HOUSING ANALYSIS PROJECT SUMMARY

Prediction, Informative Recommendations, Plotly Graphs, Frequency Word Cloud, Sentiment analysis, Topic Modeling, Recommendation Website

Introduction

Housing sales have been a hit after covid. According to CNBC,[1] Sales in April 2021 were 33.9% higher than in April 2020. The supply of homes for sale at the end of April was down 20%. There were 1.16 million homes for sale, representing a 2.4-month supply at the current sales pace. High demand and rock-bottom supply will continue to push prices higher. The median price of an existing home sold in April was $341,600, an increase of 19.1% from April 2020. That is both the highest median price on record and the largest annual increase on record. Much of that large increase in the median price is due to the mix of homes that are selling. There is far more activity on the higher end of the market, where supply is more plentiful, and very little activity on the low end, where the shortage is most severe.

As an example, sales of homes priced between $100,000 and $250,000 were unchanged from a year earlier. Sales of homes priced between $750,000 and $1 million rose 146% from a year earlier, and sales of homes over $1 million jumped 212%.

Our Project: The differences between the demand in high-end homes and low-end ones prompt us to consider the social inequality in purchasing houses, especially in races. We also looked at the discussion of the overall US housing market. We tried to find what are the topics that people may discuss, and what are people's common sentiments. Besides all the analysis part, we incorporated data in crime rates, city population, public high schools, housing locations, people's budgets, longitude and latitude, and future housing price predictions to shuffle our the ideal locations for people who wish to invest in houses.

[1] CNBC, April existing home sales drop, marking three straight months of declines: https://www.cnbc.com/2021/05/21/april-existing-home-sales-drop-marking-three-straight-months-of-declines.html?__source=iosappshare%7Ccom.tencent.xin.sharetimeline

Racial Inequality in Housing

Races Compositions in Poverty and Total Population in California

Comparing these two graphs, we are able to see that compared to the composition in total population, white people and Asian people have a relatively low poverty composition whereas Hispanic or Latino(of any race) and people of Black have a relatively bigger composition in poverty.
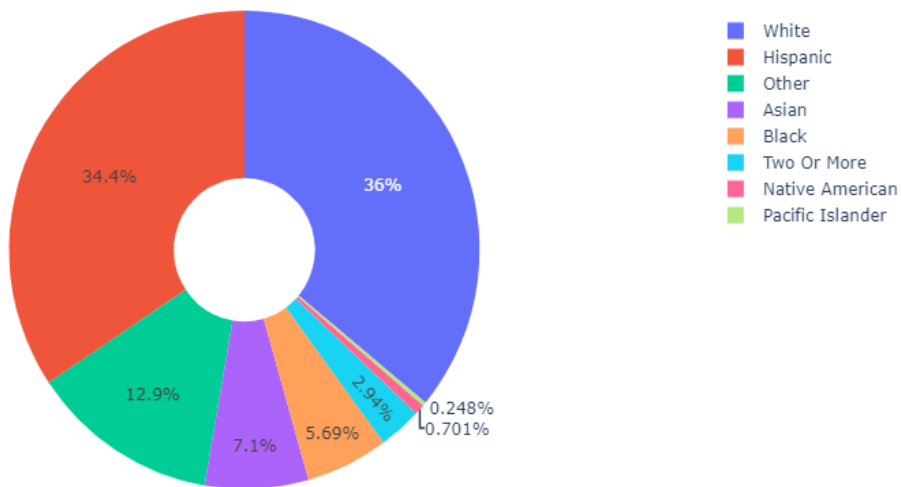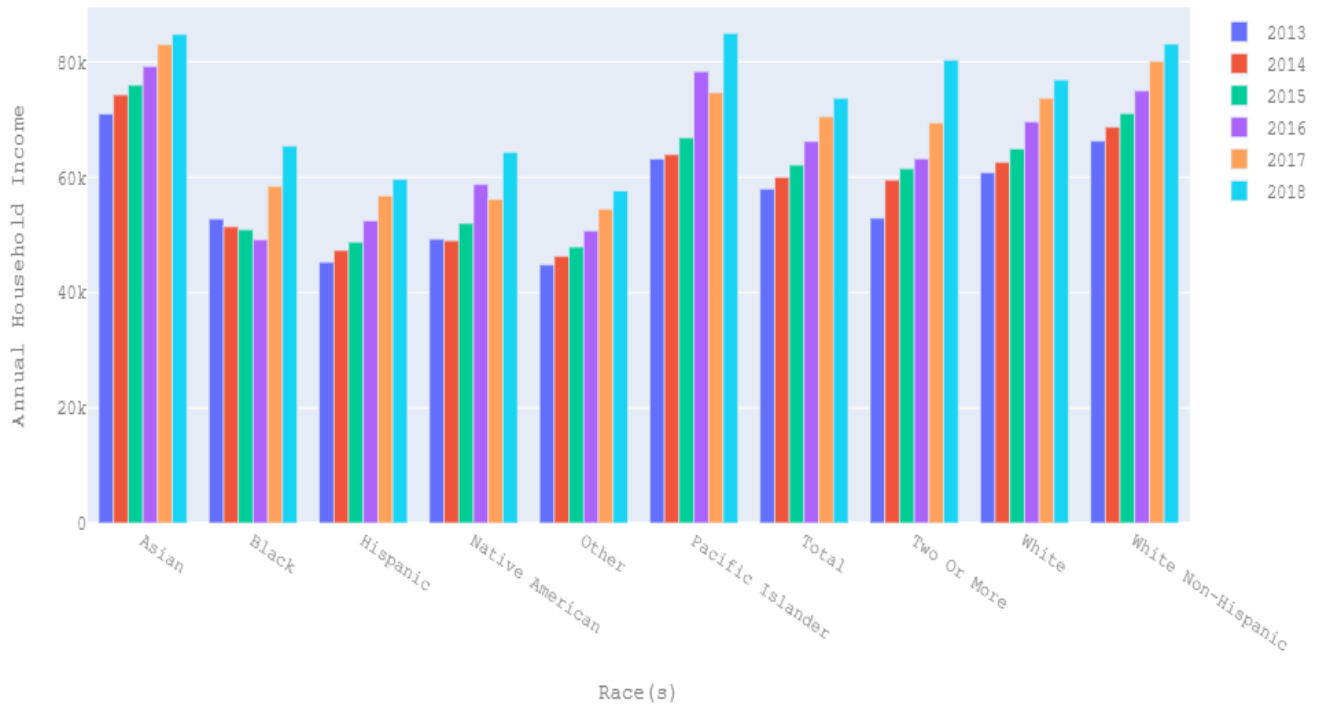

The left picture comes from data of California Races Composition in Poverty in 2018.
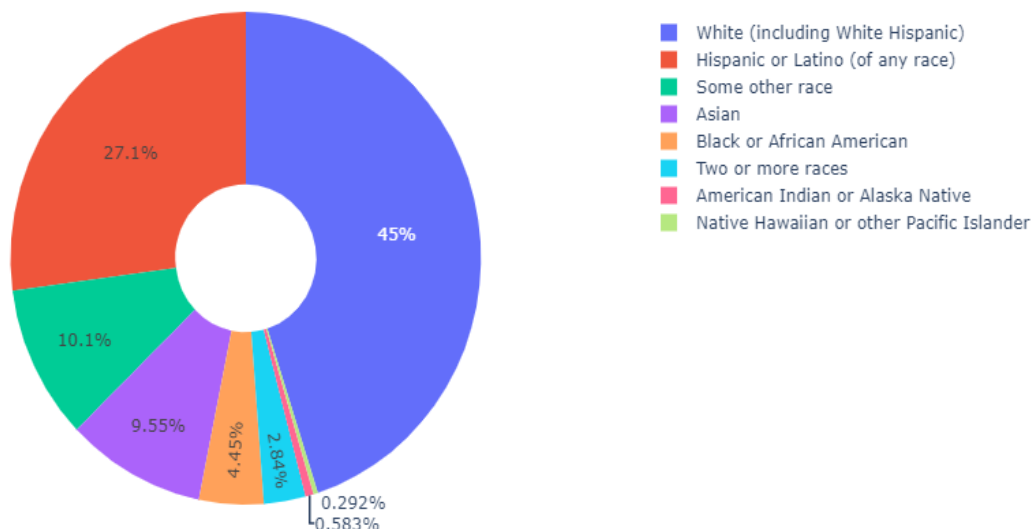
This left picture comes from the population composition of California in 2018.

Races Income Distribution in California

Annual Household Income Mean By Races from 2013-2018 in California

This graph shows the annual household income mean distribution of different races in California from 2013 to 2018. We can see that Races of White, White Non-Hispanic, Asian, Pacific Islanders have significantly higher annual incomes than other races in California.

This income distribution corresponds to the trend identified earlier that these races of higher annual mean incomes have lower poverty compositions.

Housing Prices Changes in Race-Popular Regions in California

Comparing housing prices in places where each race has the highest composition, we can see that housing prices all go up in an upper ward trend. Nevertheless, housing prices differ across races. The highest housing price of Native American popular areas is around 700k, of Hispanic/Latino is 250k, of Black/African American is 1.5M, of White is 800k, of Asian is 2.5M. Still, in this setting, Hispanic/Latino displays a lower purchasing power.
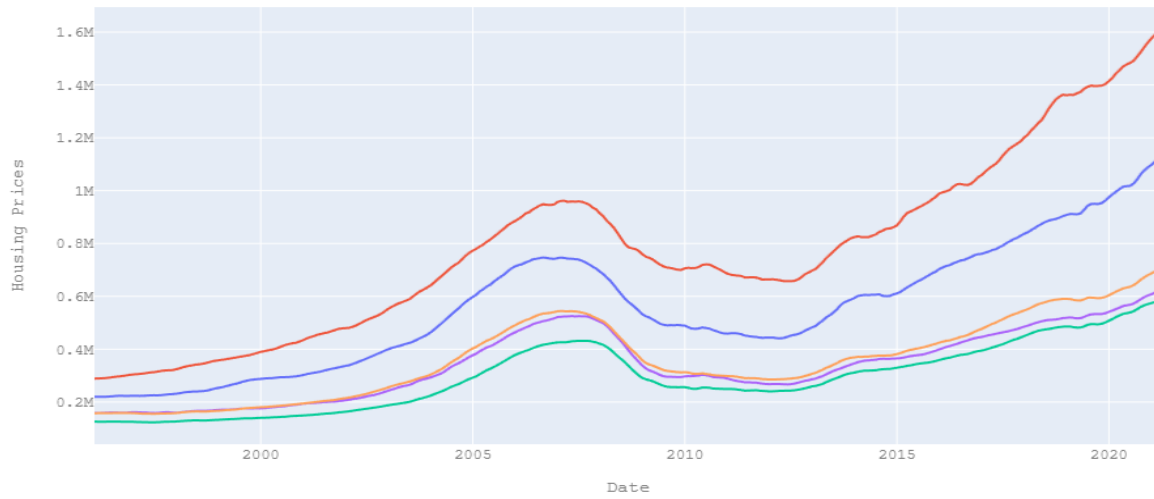
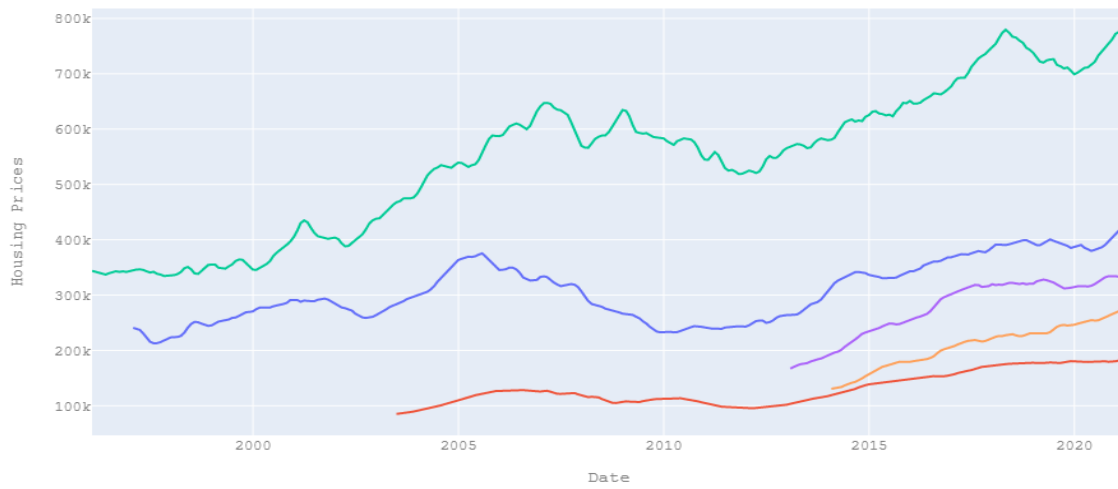Housing Prices Change for Top 5 Native American popular areas
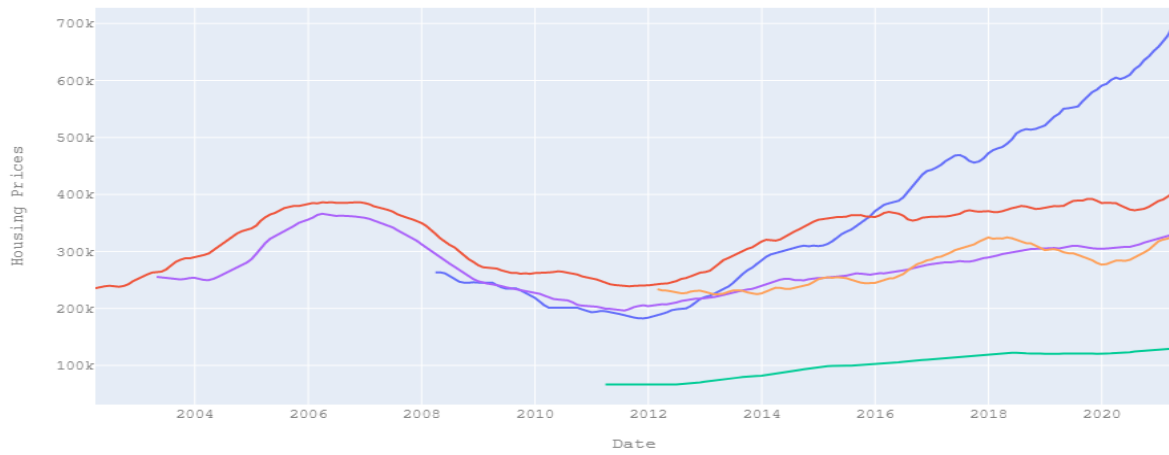
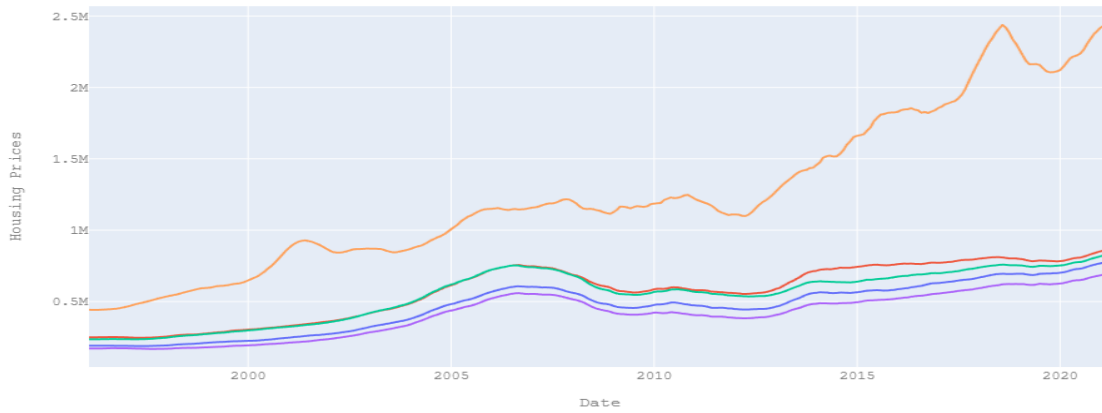Housing Prices Change for Top 5 Hispanic/Latino popular areas
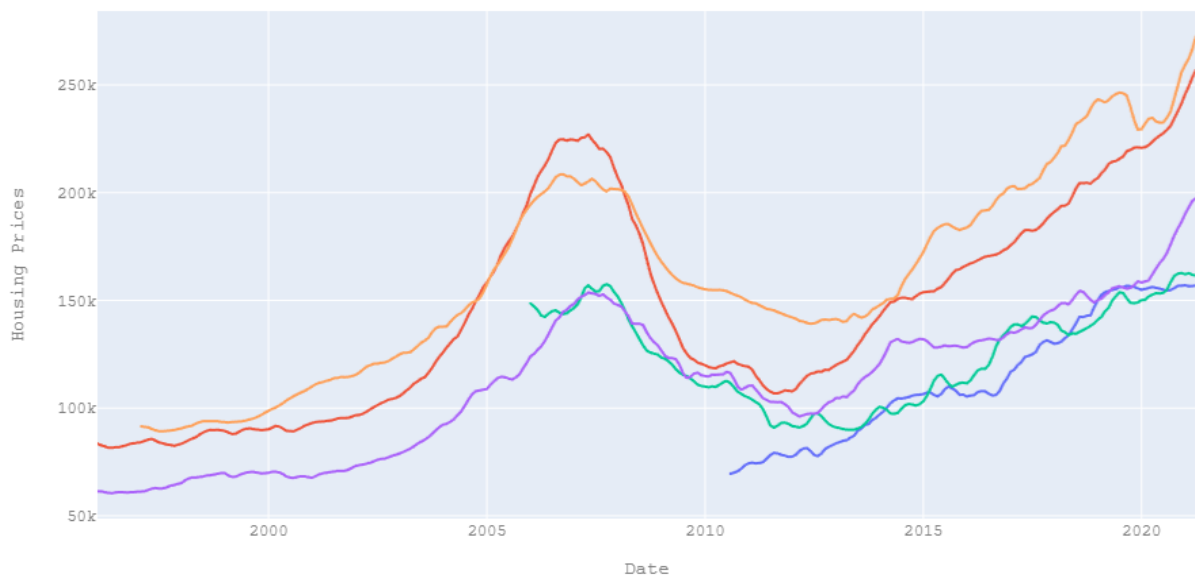
Housing Prices Change for Top 5 Black or African American popular areas

Housing Prices Change for Top 5 White popular areas

Housing Prices Change for Top 5 Asian popular areas

Frequent Words in Reddit "House" Discussion

We web scraped the house discussion from Reddit from 2013 to 2021/2/1 and did a vast series of data cleaning and preparation on the text data. Then by the 1-word frequency word cloud picture according to word frequency appeared in discussions, we surprisingly find that words related to music appear such as "song", "mix", "track". By the 2-word word cloud, besides music, we find words such as "french house", "tech house", "new house".

Sentiments in "House" and sub "House" Discussions

We then looked at the distribution of different emotions in house discussions:

We can see that Joy and Trust are the most frequent emotions, whereas disgust and surprise are the least frequent emotions.

The left above graph shows what are the words that appear in the sentiment of joy: we can see that when people discuss house with the emotion of joy, they mention "party", "cheer", "festival". The right above graph shows what are the adjective words that appear in the sentiment of "trust" in house discussions: they use words that are used in social relationships relating to houses, such as "recommend", "advice", "friend", "hope".

Then we subset the overall house discussion datasets into smaller datasets by searching for keywords important in housing, including "buy", "sell", "race", "price". When we find sentiments for each subset of House discussions, different compositions of sentiments for each sub dataset appear. The bar graphs on the left show the frequency of each sentiment in each dataset. The word cloud pictures on the right are the frequent words in the top sentiment for each dataset. In terms of buying and selling, people reveal more joyfulness than sadness, which might correspond to the earlier part where there are trends of selling and buying. Relating to races and prices, people show the top sentiment of "trust", and descriptive words in the "trust" relate to supportive social relationships such as "embrace", "friend" and also housing quality such as "guarantee", "proof", "good".
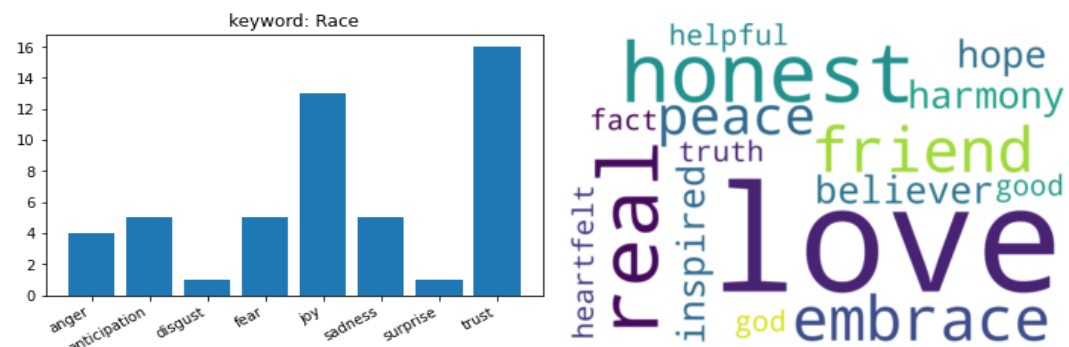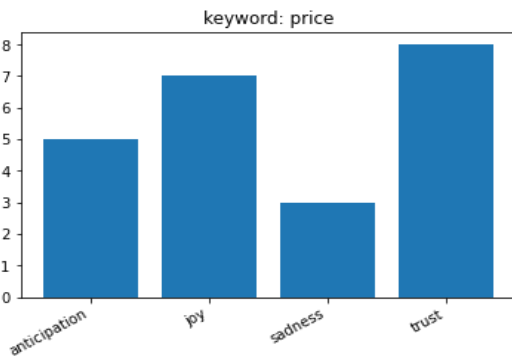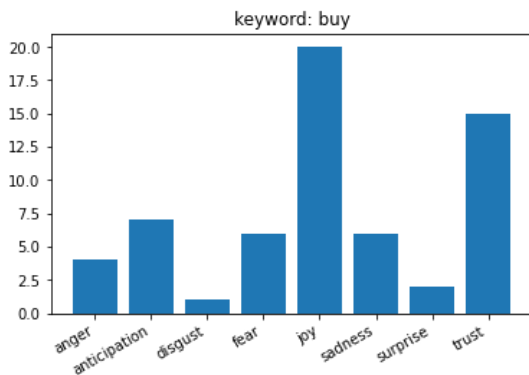
keyword: price
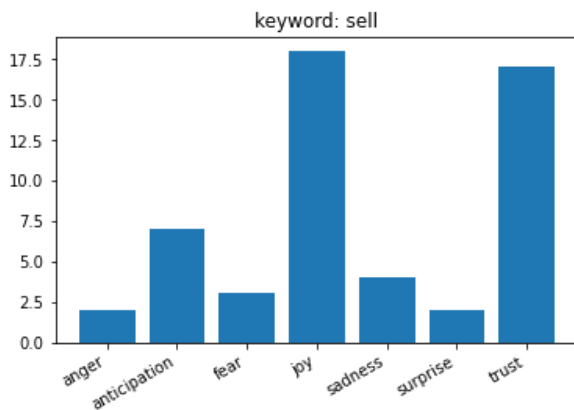

keyword: buy

keyword: sell

Topic Modeling in "House" Discussions

We first do topic modeling in the overall house discussions. We may summarize the first topic as "life in house", the second topic as "music", the third topic as "mood".

	Topic 0	Topic 1	Topic 2
0	life	looking	mind
1	come	guy	tell
2	comment	really	life
3	party	love	day
4	edit	know	love
5	original	mix	party
6	remix	like	like
7	mix	song	house
8	love	track	xb
9	amp	house	amp

Then, when doing topic modeling under each sentiment we get from the overall house discussions, we have interesting topic modeling results in the following two sentiments:

	Topic 0	Topic 1	Topic 2
0	house genre	driving mad	track mix
1	facebook page	shadow child	year ago
2	long shot	looking track	mix soundcloud
3	house tech	nick warren	new mix
4	chill house	melodic house	house techno
5	nick warren	mix soundcloud	thanks advance
6	dj koze	new mix	acid house
7	year ago	long shot	house tech
8	twitch tv	house set	classic house
9	hey guy	classic house	house playlist

Topic modeling by Sentiment Fear: the first topic may be related to houses equipments whereas the second topic is related to different types of houses. These two topics reveal that people may have negative feelings in either looking for houses decorations or experiencing different types of houses. The third topic is related to house tech and music.

	Topic 0	Topic 1
0	looking deep house	looking track similar
1	looking deep	sacha robotti
2	house tech house	looking house
3	house tech	guy looking
4	thanks lot	hey guy looking
5	similar track	hey guy
6	hello guy	similar vibe
7	tech house	house looking
8	track looking	house label
9	deep house	house tune
10	hi guy	house playlist
11	progressive house	looking good house
12	looking house track	looking good
13	looking house	looking new
14	house track	good house

Topic modeling by Sentiment Disgust: twos topics all apear to be related to looking for houses and decorating houses. We surprisingly find there are positive words such as "looking good" and "looking new". These two topics may reveal that people may have negative feelings in how to make the house look better and how to make the houses look better.

Appendix :

Average Annual Household Income By Locations in California in 2018

Races Composition by Counties in California

Legend: ■ White ■ Other ■ Asian ■ Black or AfricanAmerican ■ Native American ■ Hispanic or Latino(of any race)

RACE COMPOSITION BY CITIES

Legend:
- White[1]
- Other[1][note 1]
- Asian[1]
- Black or AfricanAmerican[1]
- Native American[1][note 2]
- Hispanic or Latino(of any race)[2]