

# Housing In Bay Area Project Report

Danielle Hu

College of Arts and Sciences

University of San Francisco

San Francisco, CA 94117

Email: eggchul1024@gmail.com

**Abstract**—Housing has always been an essential concern for people the whole time. In this paper, I present my website, which visualize data collected from popular housing websites. Readers can see information of some cities, as well as the monthly median value of a house with different number of bedrooms change in different city. The website also shows the monthly inventory change of new listing, monthly inventory change of pending listing and monthly change of Bitcoin.

## I. INTRODUCTION

A recent news, published in SFGATE[1], reported that a number of home buyers camped overnight nearby the listing office for a chance to buy a townhouse, which cost about \$1.2 million. The housing market is getting extremely intensive in some cities in the bay area, which might due to the impact from COVID-19. Ever since the pandemic of COVID-19, many people start working remotely, companies are moving out of San Francisco. In March 4 this year, East Bay Times[2] reported that there is a 30 percent increase in people moving out of California compared from 2020 to 2019. It also mentioned that the historical pattern of migrations out of California. This draw my attention to the housing market, if COVID-19 could change the housing market pattern in bay area as well.

In this paper, I present my website, which visualize data collected from popular housing websites, including **Zillow**[3], **Realtor**[4] and **Kaggle**[5] to analyze the recent housing market in the San Francisco bay area. Information of some randomly selected cities in the San Francisco bay area is collected from **Niche**[6] and presented in the website.

### A. Goal

This project aims to analyze the recent housing market in the San Francisco bay area. Some of the cities listed on Niche's[6] "2021 Best Cities/Suburbs to Live in the San Francisco Bay Area" were chosen as the target cities for this project.

### B. Project Objectives

- Help reader get to know about some cities in the bay area. Readers will be able to identify the ranking of each selected cities as well as the city's population and description.

- Provide an interactive line chart for readers to investigate if the number of bedrooms that a house have would relate to the house value change of different cities from the historical data.
- Identify if in long term, will a house with more bedrooms would be more valuable.
- Identify or investigate if there is any pattern about the trend of inventory new listing over time.
- Compare Bitcoin monthly change and the monthly pending listing of inventory change of two areas in bay area.

## II. RELATED WORK

The related work goes here.

### A. About Housing In San Francisco Bay Area

- In 2018, San Francisco Planning Department[7] published a report about housing needs and trends in San Francisco. This report analyzed housing related data using pie charts, bar charts, line charts and map to identify the some major ongoing challenges at that period, also highlight the policy of rent control is actually affecting the housing market in San Francisco.
- Redfin[8], as one of the most popular real estate brokerage, it visualizes its real time housing data with Tableau and shows the graphs in their website's data center page.

### B. Related D3 Work

- There is a sample of bar chart in Codepen by Jesse Howe[9]. This sample is later used as a reference in implementing the diverging colored bar chart in website when analysing monthly new listing of inventory change.
- Philipp Koytek published "Bar Chart with Time Scale"[10] in Observable in 2020. This page helps in supporting the bars in aligning with time label on x-axis.
- Susie Lu's D3 Legend[11] and D3-annotation[12] also used in visualizing the bar chart.
- D3.js Graph Gallery[13] provides a sample of zoom with axis.

## III. APPROACH

The research was originally started by researching some cities listed in the Niche[6] website. After having these cities' information and visualized the ranking in scatter plot[Fig.1],

I then started the data cleansing process for other following charts. Due to San Francisco belongs to the city level rank, it is not included in the first chart, however, it is covered in other charts.

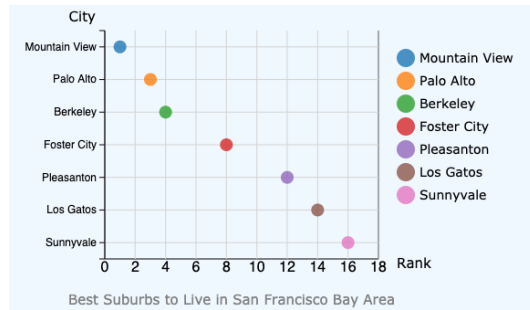


Fig. 1. Best Suburbs to Live in San Francisco Bay Area 2021.

## A. Data Cleansing and Data Visualization

1) *Price Trend*: Data for the price trend section is collected from 5 csv files downloaded from Zillow[3]. In Fig.2, I combined those csv files into one data frame and then performed data processing duties including filtering, transforming, renaming and formatting.



Fig. 2. Jupyter notebook cleansing data for price trend section

With the finalized data, I have the first version of interactive line chart for alpha release. Shown in Fig.3.

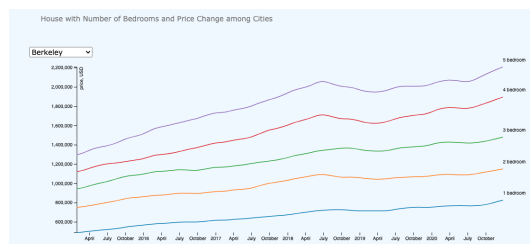


Fig. 3. Monthly House Value Change of Cities – Alpha Release

2) *New Listing*: The original design for this section is using data collected from Redfin[8] and visualize monthly inventory change of different cities over time. Fig.4 shows the data preparation process for analysing listing inventory. Due to the data set from Redfin has not enough data, I used data from

Realtor[4] as substitute and visualize new listing change based on groups of cities instead of each city. I selected only data for "san-francisco-oakland-hayward" and "san-jose-sunnyvale-santa-clara".

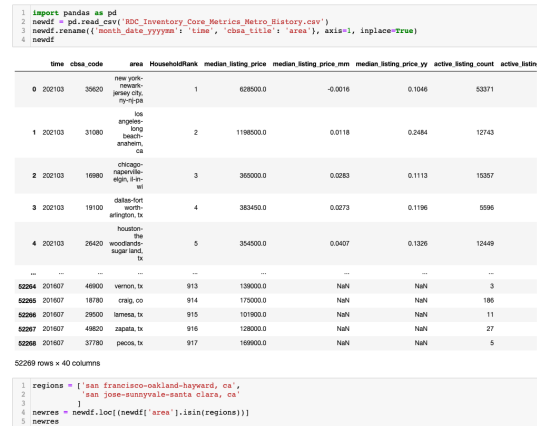


Fig. 4. Jupyter notebook cleansing data for listing inventory data

Fig.5 is the first version of inventory monthly new listing change bar chart published in alpha release.

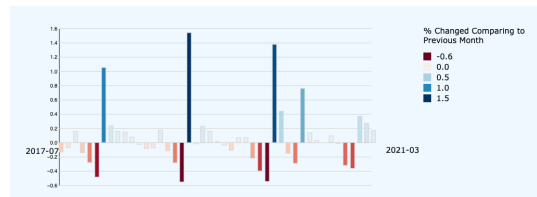


Fig. 5. Monthly New Listing Change in Ratio – Alpha Release

3) *Pending Listing and Bitcoin*: In the last section of the website and research, I put the monthly pending listing change along with Bitcoin monthly daily change, using box plot, in one chart. This is the first time for me to combine two different visualization methods together and present in one graph. Bitcoin daily change data is collected from Kaggle[5], before visualizing, I performed data trimming and calculation on the original data set. Steps are shown in Fig.6. This process helps to maintain the data consistency when implementing visualization tools later. All the change are shown in ratio, due

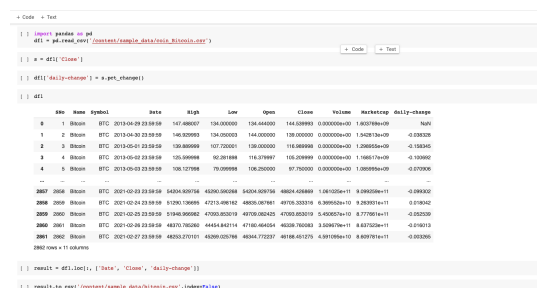


Fig. 6. Jupyter notebook cleansing data for "pending listing and bitcoin" section

to listing count and Bitcoin price are represented in different measure of value. Using ratio for comparison is a proper way to connect these two data sets. Fig.7 is the first version of this line and box plot chart which is published in beta release.

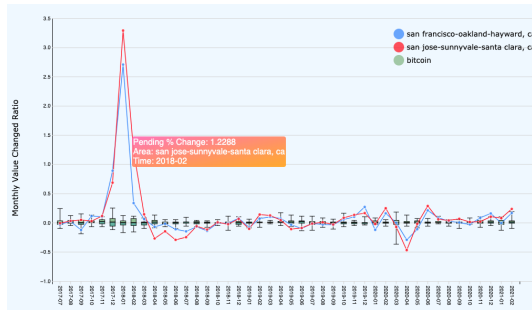


Fig. 7. Monthly Change of Pending Listing and Bitcoin in Ratio – Beta Release

### B. Correction/Improvement

There are two releases during the whole development process before this final release. In each of the release, I went through review from peers and instructor. Based on the feedback, couple of corrections are made to improve this website. Changes that I made including but not limited are adding interactive features such as zooming and hovering, adding d3-annotation and y axis labels, enlarge visibility of font and lines, etc. When implementing the diverging bar chart, making bars correctly aligning along with labels on x-axis was an issue, shown in Fig.5. In the correction process, I used D3 scaleBand and timeScale together since scaleBand could calculate bandwidth automatically. Bars then could be drawn in place.

## IV. RESULTS

1) *City Ranking*: In Fig.1, readers could identify each cities' rank and population. Mountain View is the top 1 city among the other selected suburb cities. And Sunnyvale is at NO.16. This section resolve the first objective.

2) *House Value, Bedroom Count, City*: This price trend section resolves objective 2 and 3. We can see that in San Francisco[Fig.8], house with 5 bedrooms had been increasing in a very fast pace comparing to other houses with fewer number of bedroom. But since closing to the end of 2019 when COVID-19 pandemic's impact starts showing. There is an outstanding drop, or turning point in 2020 January for house with 5 bedrooms in San Francisco.

On the contrary, if we look at the data of Sunnyvale[Fig.9], the value of houses with more than 1 bedroom had been growing very fast until mid 2018. Before the pandemic began, value all types of houses began to drop down. The reason of this breakdown might due to the housing bubble as Patrick May[14] reported in early 2018, when housing price was getting extremely high and people could hardly afford the price. However, when closing to the end of 2019, the house value started to bounce back, or recover. Later in August 2019,

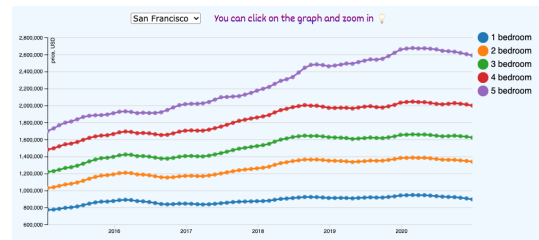


Fig. 8. Monthly Change of House Value – San Francisco

Wolf Richter of Wolf Street[15] published a report with graphs to have a deeper analysis of the bay area housing market. He also mentioned that many tech companies announced IPO during that period, which causing a number of billionaires and millionaires appear.

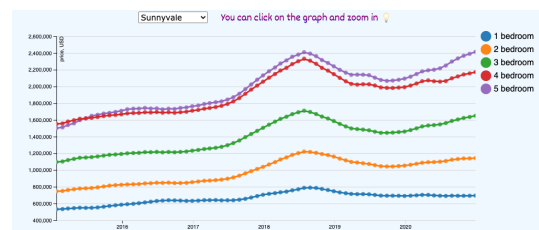


Fig. 9. Monthly Change of House Value – Sunnyvale

3) *New Listing*: This section resolves objective 4. Readers can identify the pattern of new listing inventory change in housing market that normally will have a peak in January and starts going down to the lowest in December. However, during this pandemic period, for both areas(Fig.10 & Fig.11), changes in May, June, July and August in 2020 are different from other years at the same period. As Wolf Richter reported in the article[16], "San Francisco is now flooded with homes for sale. "Active listings" surged to 1,344 homes in the week ended July 5, up 65% from the same week last year". One of the reason might due to the increase amount of people work-from-home, they moving to suburbs and smaller towns. We could also recall the housing market heat mentioned in the last section of price trend.

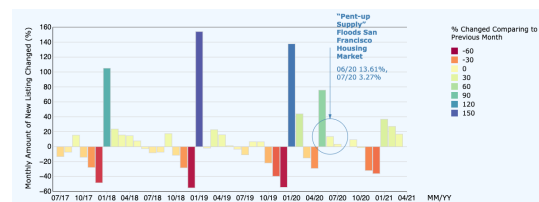


Fig. 10. Monthly Change of New Listing – Area of San Francisco, Oakland and Hayward

The similar situation happens to area of San Jose, Sunnyvale and Santa Clara. Both these areas have a significant historical low new listing in January 2020. This pandemic definitely has changed a lot bay area people's life.

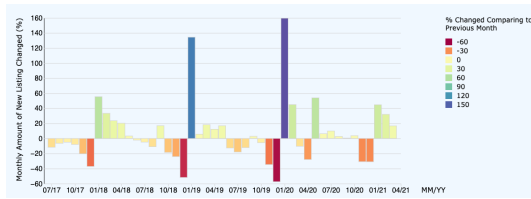


Fig. 11. Monthly Change of New Listing – Area of San Jose, Sunnyvale and Santa Clara

4) *Pending Listing and Bitcoin*: This section resolves objective 5. Fig.12 presents an overview of the monthly inventory change ratio of pending listing for both areas motioned in the previous section, as well as monthly view of daily Bitcoin value change in box plot format. With the zoom in and drag

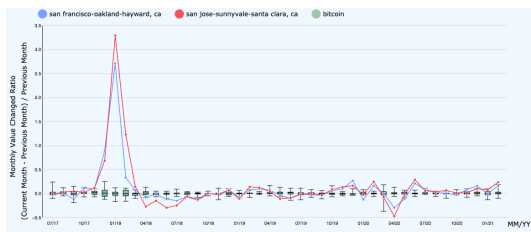


Fig. 12. Monthly Inventory Change of Pending Listing and Monthly Daily Change of Bitcoin – Overall

feature, readers can enlarge the graph to check a specific time period and see the data detail. Fig.13 and Fig.14 are the effect pictures showing when a user zooms in the abnormally high change peak of pending listing in January 2018 and significant low in April 2020.

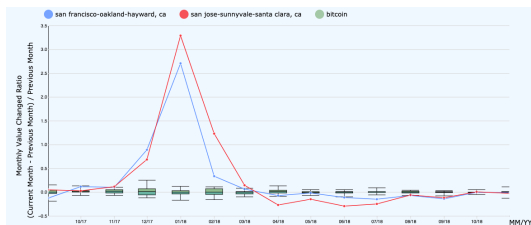


Fig. 13. Monthly Inventory Change of Pending Listing and Monthly Daily Change of Bitcoin – Zoom In 1

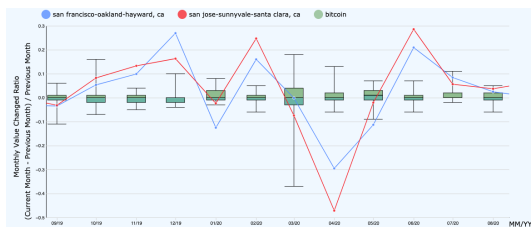


Fig. 14. Monthly Inventory Change of Pending Listing and Monthly Daily Change of Bitcoin – Zoom in 2

Readers can also see in Fig.14, Bitcoin price has a significant daily drop in March 202, which is about -0.37. Overall, all the data is properly shown in a clean format. Readers can easily identify all the objectives.

## V. CONCLUSION

### A. Discussion

- Using bar chart to visualize time series related data should use both scaleBand and timeScale for x-axis. If only using bandScale or timeScale, the bar would not be able to be drawn in place.
- Using spectral instead of red blue for displaying diverging color is better when clarifying larger range of value.
- When the data range is large and each value point is so different from the others, or we could say the data set uncertainty is high, using interactive method letting users too zoom in and out the check can resolve the issue.

### B. Future Work

To extent this research to a higher level, here are some approaches that might be doable.

- Add timeline of historical events. This is not limited to housing related news, but could be some big event. Some readers might have a change to identify some potential relationships with the news and housing market change.
- A bar chart race of housing price change over time among cities. This could save some readers time in investigating data piece by piece. Readers can have a general idea of the bay area housing market more complete.
- Introduce discussion section in the website. Readers can leave comment and share their thoughts. This approach could help improving the website quality.

## REFERENCES

- [1] Amanda Bartlett, SFGATE. *Dozens camp overnight for chance to buy \$1.2 million Bay Area townhouses* Apr. 25, 2021. Available: <https://www.sfgate.com/realestate/article/2021-04-Santa-Clara-camping-overnight-townhome-16127719.php>
- [2] GEORGE AVALOS, EAST BAY TIMES. *Exodus: Bay Area migration accelerated in recent months* Mar. 4, 2021. Available: <https://www.eastbaytimes.com/2021/03/04/covid-economy-bay-area-residents-exit-region-growing-numbers-jobs-tech/>
- [3] Zillow, *Housing Data, "HOME VALUES"*, seen in May. 16, 2021. Available: <https://www.zillow.com/research/data/>
- [4] Realtor, *Real Estate Data, "INVENTORY Monthly"*, seen in May. 16, 2021. Available: <https://www.realtor.com/research/data/>
- [5] Kaggle, *Cryptocurrency Historical Prices, "coin\_Bitcoin.csv"*, seen in May. 16, 2021. Available: <https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory>
- [6] Niche, *2021 Best Suburbs to Live in the San Francisco Bay Area*, seen in May. 16, 2021. Available: <https://www.niche.com/places-to-live/search/best-suburbs/m/san-francisco-metro-area/>
- [7] San Francisco Planning Department. *San Francisco Housing Needs and Trends Report*, 2018. Available: [https://default.sfplanning.org/publications\\_reports/Housing-Needs-and-Trends-Report-2018.pdf](https://default.sfplanning.org/publications_reports/Housing-Needs-and-Trends-Report-2018.pdf)
- [8] Redfin Real Estate News, *Data Center*, seen in May 14, 2021. Available: <https://www.redfin.com/news/data-center/>
- [9] Jesse Howe, *D3: Diverging Bar Chart (Horizontal) - Quantize Scale*, seen in May 14, 2021. Available: <https://codepen.io/jchowen/pen/WLNXqB>
- [10] Phillipp Koytek, *Bar Chart with Time Scale*, published in Aug. 13, 2020. Available: <https://observablehq.com/@phillippkoytek/bar-chart-with-time-scale>

- [11] Susie Lu, *d3 SVG Legend (v4)*, seen in May 14, 2021. Available: <https://d3-legend.susielu.com/>
- [12] Susie Lu, *d3-annotation*, seen in May 14, 2021. Available: <https://d3-annotation.susielu.com/>
- [13] D3.js Graph Gallery, *Zoom with Axis*, seen in May 17, 2021. Available: [https://www.d3-graph-gallery.com/graph/interactivity\\_zoom.html#axisZoom](https://www.d3-graph-gallery.com/graph/interactivity_zoom.html#axisZoom)
- [14] Patrick May, The Mercury News, *Bay Area homebuyers can't catch a break – and they unleash on Twitter*, published in May 25, 2018. Available: <https://www.mercurynews.com/2018/05/25/bay-area-homebuyers-cant-catch-a-break-and-they-unleash-on-twitter/>
- [15] Wolf Richter, Wolf Street, *Housing Bubble 2 in San Francisco Bay Area & Silicon Valley Pops Despite Startup Millionaires & Low Mortgage Rates*, published in Aug. 15, 2019. Available: <https://wolfstreet.com/2019/08/15/housing-bubble-2-in-san-francisco-bay-area-silicon-valley-is-cooked/>
- [16] Wolf Richter, Wolf Street, *“Pent-up Supply” Floods San Francisco Housing Market, Most Since Housing Bust*, published in Jul 17, 2020. Available: <https://wolfstreet.com/2020/07/17/pent-up-supply-floods-san-francisco-housing-market-most-since-housing-bust/>