

# Battle of California Cities

IBM Applied Data Science: Coursera Capstone  
By: Shannon Welch





# Business Problem:

- Traveling across California not knowing what are the best places to stop and see along the way.
- This Project gives insight as to where various types of venues are located as well as the frequency of them in each city in California.
- This can be used to pick the best place to go as well as real time trending spots in a city of your choosing.



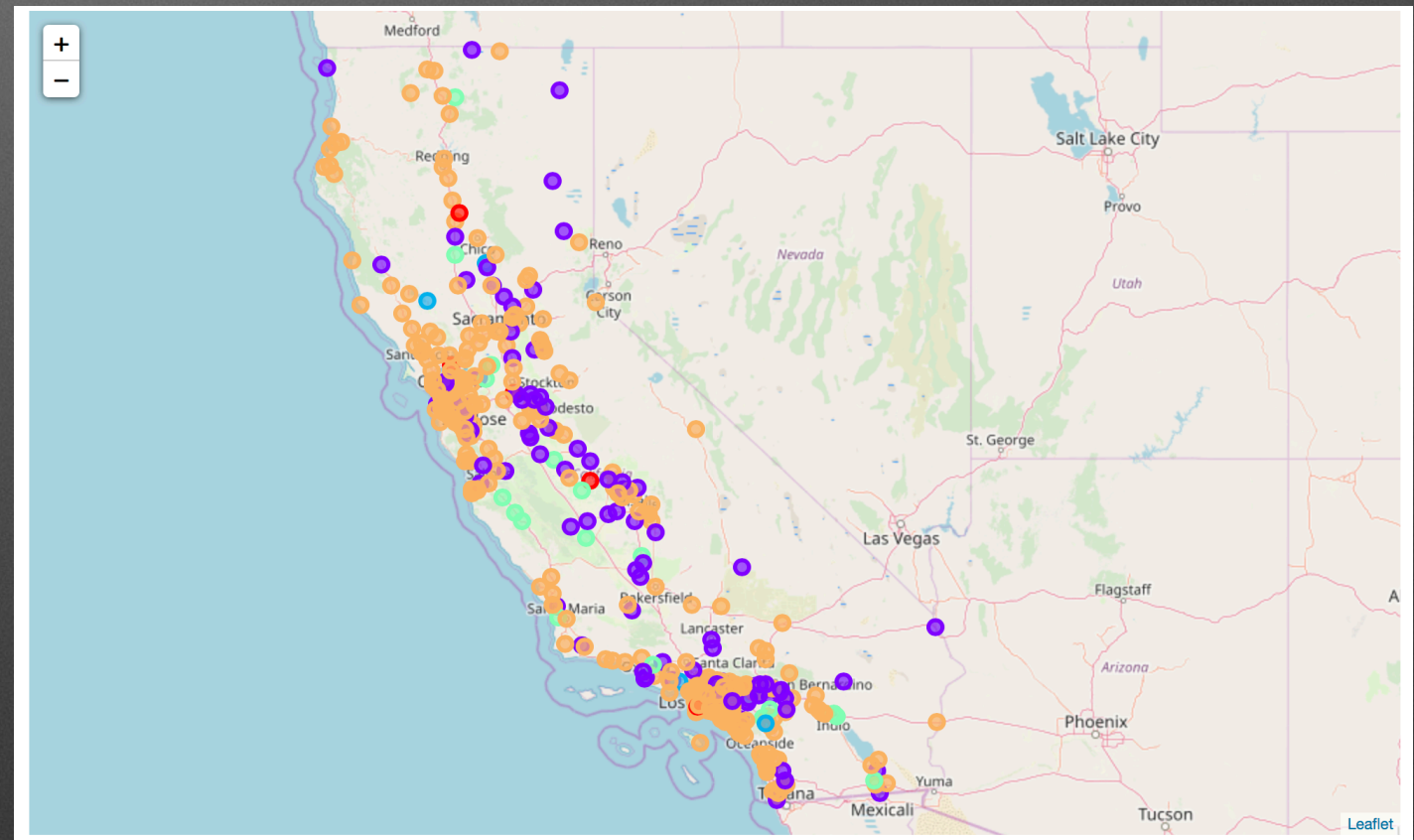
# Data Sources:

- The California Cities dataset came from Kaggle.
- <https://www.kaggle.com/camnugent/california-housing-feature-engineering>
- The dataset used could be found using the above link.
- This dataset returns 459 cities across California.
- Foursquare data was used to show venues in each City.



# Clustering Data:

- The venue data was split up into 5 different Clusters using the most common venue data.
- It returned clusters some being more suburban type areas and others being larger more cultured cities.





# Results:

- The results allow a consumer to pick a city and then get venue results. From those results they can then chose different categories of venues.
- There is also code to return real time trending venues in the desired City.



# Conclusion:

This project has gone through the process of identifying the business problem, specifying the required data, extracting and preparing the data, performing machine learning by clustering the data into 5 different clusters throughout California and giving people potential ideas as which cities have the most abundance of interesting venues to them.