

Stat 420 Final Project Proposal

STAT 420, Summer 2022, D. Unger

Stat 420 Final Project Proposal

1. Team Member information

Team Member's Names:

- Kathryn DeWitt
- Shashank Thakur
- Avinash Tiwari

NetIDs:

- kdewitt3
- sthakur5
- tiwari6

2. Proposed Title:

We propose calling our project **Modeling California Housing Prices**.

3. Description of Data File

/*Avi

'California Housing Prices' dataset pertains to the houses found in a given California district and some summary stats about them based on the 1990 census data. There are 10 columns and 20.6K rows in the data. The columns are as mentioned below:

1. longitude: Longitude of the house. A measure of how far west a house is; a higher value is farther west
2. latitude: Latitude of the house. A measure of how far north a house is; a higher value is farther north
3. housing_median_age: Median age of a house within a block; a lower number is a newer building
4. total_rooms: Total number of rooms within a block
5. total_bedrooms: Total number of bedrooms within a block
6. population: Total number of people residing within a block
7. households: Total number of households, a group of people residing within a home unit, for a block
8. median_income: Median income for households within a block of houses (measured in tens of thousands of US Dollars)

- 9. median_house_value: Median house value for households within a block (measured in US Dollars)
- 10. ocean_proximity: Location of the house w.r.t ocean/sea

*/

//TODO: Description of the data file (what they contain including number of variables and number of records). You do not necessarily have to list all the variables, but at least mention those of greatest importance.

4. Background on datasets

//Shashank //TODO: Background information on the data sets, including specific citation of their source (so that I can also access it).

5. Research Question/Interest in Data Set

//Avi //Kathryn

//TODO: A brief statement of the business, science, research, or personal interest you have in the data set which you hope to explore

6. Evidence of Loading into R

//Shashank

//TODO: Evidence that the data can be loaded into R. Load the data, and print the first few values of the response variable as evidence.

```
#TODO: load in the dataset and print first few obs
```