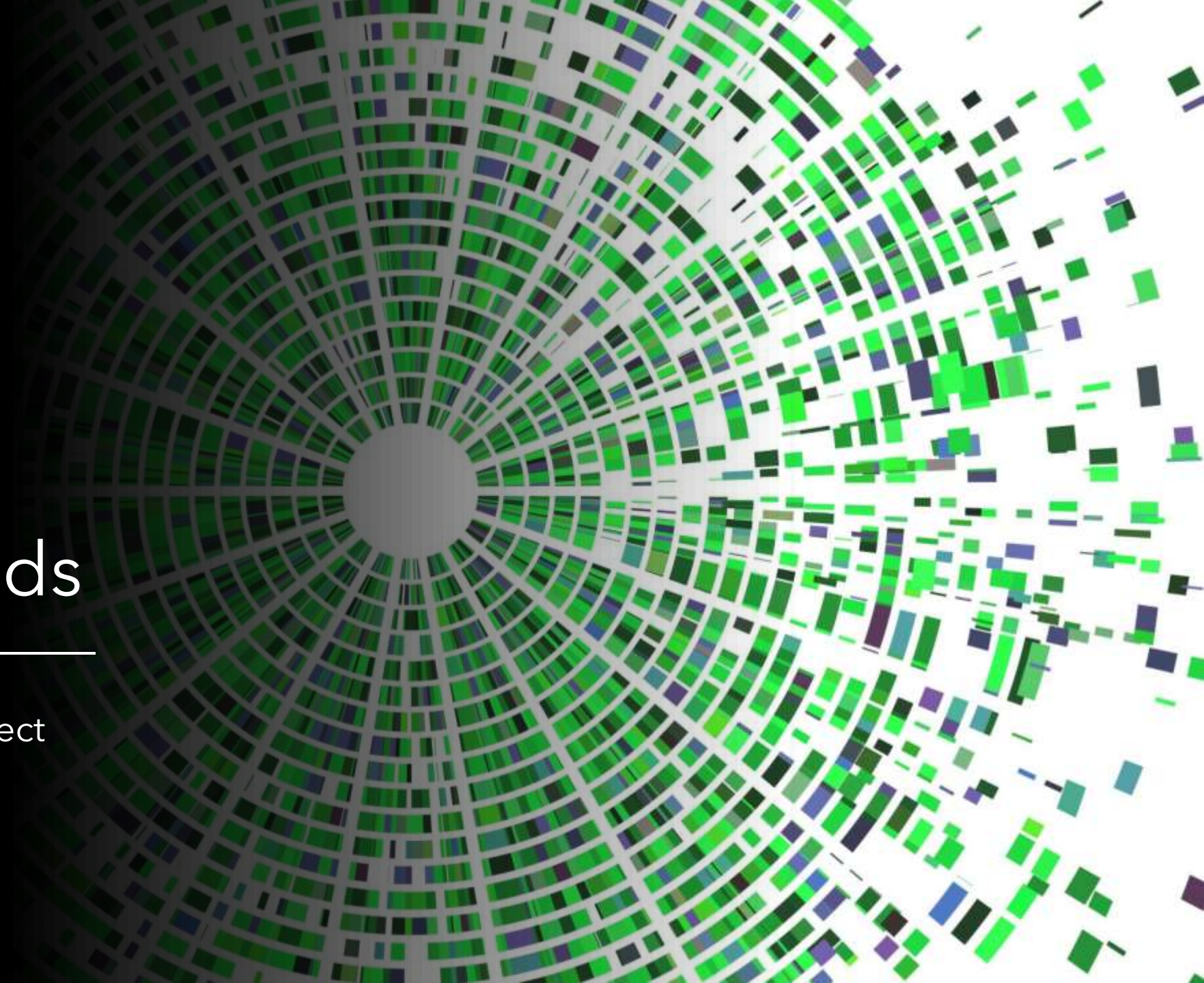




Battle of Neighborhoods

Data Science Capstone Project



Problem Statement

- In this project we will try to find a good place to **retire in California**. Ideally a place that has diverse population and lots of **parks and cafes**. Also, with **low crime rate**.
- As there are more attributes to a county that average person's mind can handle, we need to utilize the tools of data science to figure this out.

Data Requirements

In order to find optimal solution for this real-life problem we need to get data from multiple sources.

- We need **Demographic** information for each County in California
- We need **Crime Data** about each County in California
- We need to find **Popular Venues** in each County in California

Data Sources

- Demographics from US Census
<https://data.census.gov>
- Crime Data from Open Justice website.
<https://data-openjustice.doj.ca.gov>
- Popular Venues from **FourSquareAPI**
<https://developer.foursquare.com/docs/>

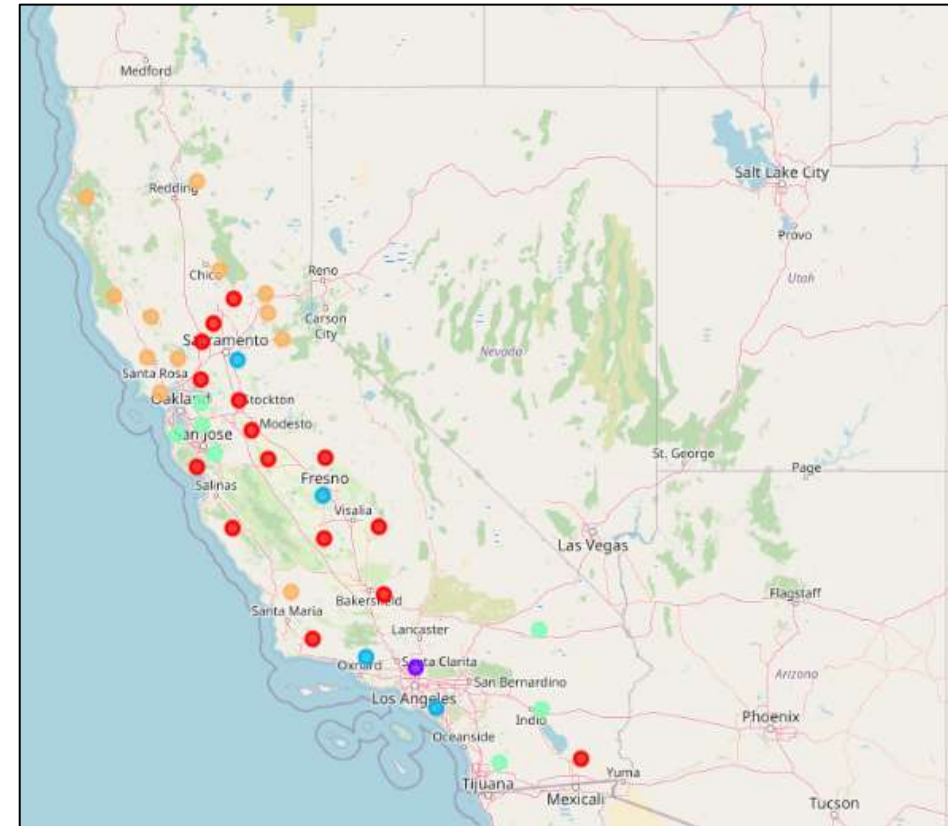


Data Wrangling

- Drop Invalid Rows and Columns
- Fix Invalid Data Elements
- Normalizing Data
- Merge Multiple Datasets

Modeling

After collecting data from all data sources, cleaning it up and merging them we used unsupervised machine learning algorithm K-Means to cluster all California Counties into five clusters.



Counties in Clusters

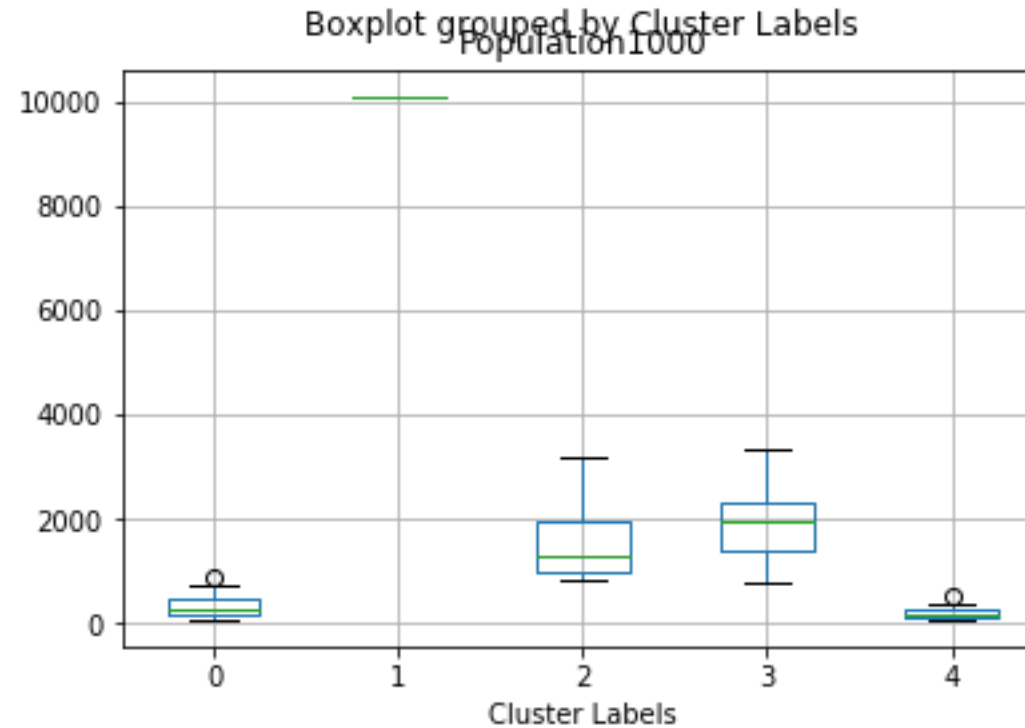
Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Kings County Monterey County Yuba County Santa Cruz County Stanislaus County Sutter County Solano County Yolo County San Joaquin County Santa Barbara County Tulare County Madera County Imperial County Kern County Merced County	Los Angeles County	Ventura County Orange County Sacramento County Fresno County	Riverside County Santa Clara County Alameda County San Bernardino County San Diego County Contra Costa County San Mateo County	Riverside County Santa Clara County Alameda County San Bernardino County San Diego County Contra Costa County San Mateo County

Understanding Results

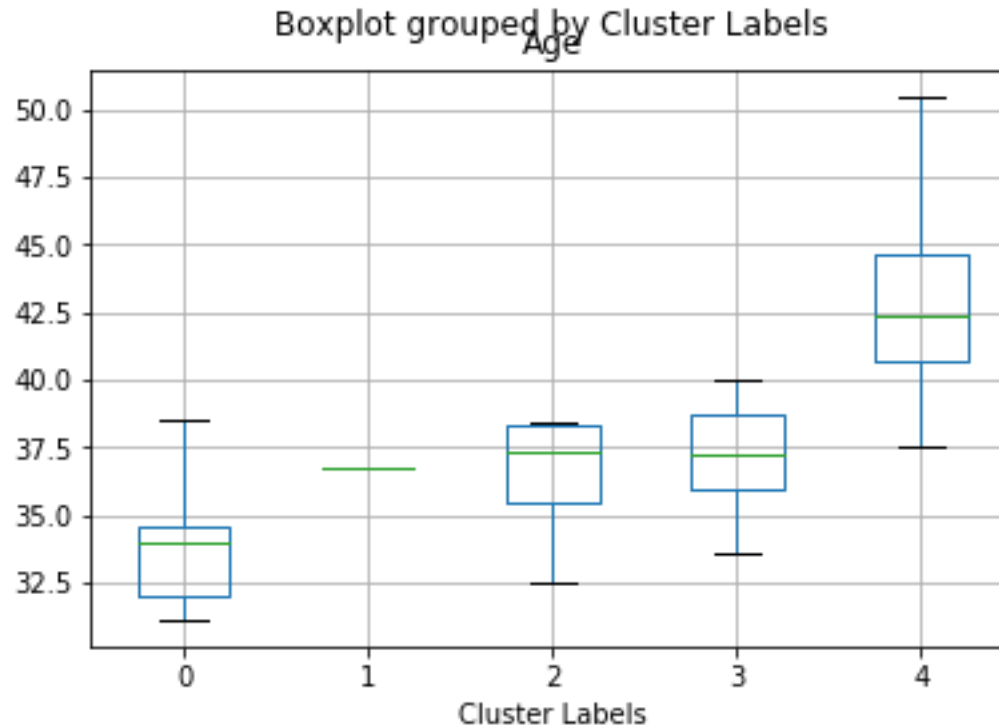


Exploring Clusters

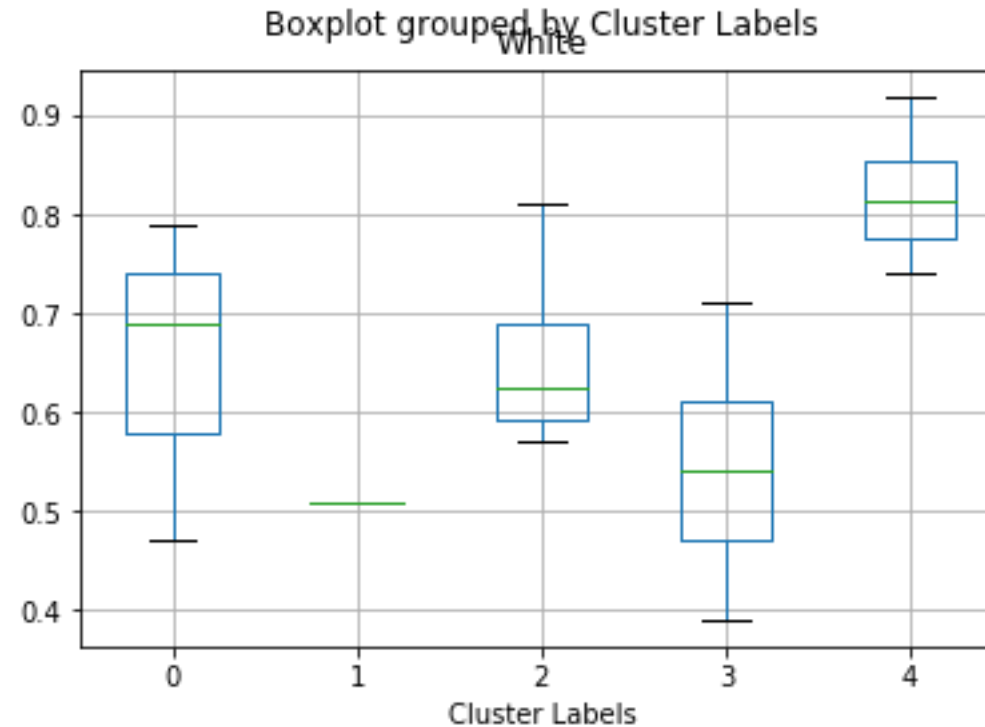
How Clusters differ in Population



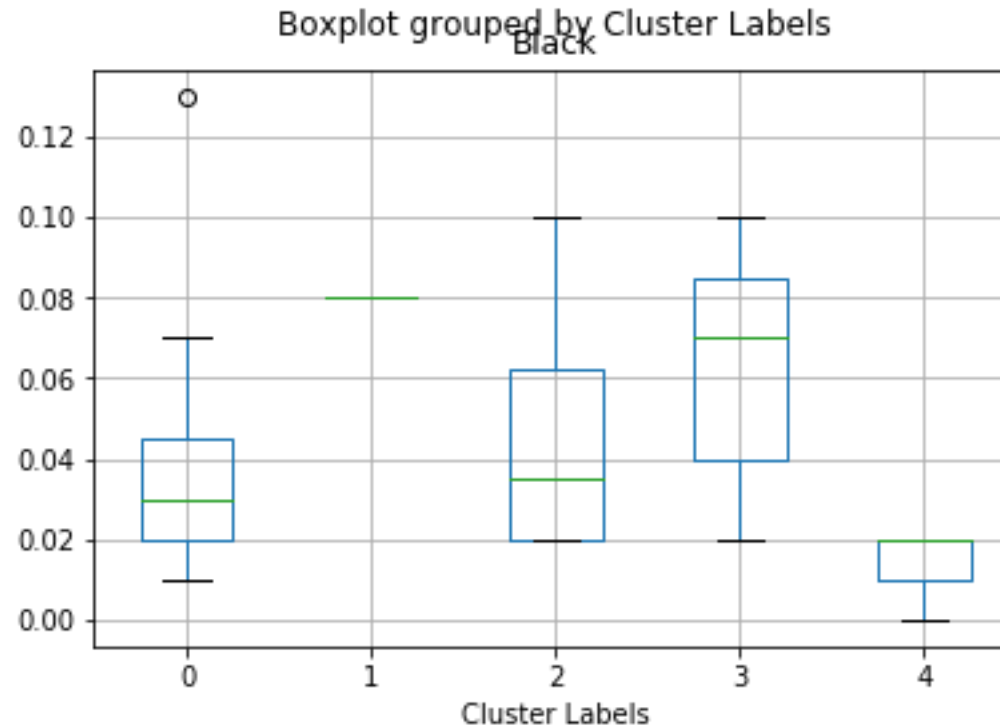
How Clusters differ in Age



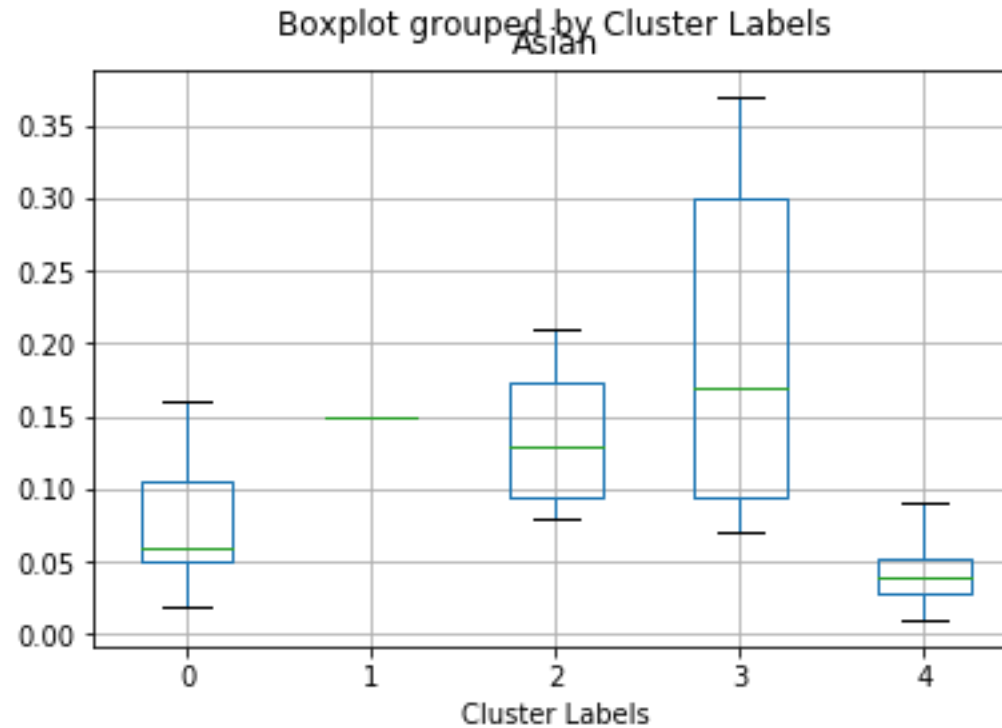
How Clusters differ in Demographics (White)



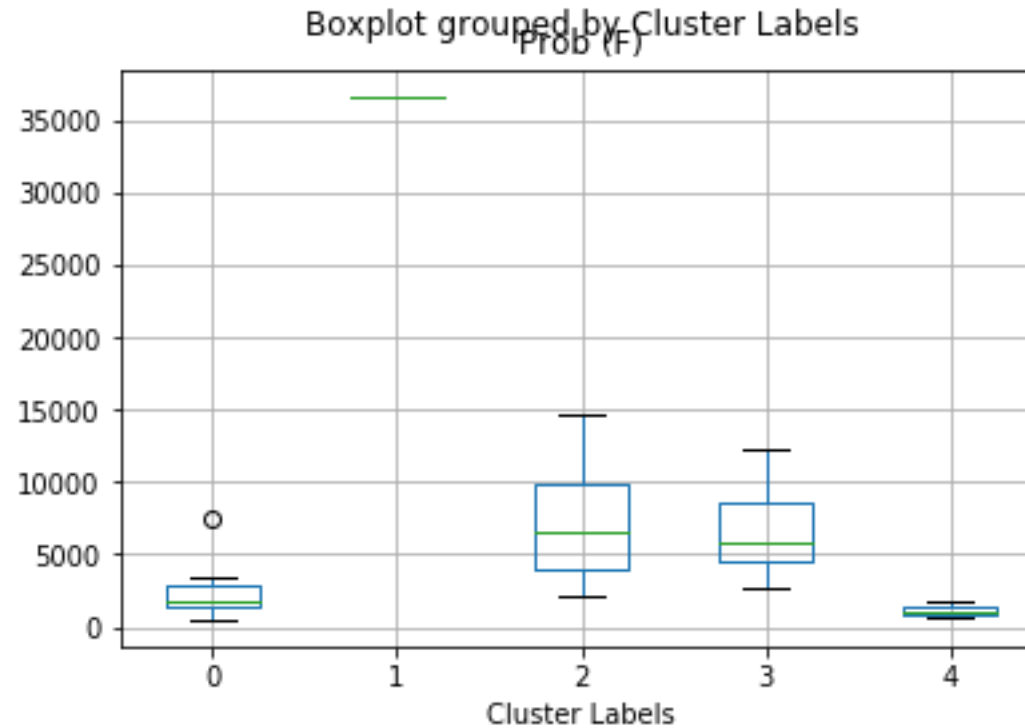
How Clusters differ in Demographics (Black)



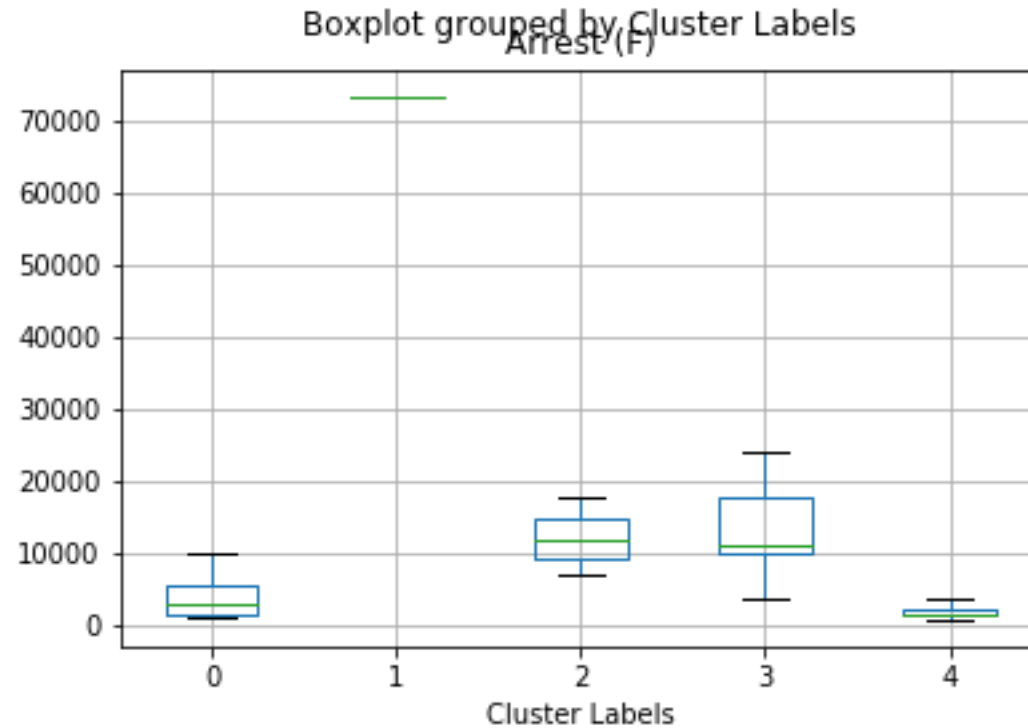
How Clusters differ in Demographics (Asian)



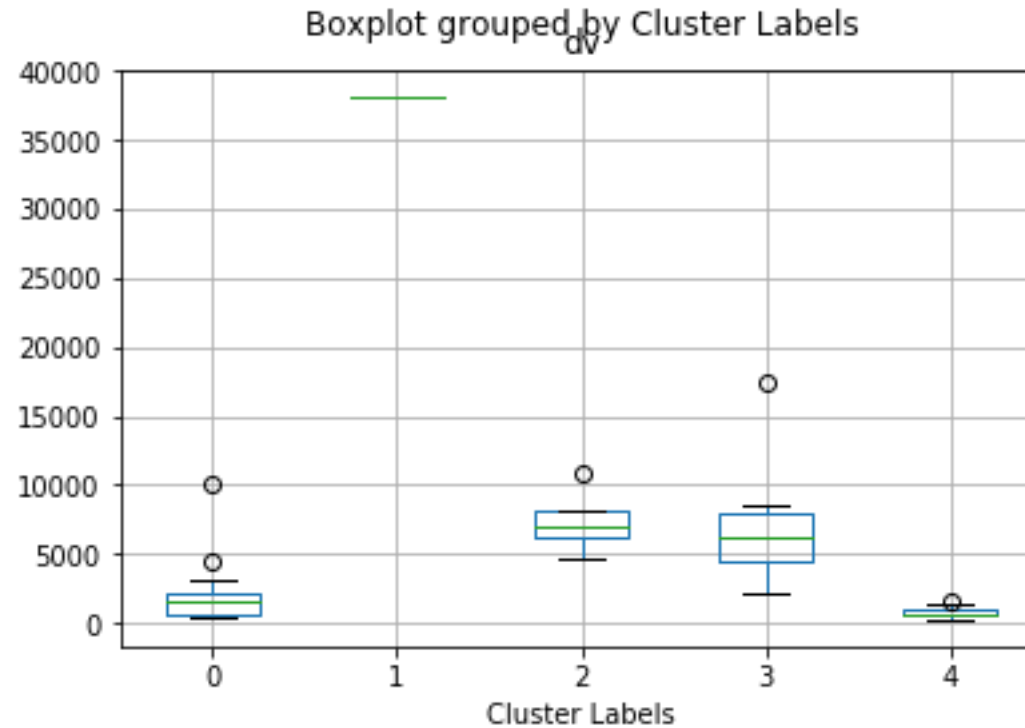
How Clusters differ in Crime (Probation)



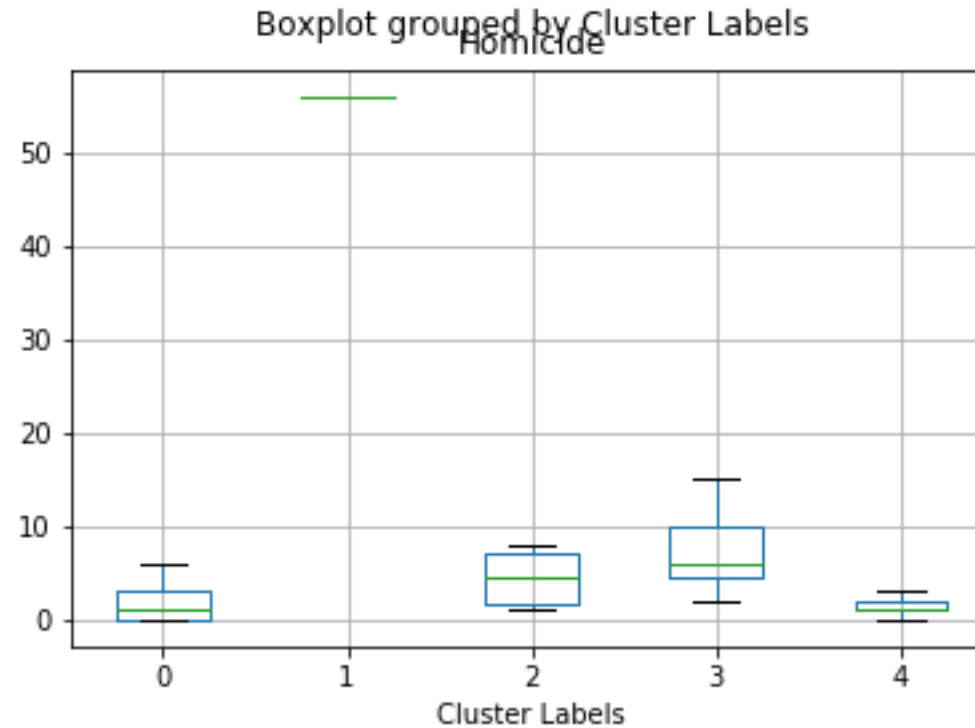
How Clusters differ in Crime (Arrest)



How Clusters differ in Domestic Violence



How Clusters differ in Homicide



Conclusion

The **most preferred place** to retire is the Counties with reasonable mix of demography has the lowest crime rates and that along with the popular places such as Winery, Parks, Trail, Farms and Restaurants makes it the. It is Cluster 4. The following are the Counties and the popular joints. See the below table and find where you like to retire!

The **least preferred place** to retire is Los Angeles, because of the high crime rate.

```
In [187]: df_merged_explore.loc[df_merged_explore['Cluster Labels'] == 4, county_names + venue_columns]
```

Out[187]:

	County Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Nevada County, California	Breakfast Spot	American Restaurant	Brewery	Dessert Shop	Coffee Shop
3	Shasta County, California	Coffee Shop	Mexican Restaurant	Ice Cream Shop	Gay Bar	Restaurant
12	Mendocino County, California	State / Provincial Park	New American Restaurant	Winery	Grocery Store	Nature Preserve
14	Placer County, California	Ski Area	Brewery	Farm	Café	Coffee Shop
18	Lake County, California	Coffee Shop	Italian Restaurant	Café	Breakfast Spot	Mexican Restaurant
19	Sonoma County, California	Grocery Store	Winery	Park	Coffee Shop	Wine Bar
28	Humboldt County, California	Bagel Shop	Brewery	Restaurant	Coffee Shop	Ice Cream Shop
30	Napa County, California	Winery	Hotel	New American Restaurant	Vineyard	Grocery Store
31	Butte County, California	Bathing Area	Pizza Place	Park	Italian Restaurant	Coffee Shop
32	El Dorado County, California	Farm	American Restaurant	Brewery	Grocery Store	Vineyard
35	Marin County, California	Trail	Beach	Scenic Lookout	Park	Coffee Shop
37	San Luis Obispo County, California	Grocery Store	Brewery	Winery	Burger Joint	Deli / Bodega