Santa Cruz Housing

By Ayush Arora, Albert Garcia, Kyle Oda, Prajna Ta

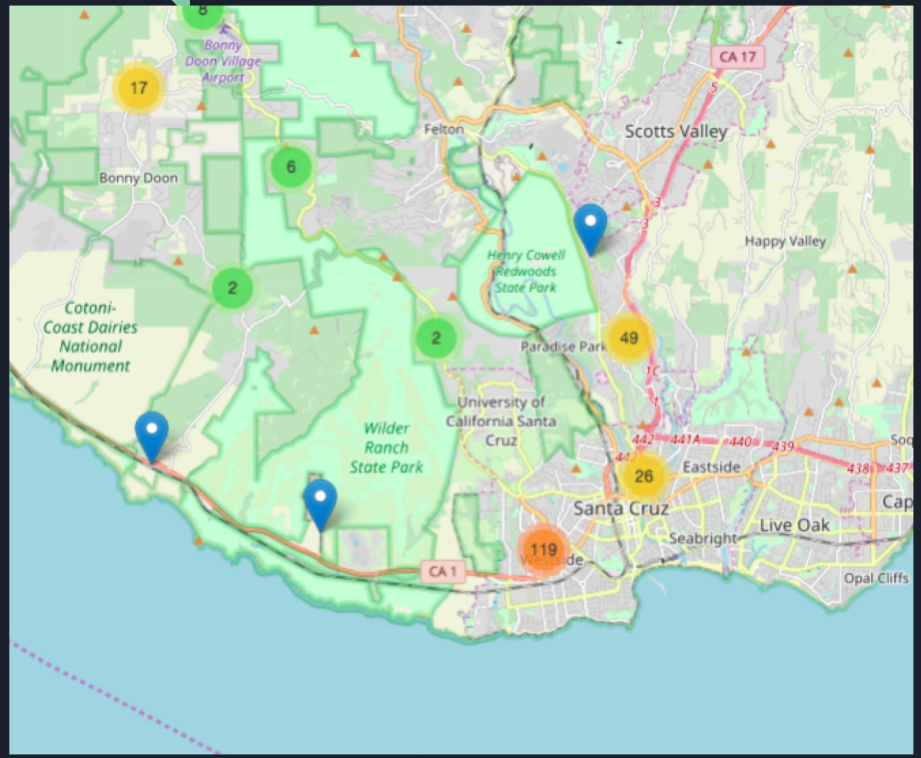
Questions and Data

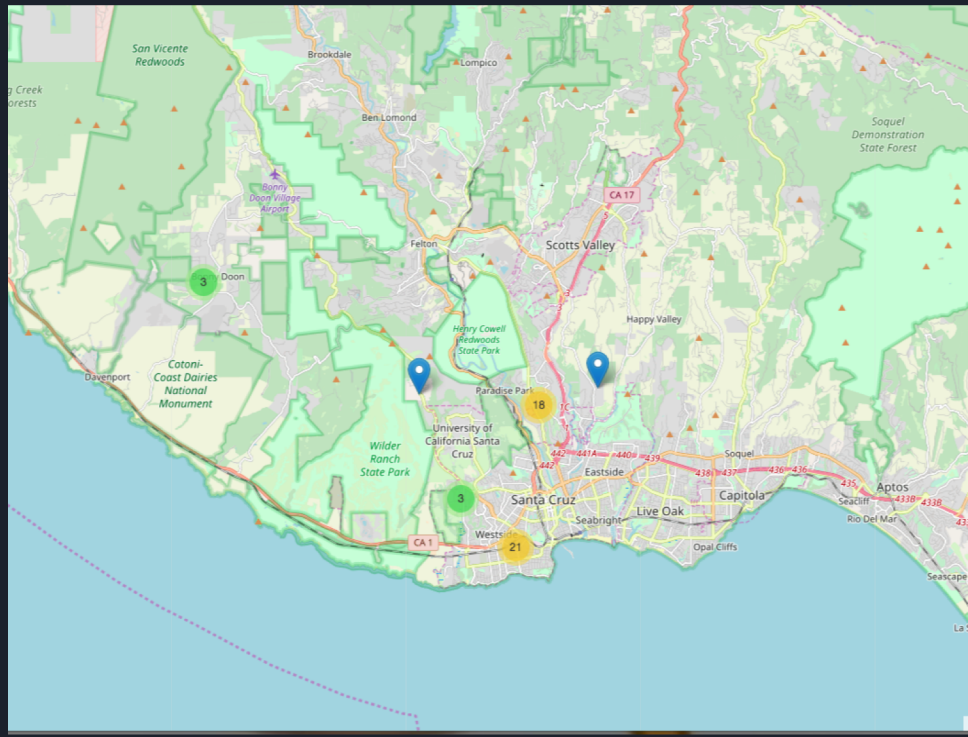
- Questions Explored
  - Is there a relationship between the location & features of a house with its price?
- Data
  - 2 regions of housing data
    - Zip Code, City, Street
    - Latitude & Longitude values
    - Beds, Baths, Sqft, Price per Sqft, Year Built, Price Estimate

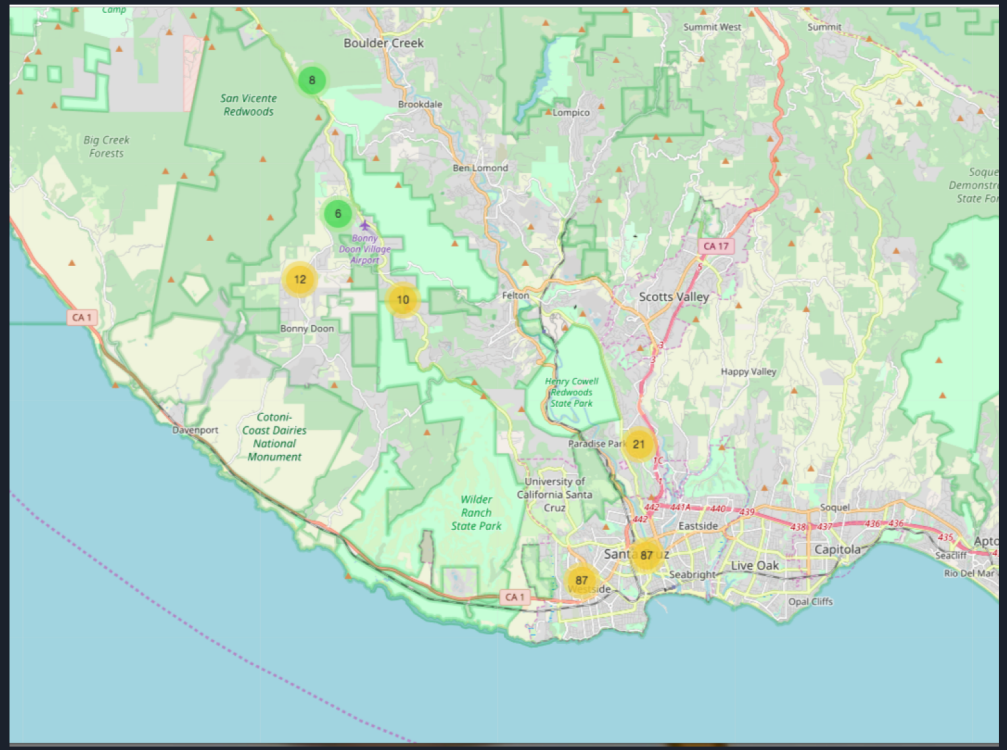
Results

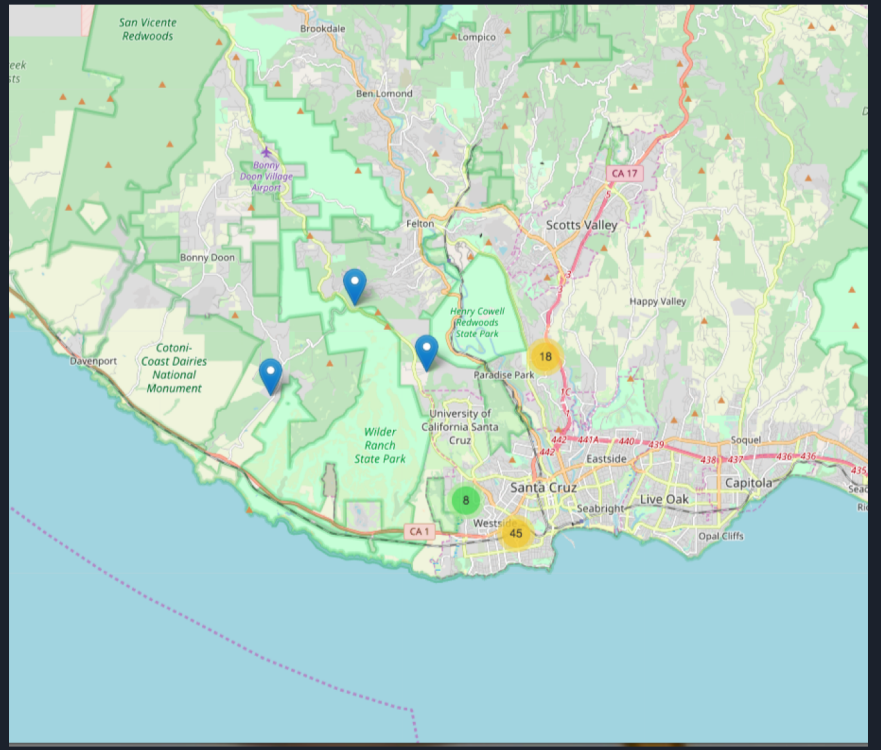
- Ran k-means cluster analysis with 6, 8, and 10 clustered neighborhoods, determined that 6 was the optimal cluster number
- Cluster Analysis created 6 different maps, each map contains houses from an individual cluster
- Due to rendering constraints, we were unable to draw concrete conclusions, but we speculate that the first cluster concentrates on downtown apartments because of similar square footage, beds/baths, and year built
- In the future, we'd like to use a different data viz package that'll be able to overlap all the clusters onto one map

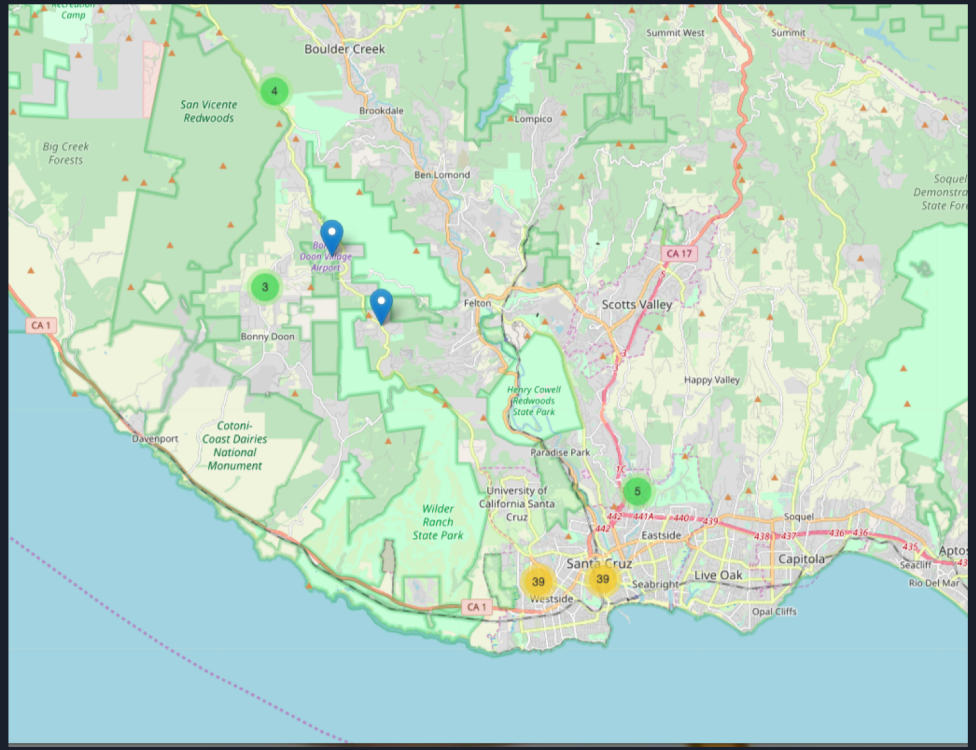
Final Results Obtained

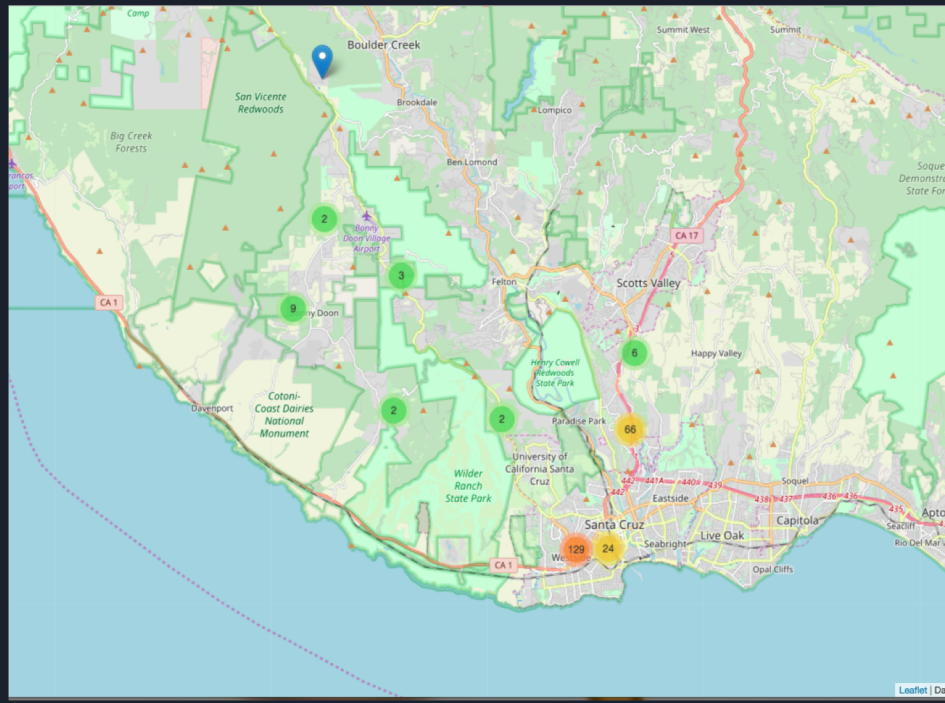





Boulder Creek
Summit West
Summit
San Vicente Redwoods
Brookdale
Lompico
Big Creek Forests
Ben Lomond
Bonny Doon Village Airport
Felton
Scotts Valley
CA 17
CA 1
Bonny Doon
Happy Valley
Henry Cowell Redwoods State Park
Davenport
Cotoni-Coast Dairies National Monument
Paradise Park
University of California Santa Cruz
Wilder Ranch State Park
Soquel
Eastside
Capitola
Live Oak
Seabright
Opal Cliffs
8
6
12
10
21
87


San Vicente Redwoods
Brookdale
Lompico
Ben Lomond
Bonny Doon Village Airport
Felton
Scotts Valley
CA 17
Bonny Doon
Happy Valley
Henry Cowell Redwoods State Park
Davenport
Cotoni-Coast Dairies National Monument
Paradise Park
18
University of California Santa Cruz
Wilder Ranch State Park
Soquel
Eastside
Santa Cruz
Capitola
Live Oak
Seabright
Opal Cliffs
CA 1
8
45


Boulder Creek
San Vicente Redwoods
Big Creek Forests
Brookdale
Lompico
Ben Lomond
Summit West
Summit
Felton
Scotts Valley
Bonny Doon
Happy Valley
Henry Cowell Redwoods State Park
Paradise Park
Davenport
Cotoni-Coast Dairies National Monument
Wilder Ranch State Park
University of California Santa Cruz
Eastside
Soquel
Santa Cruz
Seabright
Live Oak
Capitola
Seacliff
Rio Del Mar
Opal Cliffs
Westside
CA 17
CA 1
4
3
5
39
39


Boulder Creek
San Vicente Redwoods
Big Creek Forests
Brookdale
Lompico
Ben Lomond
Summit West
Summit
Felton
Scotts Valley
Happy Valley
Henry Cowell Redwoods State Park
Paradise Park
Davenport
Cotoni-Coast Dairies National Monument
Wilder Ranch State Park
University of California Santa Cruz
Eastside
Soquel
Santa Cruz
Seabright
Live Oak
Capitola
Seacliff
Rio Del Mar
Opal Cliffs
CA 17
CA 1
2
3
9
6
2
2
66
129
24
Leaflet

Libraries

- Pandas
- RegEx
- Numpy
- SKLearn
- BS4
- Folium
- Requests
- Matplotlib

Group member tasks

- Ayush
  - Assisted in pre-processing the data (parsing & cleaning) & labeled each house with a clustering number.
- Albert
  - Web Scraping and geocoded data.
- Prajna
  - Did data visualization and applied Folium to visualize the cluster maps and made changes to the csv files.
- Kyle
  - Helped create a parser to wrangle the data obtained from web scraping. Created data visualizations to better understand the dataset.