HOUSING VALUE & ACADEMIC PERFORMANCE IN LOS ANGELES

04/15/2021

Mai La, Sitao Chen

Overview

While the old real estate adage is "location, location, location", perhaps the new motto should be "school, school, school", as school district ranking is one of the most important factors that many people with children consider when purchasing a house. Our analysis will provide insights of how housing value correlates to school academic performance in Los Angeles Metropolitan (LA). For academic performance evaluation, we base on how each school performs on the Standard State tests, high school graduation rate and going to college rate.

Analysis Questions

1. How to evaluate school academic performance?
2. Is there any correlation between housing value and academic performance in LA?
3. For the house value changes over the last 10 years, do areas with better school performance also have higher improvement in their housing value?
4. What are the top ten zip codes having the best school performance for some of the most common house value ranges in LA?

Dataset

To evaluate school performance and housing value, we used 5 different data sources as follows:

1. Single house value:
   - Data Source: Zillow
     Data downloaded: csv file, ZHVI Single-Family Homes Time Series, Geography - Zip Code
   - Description: The dataset includes monthly averaging value for single-family homes for each zip code in the US from Jan-1996 until Feb-2021.
   - Raw data size: 30,311 entries x 311 columns
   - Variables used: zip code, state, metro, monthly house value ("1996-01-31" until "2020-01-31").
2. School location:
   - Data Source: California Assessment of Student Performance and Progress
     Data downloaded: txt file, 2018-19 Entities List
   - Description: California (CA) school information and addresses, 2019
   - Raw data size: 2,992 entries x 10 columns
   - Variables used: school code, school name, district code, district name, zip code, county name.
3. California standard test score:
   - Data Source: California Assessment of Student Performance and Progress
     Data downloaded: txt file, Countywide/Districtwide files - Los Angeles & Orange County
   - Description: detailed test scores for each subject, each grade, each school in CA, 2019
   - Raw data size: 3,576,491 entries x 32 columns
   - Variables used: school code, county code, subgroup ID, school code, grade, number of students tested, test ID, test score.

4. Graduation rate:
   - Data Source: Department of Education
     Data downloaded: csv file, Year of Data 2019-20
   - Description: graduation rate for all CA high schools
   - Raw data size: 241,058 entries x 34 columns
   - Variables used: school code, school name, number of cohort students, number of high school diploma completers, graduation rate, reporting category, aggregate level, county name.
5. College going rate:
   - Data Source: Department of Education
     Data downloaded: csv file, year 2017-2018
   - Description: going to college rate for LA high schools
   - Raw data size: 452 entries x 11 columns
   - Variables used: school name, number of students graduate, number of students enrolled in colleges.

Data Exploratory

Assumptions

1. House value:
   - Using single-family home value to represent house pricing for each zip code, the report does not include other home types (like condo or co-op, town house)
2. School:
   - Using standard public-school data for the analysis and excluding other non-standard public schools. The school types below are not considered as standard public schools:
     - Continuation Schools
     - Alternative Schools
     - Home & Hospital Instruction
     - Juvenile Court Schools
     - Opportunity Education
     - Community Day Schools
3. Academic Performance:
   - To indicate how well a school performs, we design an academic performance index (API). The API is calculated based on how well a school performs in the standard state tests (the test subjects are Math and English, Literacy & Language Art), the high school graduation rate and the going to college rate. We do not include other factors to calculate the API, like how students experience school, how well the schools serve students from different racial, ethnic, socioeconomic backgrounds, etc.

Data Cleaning & Sanity checks

The datasets we have downloaded are from multiple resources. Each of these files has its own naming convention, and the data format and data type are not consistent across all files. So, our first step is to review and understand the information from each data file, reformat the columns, and take out irrelevant data; then standardize the data format and data type; identify the common key columns; finally merge subset of the data from each file into one consolidated dataset. Within a single dataset, it is much easier for us to explore the relationship between variables that we are interested in.

1. House value:
   - Filter applied: “State” = “CA”, “Metro” = “Los Angeles-Long Beach-Anaheim”

2. School location:
   - Filter applied: “County Name” = “Los Angeles”, “Orange”
   - Remove entries with blank school names and non-standard public schools from the dataset. Excluding any school having these special words in their names: “Continuation”, “Alternative”, “Home”, “Juvenile”, “Opportunity”, “Community Day”.
3. Standard test score:
   - Filter applied:
     - “County Code” = “19”, “30” (stands for Los Angeles & Orange)
     - “Subgroup ID” = 1 (stands for all students, we do not include other groups for genders & ethnicity, etc.)
     - “School Code” ≠ 0 (exclude all records that are not schools)
   - Remove all entries with no value or “*” for the test score.
4. Graduation rate:
   - Filter applied:
     - “countyname” = “Los Angeles”, “Orange”
     - “reportingcategory” = “TA” (stands for all students & not include other demographic groups)
     - “aggreatelevel” = “S” (stands for schools)
     - “schoolcode” ≠ 0 (exclude all records that are not schools)
   - Remove all entries with “*” value for the graduation rate.

Data Analysis

Calculations

1. House value:
   - Calculate each year averaging house value from the monthly values for all years.
2. Academic performance:
   - Define API score using base 10:
     - From 1 – 3: below average
     - From 4 – 5: nearly met
     - From 6 – 8: standard met
     - From 9 – 19: standard exceeded
   - Calculate the API for each school type: elementary school, middle school and high school, based on each test subject and grade level, graduation rate and college going rate. We use different weights as the following table to get the total API score for each school. For the standard test score, we refer to the criteria listed by the CA Department of Education for ranging which test score means failed, pass, and exceed standard. Whereas, for the graduation rate and college going rate, we decide the API range based on the distribution of these data (i.e. by the values at min, 50%, 75%, max).
   - For example: API for a high school is calculated by

API = (API_English_test_grade11 x 35% + API_Math_test_grade11 x 35%
+ API_Graduation_rate x 15% + API_College_going_rate x 15%)

API score for each group	Elementary rating						Middle School rating						High School rating			
	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 11		Graduation	College
	English	Math	English	Math	English	Math	English	Math	English	Math	English	Math	English	Math	Rate	Going
1	2114	2189	2131	2204	2201	2219	2210	2235	2258	2250	2288	2265	2299	2280	0	0
2	2198	2253	2226	2273	2281	2298	2292	2314	2332	2328	2354	2345	2364	2368	40	20
3	2283	2317	2321	2342	2362	2376	2375	2394	2405	2406	2421	2424	2428	2455	70	35
4	2367	2381	2416	2411	2442	2455	2457	2473	2479	2484	2487	2504	2493	2543	83	50
5	2400	2409	2445	2448	2472	2492	2494	2513	2516	2526	2527	2545	2538	2586	93	60
6	2432	2436	2473	2485	2502	2528	2531	2552	2552	2567	2567	2586	2583	2628	94	70
7	2447	2452	2488	2501	2522	2541	2553	2567	2576	2584	2592	2603	2608	2651	95	75
8	2461	2469	2503	2517	2542	2554	2575	2581	2601	2601	2618	2620	2633	2673	96	80
9	2476	2485	2518	2533	2562	2566	2596	2596	2625	2618	2643	2636	2657	2696	98	85
10	2490	2501	2533	2549	2582	2579	2618	2610	2649	2635	2668	2653	2682	2718	99	90
MAX	2623	2621	2663	2659	2701	2700	2724	2748	2745	2778	2769	2802	2795	2862	100	100
API Weights	17%	17%	17%	17%	17%	17%	17%	17%	17%	17%	17%	17%	35%	35%	15%	15%

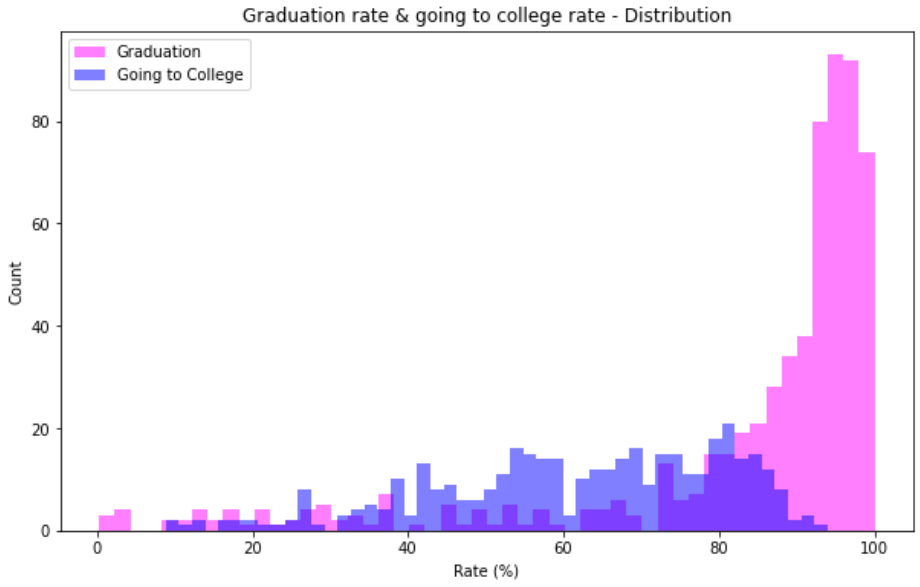
- When grouping the schools together by each zip code the total API of each zip code is the average of elementary school API, middle school API and high school API. We use a similar approach when grouping the schools together by school district.

  API = (API_elementary + API_middle + API_high) / 3

- The averaging graduation rate when grouping the data together by zip code (or by school code) is calculated by the total number of students graduated divided by the total number of grade 12 cohort students in that zip code (or school code). Similar approach is used for calculating the going to college rate. We don't use the mean values of all schools, due to the fact that different schools have different numbers of students, and the result will not be correct without considering the weights of each school population.

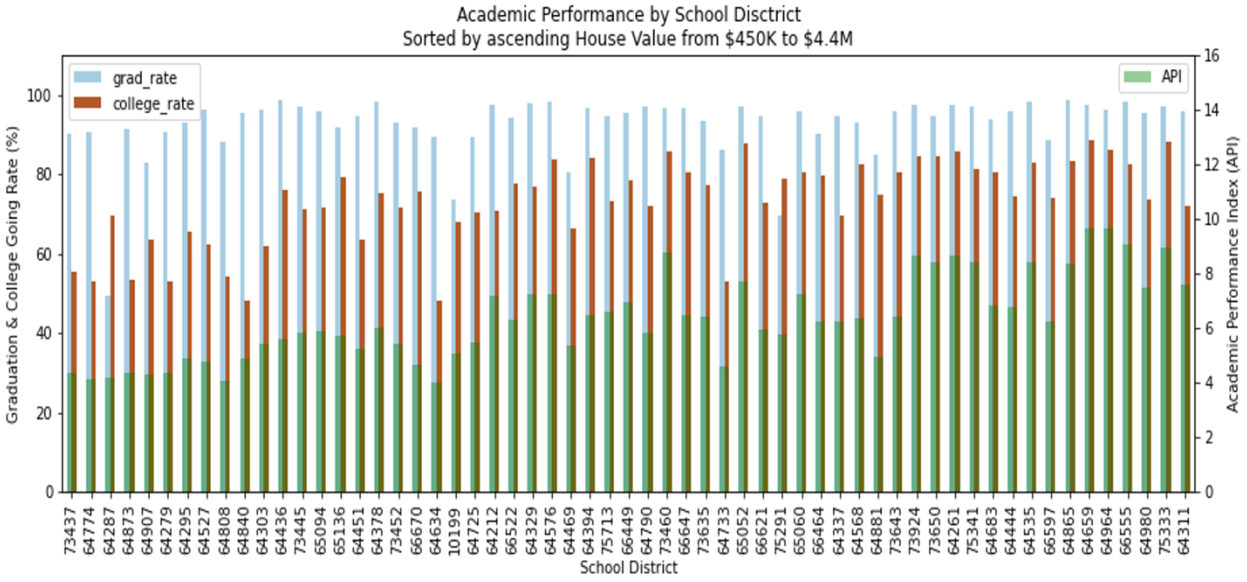
What is the distribution of graduation rate and college going rate?

The average graduation rate for all LA high schools is 93%, most schools have high graduation rate >90%. Whereas, the going to college rate is ~ 65% on average and the distribution is spread across a wider range (mostly from 40% - 90%). Based on this information, we cut the API score for these data accordingly as the table above.



How does each school district perform?

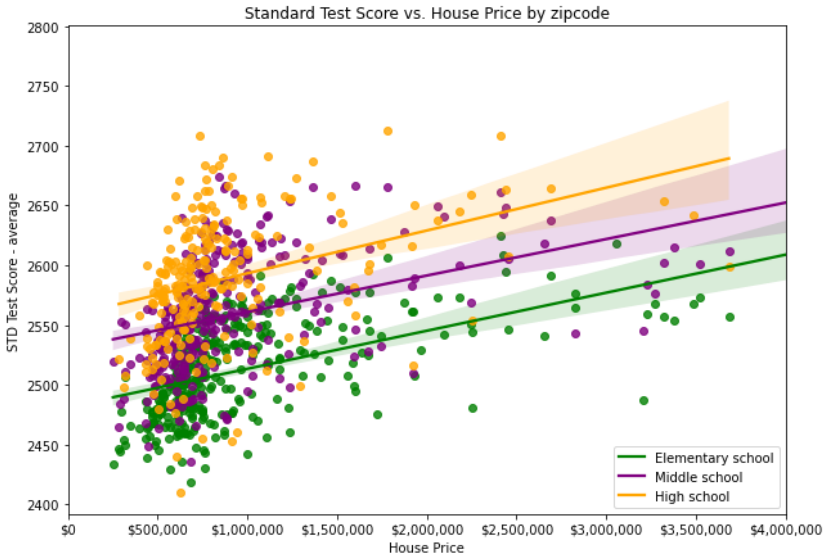
By grouping the data by school district and sort the house values by pricing increased gradually (from minimum $450,000 to $4,500,000), we can see from the graph below that the high school graduation rates are pretty much the same across all school districts regardless of the house value; however, the going to college rates are increased from low house value school districts to high house values school districts. So does the API.



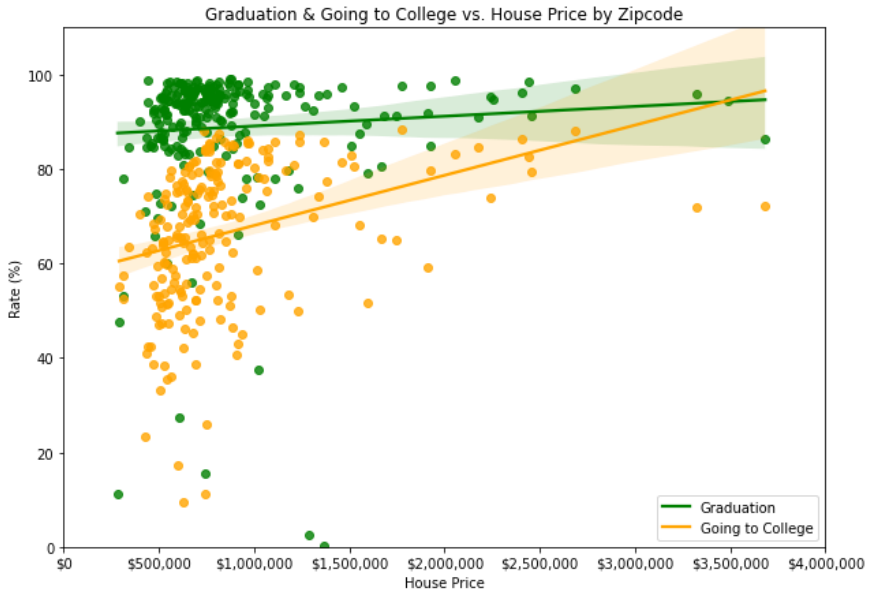
The data points to a fact— schools that have high Academic Performance Index, so do home values. As that being said, when it's time to sell and upgrade your home, doing so inside a decent school district will likely reap dividends.

Is there any trend between house value and school performance in each zip code?

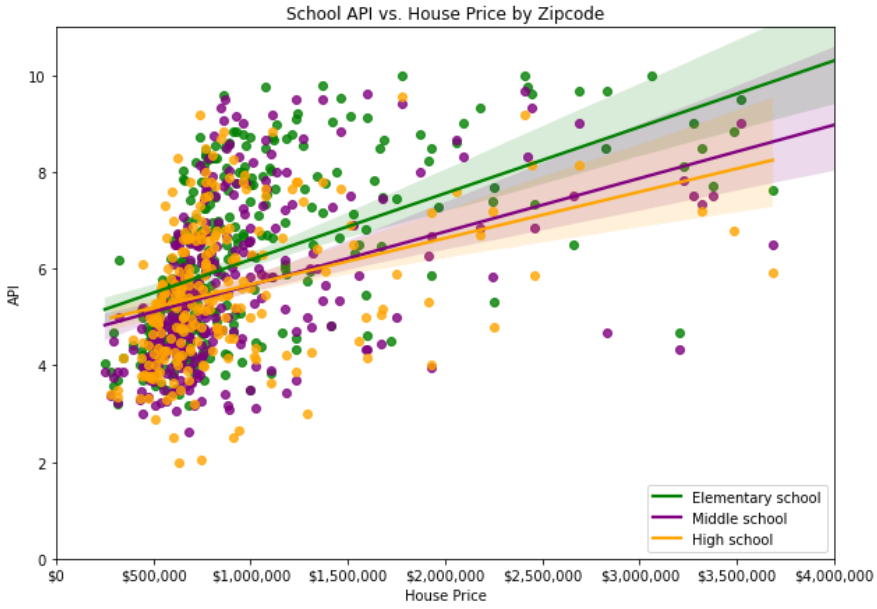
We group the data by zip code to explore the trends between the house values with the standard test scores among all grades level, and the result shows that as house price increases, the average test scores also increase as the graph below shows. Each data point represents the average house price and test score for each zip code. We could also see that most house values are in the range of $450,000 to $900,000.



The below graph shows the relationship between the high school graduation rate (in green) and college going rate (in orange) with house price grouped by zip code. It shows the graduation rate does not change that much by house price. Whereas, going to college rate shows areas with higher house price is likely to have higher going to college rate.



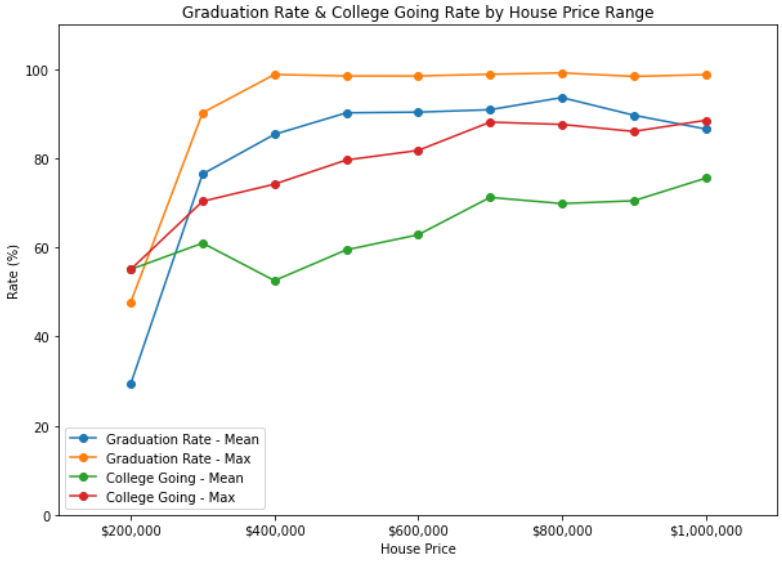
Combining the standard test score with the graduation rate and college going rate to calculate the school API for each zip code by grade level, we also see an increasing trend between house price and school API as the following chart. And most schools have the API scores in the range of 3 to 7.



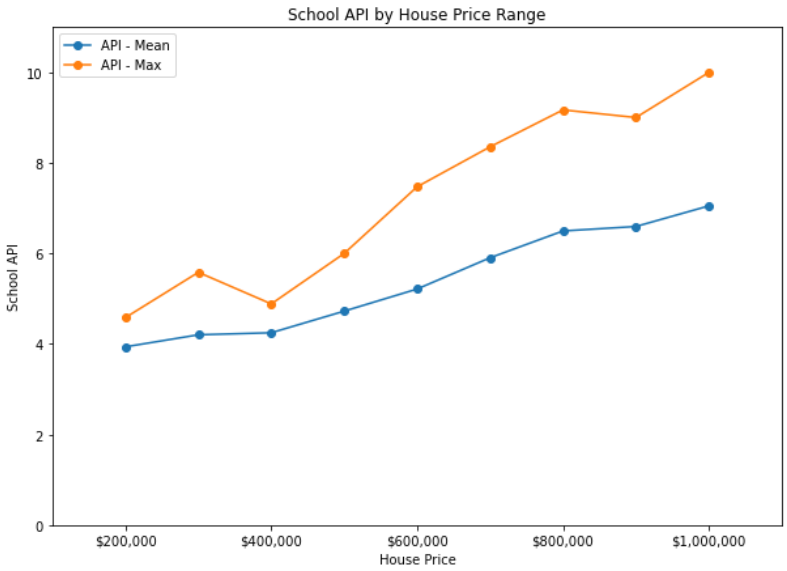
What is the school performance for different house value ranges?

Further grouping the data by each house pricing ranges within $100,000 for each group (i.e., the $500,000 group has houses with value from $500,000 until $599,999). For the house that is $1,000,000 or above we group them all together since there are not many houses in this range, and the school performance was not much different within this group as observed from the data above.

The chart below shows a correlation between going to college rate shows and the house value. Areas with higher house values have more students going to colleges after graduation (green line for the mean value of going to college rate). The graduation rate does not show much difference in the graduation rate for houses in the range of $400,000 to $1,000,000, however, we could see that houses in the $200,000 and $300,000 range have low graduation rate and college going rate (blue line for the mean value of graduation rate).

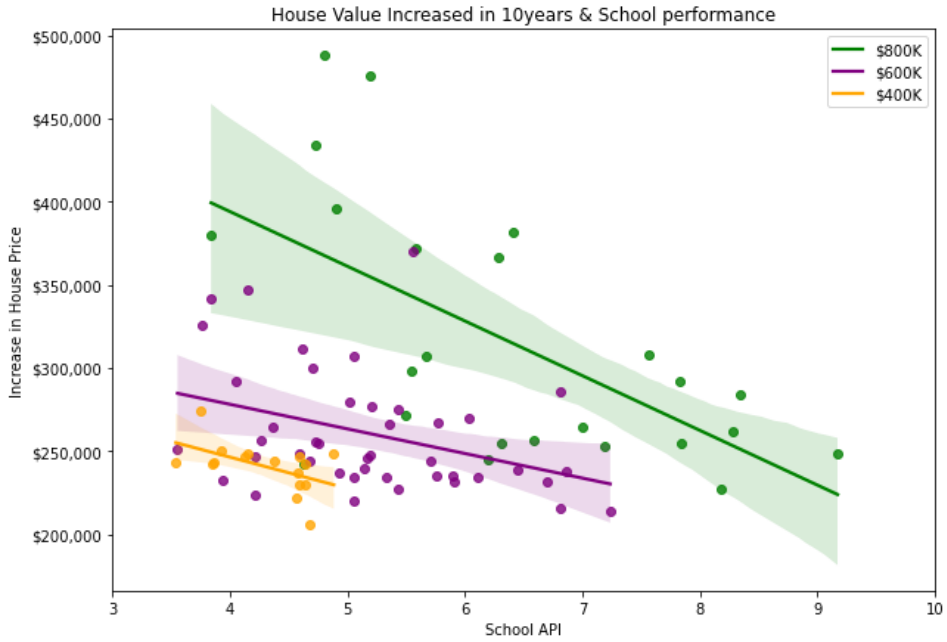


For the overall school performance API (including all schools from elementary to high school), the data show a clear trend that when the house price increases, the school API also increases as the chart below.



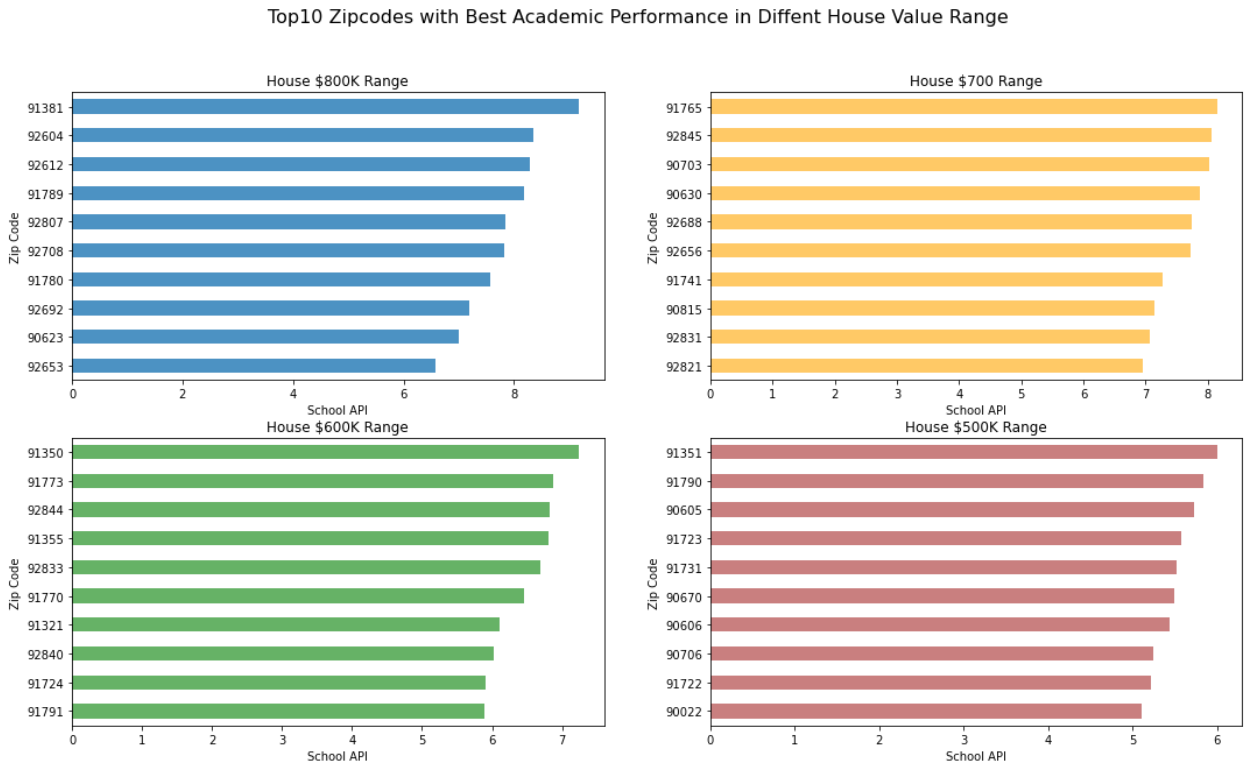
How does school performance correlate with the increase in house value over the past 10 years?

Our hypothesis was areas with good schools could be good investments for home buyers, and it might have better value increasing through years compared to areas with poor school performance. However, while looking at the trend between school performance and the changes in the house value during the past 10 years in this dataset, we see an adverse trend that is contradictory to our hypothesis. The following graph shows the amount of housing price increased from 2010 until 2020 versus the school API performance for some common price ranges in LA (i.e. houses at $400K, $600K and $800K). Taking the $800K range as an example (green dots), areas with school API 9 had ~$250K increased in their values after 10 years, however areas with school API 4 had ~$400K increased in their values after 10 years.



What are the top 10 zip codes with the best API for the common house value ranges in LA?

The following charts are the top 10 zip codes having the best school API performance for houses in the range of $500K, $600K, $700K and $800K (the most common house value ranges in LA), which was created for home buyers' references during shopping for houses.



Conclusions

From our analysis, below are the key findings:

- Most LA houses are in the range of $500K to $900K
- School performance: most schools have API from 4 – 6 (>60%), about 30% of the schools have API from 7-10.
- High graduation rate does not mean high going to college rate.
- Areas with higher house values also have higher school API & college going rates.
- The increased amount of house price over 10years does not proportional to the school API performance. (higher API – lower value increased)

Clearly, housing prices are affected by school districts, a good school district can contribute to the overall value of the real estate within the district.

Appendix

The downloaded raw datasets were not updated on GitHub, due to exceeding the size limitation. They could be accessed through our shared drive as following:
https://drive.google.com/drive/folders/1cyYQ74kCumjcq6N98Tv8FmlRvAYOURF3?usp=sharing

Link to Data Sources:

- House Value: https://www.zillow.com/research/data/
- CA standard test score and school location: https://caaspp-elpac.cde.ca.gov/caaspp/ResearchFileList?ps=true&lstTestType=B&lstTestYear=2019&lstCounty=19&lstCntyNam=Los%20Angeles&lstFocus=btnSearch
- Graduation rate: https://www.cde.ca.gov/ds/sd/sd/filesacgr.asp
- Going to College rate: https://dq.cde.ca.gov/dataquest/DQCensus/CGRLevels.aspx?cds=19&agglevel=County&year=2017-18