Best Venue for Barber/Stylists & Customers in Sacramento, CA

How do I know what place is “best?”

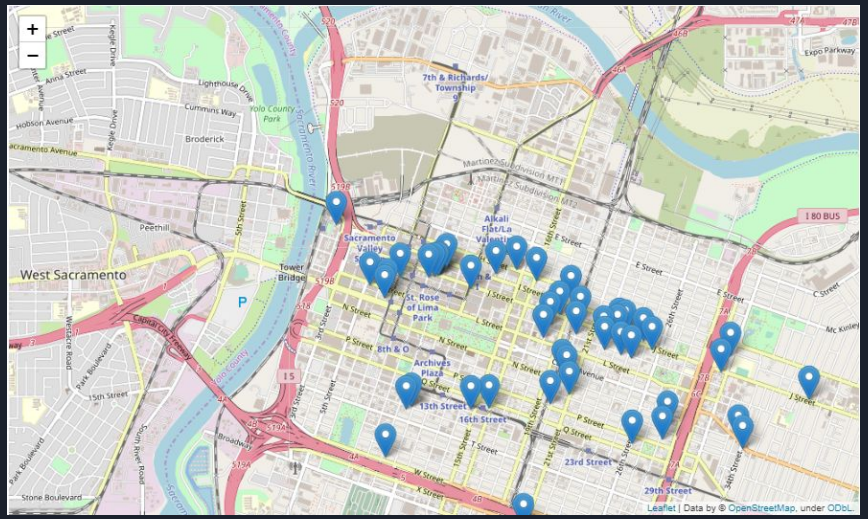
- Generally we check the ratings of the venue in question
- We also consider the total number of ratings
- Usually more ratings means safer choice and must be good if it that many people reviewed it.

Where did the data come from? Cleaning?

- Foursquare API for Venue name, location, category type
  - (https://api.foursquare.com/v2/venues/search)
- Google Places API for average rating and total rating count
  - (https://maps.googleapis.com/maps/api/place/nearbysearch)
- Duplicate, highly similar or highly correlated features were dropped.
- Final count of observations after were 54

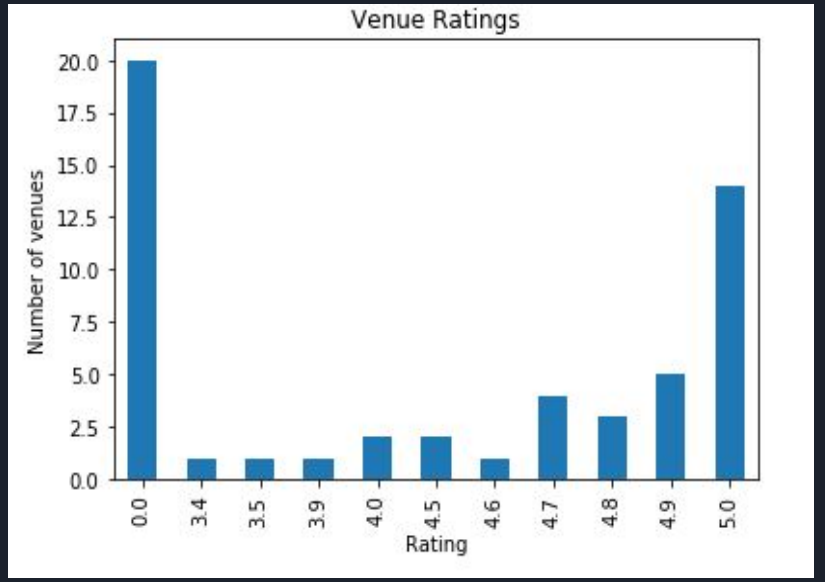
Market Concentration

- After the data was cleaned we validated that all observations were in the same location



Rating Distribution

- Polarized at the extremes
- 0-5 values were most common
- Outliers?
- Surge of new participants entering market?



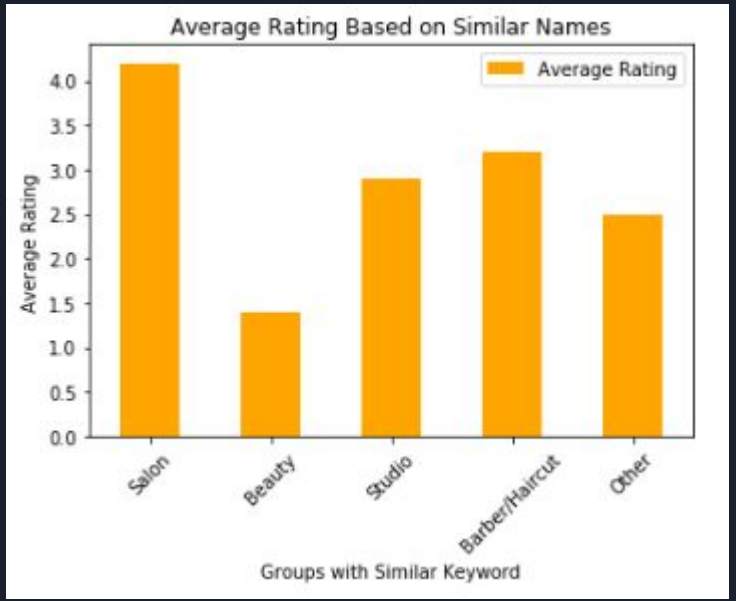
Which venues account for these ratings?

- Clustered by similar keyword in venue name and given a value
  - Eg. Salon, Beauty, Studio, Haircut/barber, other

Allure Salon & Spa	4.7	27.0	1
The look threading salon	4.9	16.0	1
Salon Entro	3.4	9.0	1
Michael Z Salon	5.0	10.0	1
Double Take Hair Gallery	5.0	7.0	4
Lust Beauty Bar	0.0	0.0	2
Nina's Studio	0.0	0.0	3
Sola Salon Studios	4.7	36.0	1

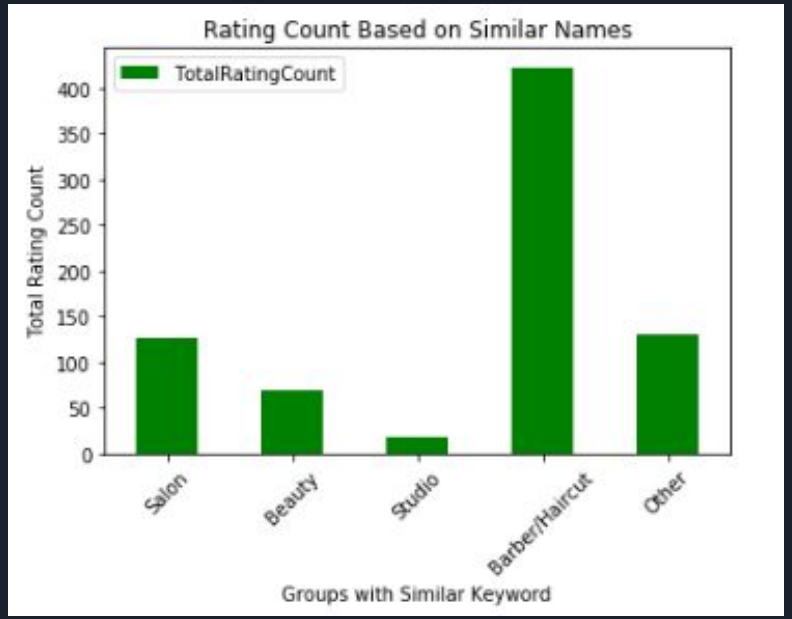
Average Ratings based on similar name

- Salon group has highest
- Others seem more closely matched
- Would rating counts normalize this?



Total rating count per group

- Barber/Haircut takes a massive lead
- Salon is now looks tied with Other group



Why is Barber/haircut so far ahead?

- Trending venue group?
- Test this by using k-Nearest Neighbors to predict the venue category of new venue if one were to be built
- Then validate accuracy of model

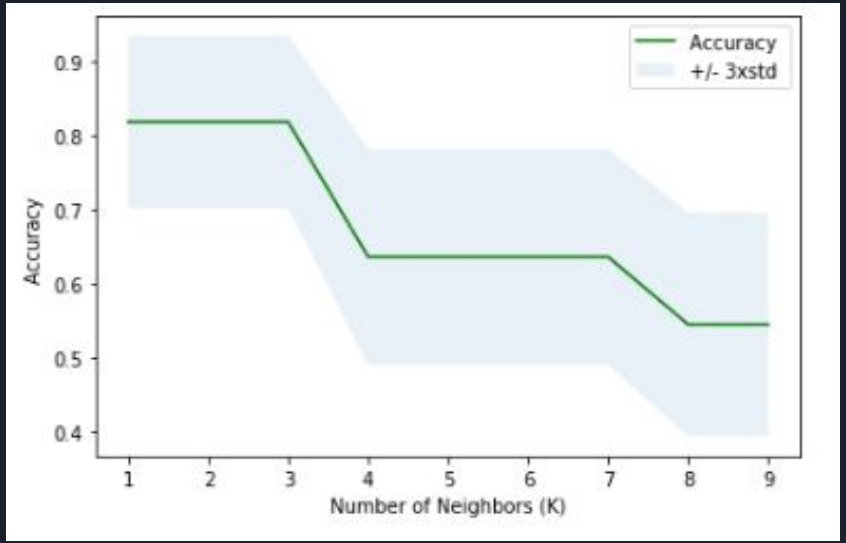
Prediction and Accuracy

- The model predicted that the next 10 venues to open, 5 will have Barber/haircut nomenclature.
- The accuracy seems pretty good for in sample testing.

```
Train set Accuracy:  0.9534883720930233
Test set Accuracy:  0.8181818181818182
```

K-value with highest accuracy

- For-loop was constructed to check all possible k-values 1-10
- It seems with this model, k values 1-3 are equivalent.



Conclusion

- We would like to dive deeper and test to see if there are any impacts of ratings on the foot-traffic and revenue.
    - Financial data from venues, count of transactions per venue
- Lastly check the financial impact on the smaller groups that have few competitors relative to the bigger groups.