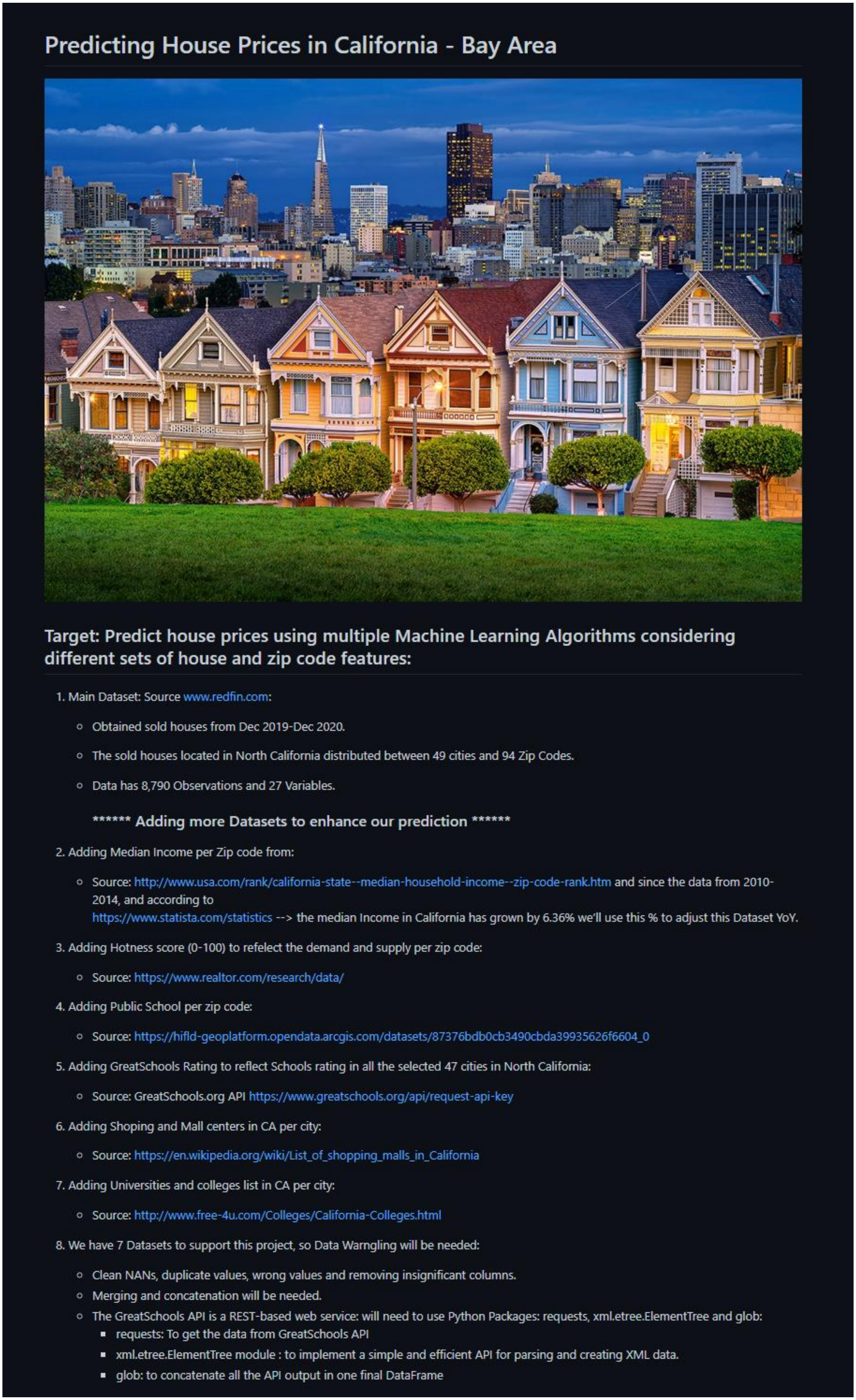

Predicting House Prices in California - Bay Area

Target: Predict house prices using multiple Machine Learning Algorithms considering different sets of house and zip code features:

1. Main Dataset: Source www.redfin.com:
   - Obtained sold houses from Dec 2019-Dec 2020.
   - The sold houses located in North California distributed between 49 cities and 94 Zip Codes.
   - Data has 8,790 Observations and 27 Variables.

   ******** Adding more Datasets to enhance our prediction ********

2. Adding Median Income per Zip code from:
   - Source: http://www.usa.com/rank/california-state--median-household-income--zip-code-rank.htm and since the data from 2010-2014, and according to
     https://www.statista.com/statistics --> the median Income in California has grown by 6.36% we'll use this % to adjust this Dataset YoY.
3. Adding Hotness score (0-100) to refelect the demand and supply per zip code:
   - Source: https://www.realtor.com/research/data/
4. Adding Public School per zip code:
   - Source: https://hifld-geoplatform.opendata.arcgis.com/datasets/87376bdb0cb3490cbda39935626f6604_0
5. Adding GreatSchools Rating to reflect Schools rating in all the selected 47 cities in North California:
   - Source: GreatSchools.org API https://www.greatschools.org/api/request-api-key
6. Adding Shoping and Mall centers in CA per city:
   - Source: https://en.wikipedia.org/wiki/List_of_shopping_malls_in_California
7. Adding Universities and colleges list in CA per city:
   - Source: http://www.free-4u.com/Colleges/California-Colleges.html
8. We have 7 Datasets to support this project, so Data Warngling will be needed:
   - Clean NANs, duplicate values, wrong values and removing insignificant columns.
   - Merging and concatenation will be needed.
   - The GreatSchools API is a REST-based web service: will need to use Python Packages: requests, xml.etree.ElementTree and glob:
     - requests: To get the data from GreatSchools API
     - xml.etree.ElementTree module : to implement a simple and efficient API for parsing and creating XML data.
     - glob: to concatenate all the API output in one final DataFrame