**IBM Data Science Professional Certificate Capstone**

Ha Nguyen

**I – Business problem**

San Francisco: tech capital, housing bubble

Stakeholder: newbies to the city looking to find a house

Care about: housing price, housing price trend, nearby venues

---

**II – Data**

This assignment uses zip code and neighborhood information from healthysf.org. Specifically, I use the BeautifulSoup4 library to scrape the zip code – neighborhood table from the webpage. To retrieve the coordinates of each zip code, I use Google geocoder. To retrieve popular venues in each zip code, I use the Foursquare API.

Also: property shark for house price

**Bibliography**

Katcher, B., Reiter, R. (2004). San Francisco Neighborhoods as ZIP Codes. Retrieved from <http://www.healthysf.org/bdi/outcomes/zipmap.htm>

Poenar, R. (2017). The most expensive zip codes in San Francisco. Property Shark. Retrieved from <https://www.propertyshark.com/Real-Estate-Reports/2017/09/28/expensive-zip-codes-san-francisco/>