets and that in 2019, 87.58% of buyers upgraded their cabinets.

Upon seeing this difference in cabinet sales, the analyst goes back to the initial data source and pulls upgrade percentages for each year this information has been logged. Again, the information is copied and cleaned up. The proportions for each year are calculated.

Summary Statistics

The data is pulled, and the analyst sees that there is not much variation during the last 6 years that upgrades have been recorded. The lowest take-rates were in 2015 and 2019 which does raise the possibility that the company has sold a statistically significant low number of upgrades in 2019. To further describe the data, the analyst runs summary statistics which are recorded in Table 1. Quickly, the analyst can see that the 2019 year is below the average and the median and that it is approximately one standard deviation below the average. But what does this really mean for the data?



mean	0.91
median	0.92
max	0.94
min	0.87
std dev	0.02

Table1 : Summary Statistics of Take-Rates

Plot 1: Upgrade Take-Rates per Year

Hypothesis Testing

While these summary statistics are useful, the analyst knows to prepare a hypothesis test to better understand if this drop is significant. The analyst decides to run a two-proportions t-test using the following code in Rstudio:

prop.test(x=c(199,134), n = c(215,153), alternative = "great", conf.level = 0.95).

The null hypothesis is that 2018's upgrades were not statistically significantly greater than 2019's upgrades. The alternative hypothesis is that 2018's upgrades were statistically significantly greater than 2019's upgrades. The output of the prop.test function is as follows:

```
        2-sample test for equality of proportions with continuity correction
Data:   c(199,134) out of c(215,153)
X-Squared = 2.0265, df = 1, p-value = 0.07729
alternative hypothesis: greater
95 percent confidence interval:
  -0.008650071          1.000000000
sample estimates:
   Prop 1         prop 2
0.9255814      0.8758170
```

The results suggest that there is not enough evidence to reject the null hypothesis. If this is the case, then the analyst can report that even though the company has sold 7 percent fewer cabinet upgrades, that this does not constitute a significantly drop.

Additional Questions

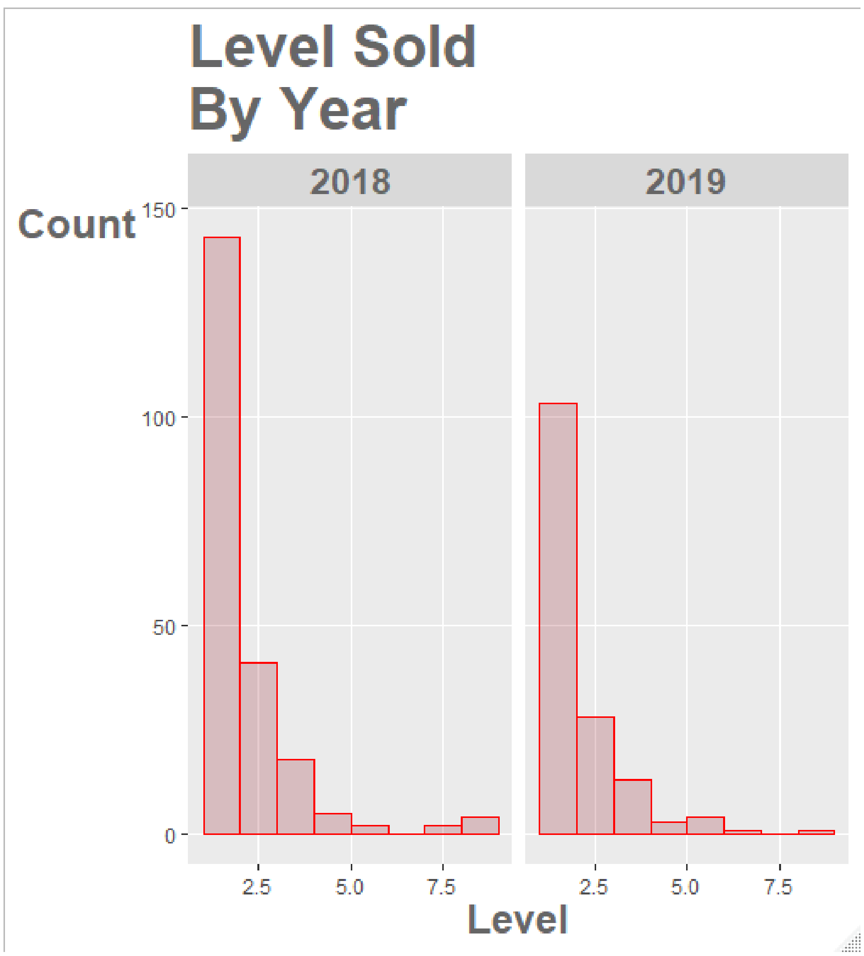
Having knowledge that the number of cabinet upgrades sold is not significantly less in 2019 is comforting but there are other questions that can also be asked. In the Tucson market there are 9 cabinet levels. Each increase in level selected generates higher revenue for the company. Richmond also wants to know if the cabinet levels sold are significantly different.

The analyst loads another table that includes information on which cabinet level was sold by year. Using the dplyr package, and running the following code produces some summary statistics on levels. These are as follows

```
group_by(df2, Level) %>%
        summarise(
                count = n()
                        mean = mean(Level, na.rm = TRUE),
                        sd = sd(Level, na.rm = TRUE)
        )
```

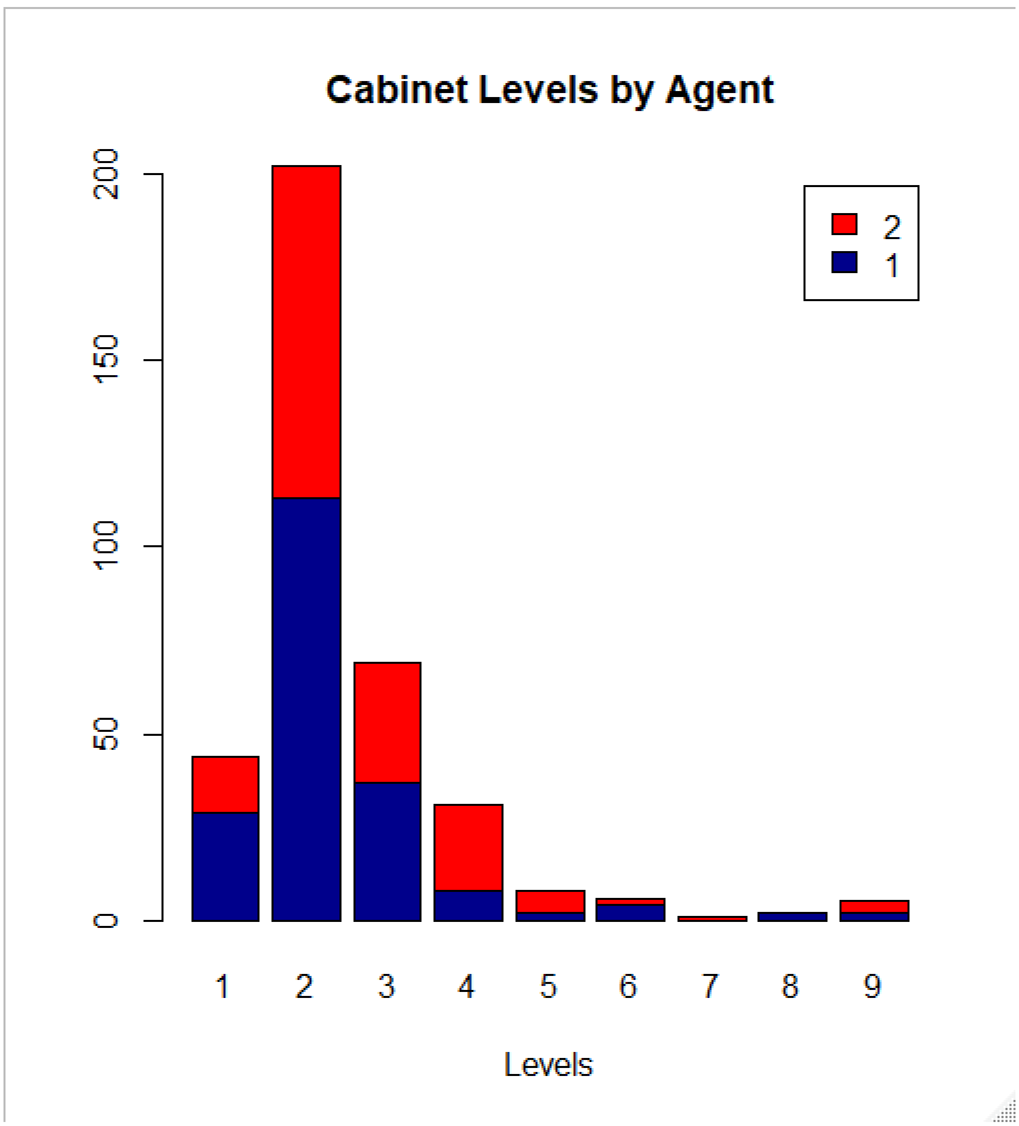
Level

To better visualize levels sold between the two years, the analyst uses the ggplot2 package to produce the visuals shown in Plot 2. The findings show that there is some variation but that the cabinet levels sold have not changed very much in the past two years.



Plot 2

Another good question to ask is if one designer is selling more or a specific cabinet than the other. By generating a stacked histogram using all the cabinet sales in the past 8 years, the analyst can quickly visualize who is selling what. What we can see is that most sales are distributed across levels 1, 2, 3, and 4 and that upgrades beyond this point do not sell particularly well. It is also interesting to note that designer #2 appears to have better luck selling beyond level 3. Perhaps we can assist the other designer with some selling techniques to drive more revenue?



Summary

In summary, statistical analyses have shown the company that while the number of sales has declined from 2018 to 2019, that the drop is not statistically significant. The company also is aware of the general distribution of cabinet sales and can see that levels 1-4 are the top sellers. Finally, we looked to see if there was any difference in the way designers were selling cabinets. It appears that both designers sold a proportionally similar amount of cabinets per level with the exception that designer #2 had more success in selling the more expensive level 4’s and 5’s.

In the future, the company may want to perform additional analyses on cabinet sales. For example, does the Spec Level of a home have any correlation to cabinet upgrades? Homes are split between 6 different specification levels (Spec S, 1, 2, 3, 4, and 5). Spec S has the least expensive included features whereas Spec 5 homes have the most expensive included features. It may be worthwhile to see if these Spec types upgrade differently. Another interesting study would be to look at each level more closely to determine which cabinets in each level are selling the most? Rather than bidding 5 different styles of Cabinet Level 6 with multiple colors per style, maybe the company could better negotiate their deals on cabinet levels 2, 3, and 4 to take advantage of their high sales volume in those categories?

It’s clear that this study is just scratching the surface of a cabinet analysis. The next steps will be to share this information with decisionmakers to apply some of this new knowledge to selling and bidding strategies.

References:

Builder. 2019. The Top 100. Retrieved from https://www.builderonline.com/builder-100/builder-100-list/2018/