**Project Title: Data Analytics of real estates in California, specifically in the L.A. area.**

**Team (#4) members: Shayan Norouzi, Sicong Wang, Ly-Bach Truong, Eric Liu**

**Project Description/Outline:**

In California, if today’s houses for sale have similar features of the top 5 percent most valuable and least valuable houses according to the 1990 California Census, then they should have similar standing against other houses.

**Research Questions to Answer:**

* According to the 1990 Census, what are the houses with values that are the top 5 percent most and least valuable? What is the average of each of their features?
* What are the houses that are currently being sold on today’s market? What are the houses with similar features?
* How do houses with similar features in the 1990 compare to their present-day counterparts? Do they cost the same? (Calculate for inflation)

**Data Sets to be Used:**

* <https://www.kaggle.com/camnugent/california-housing-prices>
* www.CRMLS.org
* Possibly will use Gmaps API

**Rough Breakdown of Tasks**

* Understand the data
* Search for any inconsistencies/empty data
* Fill in or remove empty data
* Sort by median house value
* Grab the top five most and least valuable
* Search for houses that are being sold currently that have similar features and determine if they cost the same