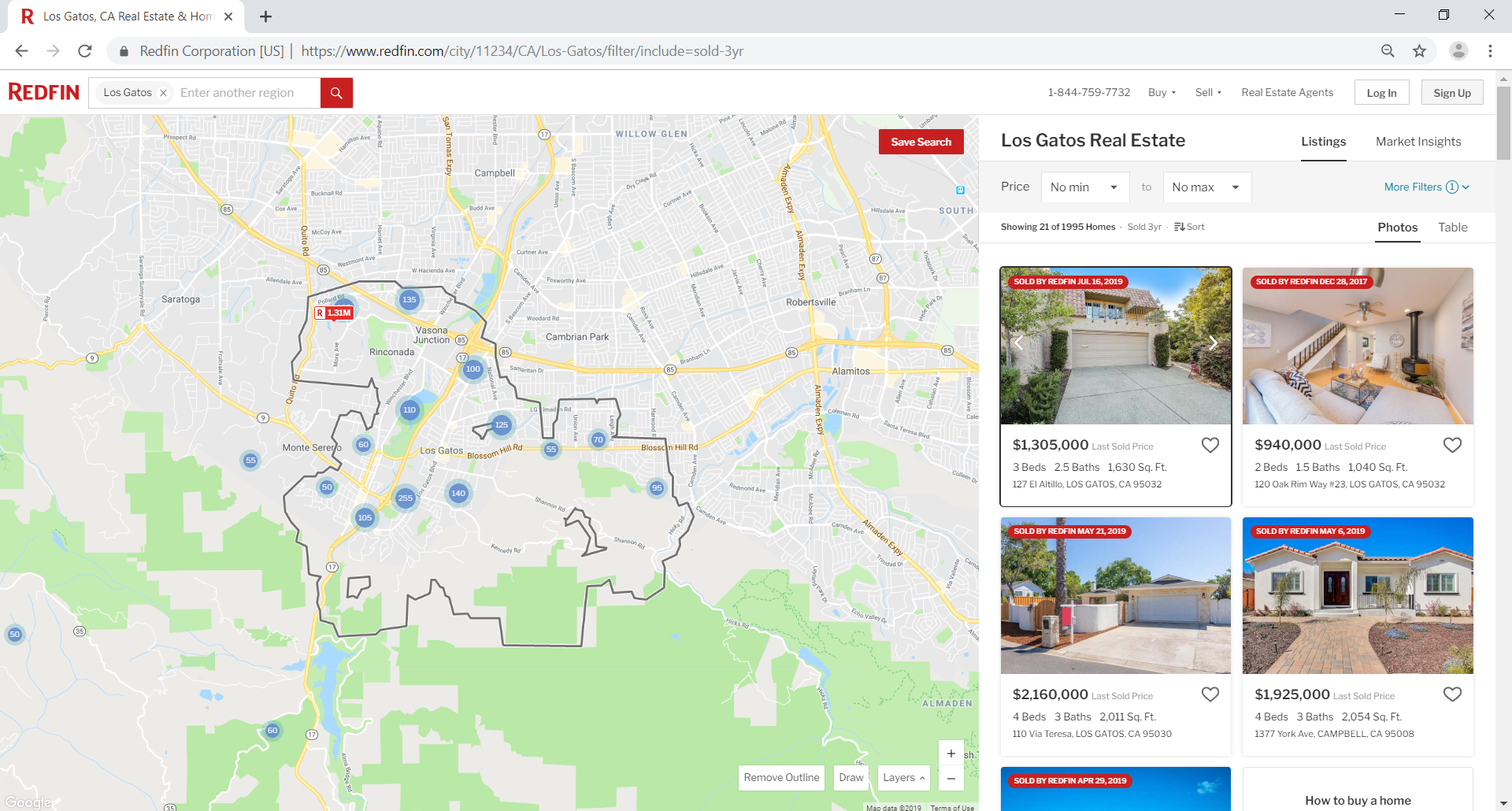
Silicon Valley’s overheated housing market has been a nightmare for incoming residents as well as the current residents. Tech companies have continued hiring, leading to the influx of people in the Bay area. This has led to increased demand and inventory has not kept up pushing the prices to unaffordable levels. In the last 5 years, the market has seen heavy bidding from $50k to $200k over the asking price. The main contributors to this competitive landscape are the high-tech salaries, tech IPOs and overseas investment. Home buyers consider affordability, commute time and good school districts as important vectors in their decision-making process.

Having lived in the Bay area for more than 7 years, I have personally experienced the upward market trajectory. I have been a resident of one of the mediocre bay area cities and have a keen desire to move to a better school district. With this goal in mind, I have closely monitored the real estate prices over the last 2 years in one of the Silicon Valley cities, Los Gatos. Los Gatos is a beautiful little town on the foothills of the Santa Cruz Mountains with a population of 30,680 as per U.S Census Bureau. Los Gatos is also home to the popular streaming content creator and provider, Netflix. The Los Gatos school district is very good and low inventory has led to skyrocketing prices in the past few years. The inventory is more compressed than other Bay Area cities due to established neighborhoods and very few new homes being built.

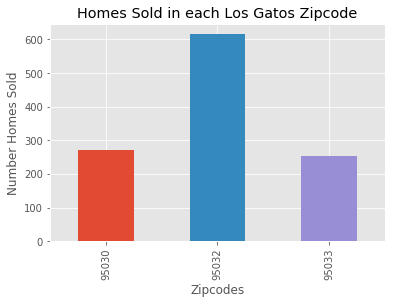
In my observations, from Spring of 2018, the market dynamics compelled buyers to make offers within 24 - 48 hours after the public listing of their desired home. Desperate home buyers were forced to bid way above listing price. Depending on their level of interest buyers would bid 50k to 200k above the asking price. This led to a huge real estate price disruption pricing a large number of buyers out of the. By Spring of 2019, the market had slowed down and buyers were not making hasty offers. Not many buyers were interested in offering more than 20k above the asking price. The time to sell homes increased from less than a week to over 30 days. In order to sell their homes, some sellers offered incentives to the buyers, some reduced the prices while others extended offers such as paying the HOA dues for a year.

The above observations and my personal desire to buy a home inspired me to consider this topic and pursue a web scraping project in analyzing a sample set of Silicon Valley’s housing market specifically for the last 2 years. Initially, I decided to scrap the real estate website Zillow. But Zillow only showed data of sold homes for approximately last one year. Realtor.com also had the same time restriction. Since I was looking for a larger dataset to compare the sold home listings for at least 2 years, I went with Redfin. The sample set considered is the listings of sold homes in the city of Los Gatos from July 2017 to July 2019. Records of sold homes were scraped from [redfin.com](https://www.redfin.com/city/11234/CA/Los-Gatos/filter/include=sold-3yr) by using Scrapy. Other tools used to clean and visualize data were Pandas, NumPy, Matplotlib and Seaborn. Around 1500 sold home records with 8 variables including street name, city, zip code, bedrooms, bathrooms and square feet, sold price and sold date were analyzed. A screenshot of the Redfin website is attached below.



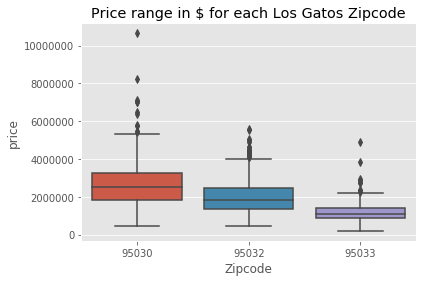
On scraping, my Scrapy spider was detected every single time and my IP address was blocked from scraping the website. To get around this, I tried different ways to slow my requests. I set download delays to 10 and CONCURRENT\_REQUESTS to 1 but was able to scrap a maximum of 40 records. My spider worked and was giving me the desired data but had limited records. For the analysis part of this project, I decided to download the data with the Redfin API. The data set had 1282 homes sold in the last two years. I cleaned the data by filtering out unwanted zip codes, missing values, duplicate columns, renamed the columns with relevant names.

**So, which is the most popular Los Gatos zip code among buyers?**



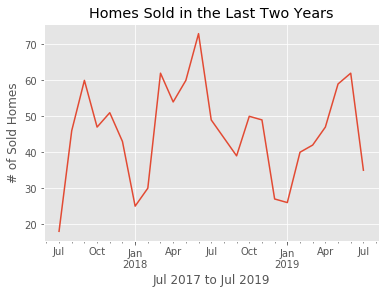
As seen in the graph above, the most homes were sold in 95032 zip code. This could be due to multiple factors such as relative affordability (as seen in the next chart) and relatively higher inventory compared to the other zip codes. 95033 is the least preferred because it lies further away from Los Gatos downtown and Silicon Valley. 95033 homes are mostly in the Los Gatos mountains with longer commute times due to windy and narrow roads.

**Which zip code is the most expensive?**



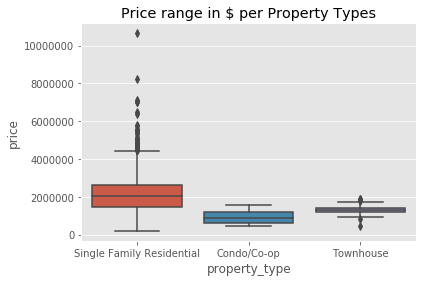
In spite of 95032 being the most popular among the three zip codes, the median price in that zip code is not the highest. As we see, 95030 has the highest median price. This particular zip code has a higher fraction of the luxury homes of Los Gatos. It’s close proximity to Los Gatos downtown also leads to higher prices.

**What has been the seasonal buying habits of buyers in this area?**



As per conventional wisdom, the real estate market picks up in Spring and slows down in winter. The above graph shows this behavior. Families in particular, start to shop for homes in March so that they are able to close before the start of summer. This basically helps them to move to a new home before the start of a new school year. Also, the above graph shows the lower demand for homes during fall and winter seasons.

**What kind of home falls in your price range?**

****

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Single Family Home** | **Townhome** | **Condo** |
| **Price Range** | $196,875 - $10,650,000 | $449,000 – $1,900,000 | $477,725 - $1,580,000 |
| **Median Price** | $2,040,000 | $1,300,000 | $875,000 |

**Next Steps:**

For my ongoing future analysis, I would like to research on

* the correlation between sold home prices and household income.
* Web scrape with the tool Selenium to collect more variables like the listed price that will help understand the price hike or price cut on the property. These are not available with the API.

This web scraping project along with use of Pandas and NumPy, gives a fair insight into the real estate market in the city of Los Gatos. All conclusions were drawn on the basis of data collected from Redfin. Hope this analysis helps you whether you are a potential buyer or a web scraper data scientist 😊. Thank you for stopping by.

**Citations:**

<https://www.redfin.com/city/11234/CA/Los-Gatos/filter/include=sold-3yr>

<https://www.cnbc.com/2019/07/01/how-much-a-san-francisco-house-bought-10-years-ago-could-be-worth-now.html>

<https://www.financialsamurai.com/how-new-tech-ipos-could-actually-accelerate-the-decline-in-sf-bay-area-real-estate-prices/>

<https://www.census.gov/quickfacts/losgatostowncalifornia>

<https://liveinlosgatosblog.com/los-gatos-zip-code-maps/>