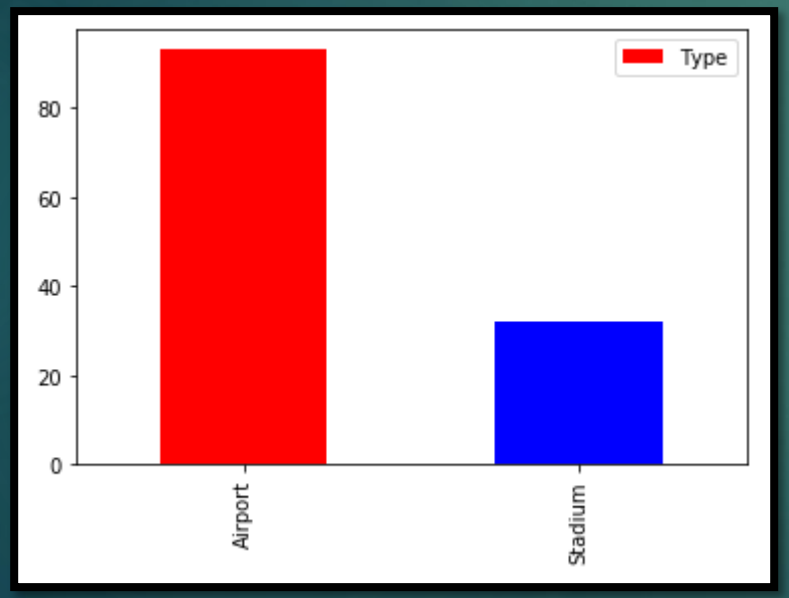
Applied Data Science Capstone

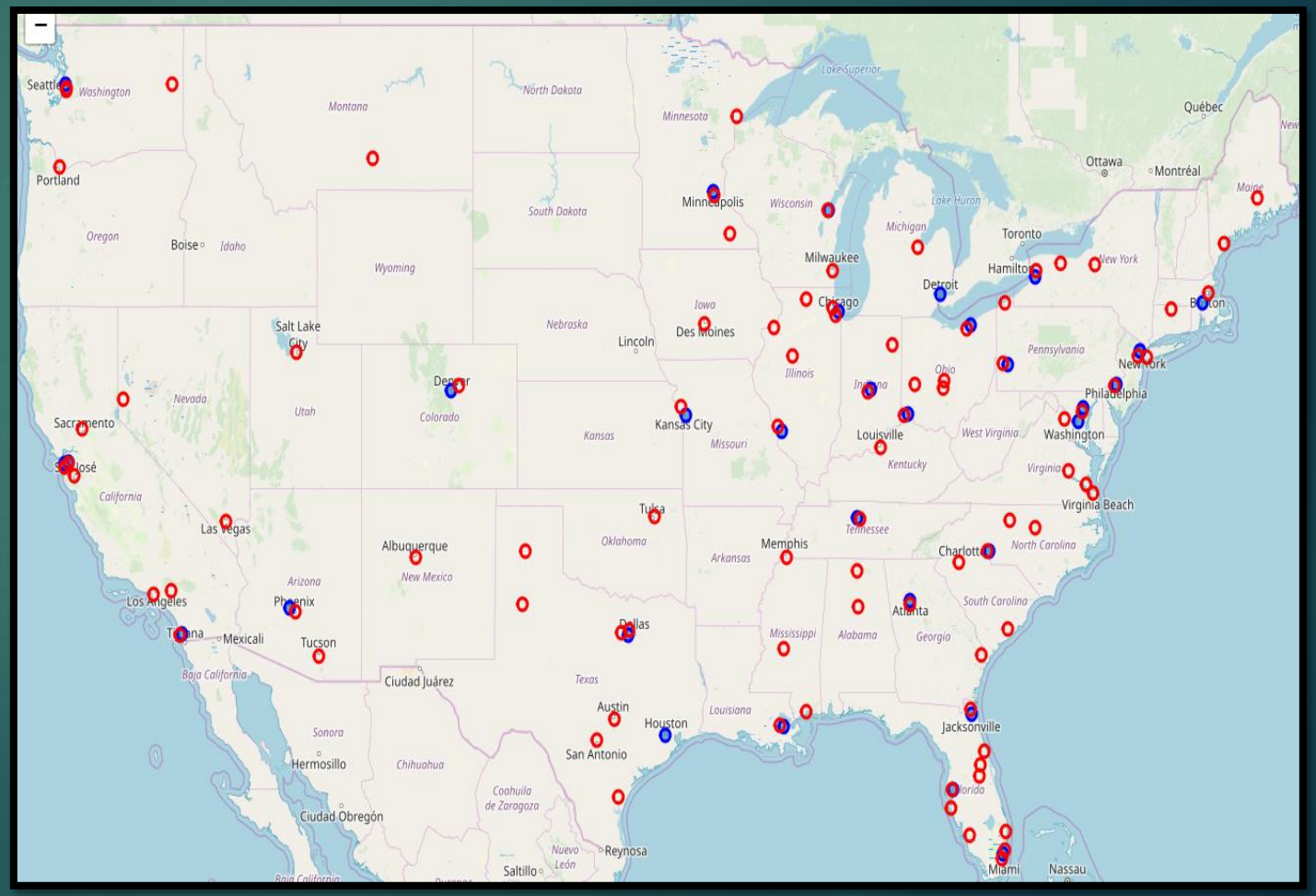
DETERMINING WHETHER A LOCATION HAS AN NFL STADIUM OR INTERNATIONAL AIRPORT USING FOURSQUARE DATA

Stadium versus Airport

This analysis compares the areas surrounding NFL Stadiums and International airports to understand if the venues surrounding them can be used to identify them.

NFL Stadiums – 32
International Airports - 62





Foursquare Categories

- The following tables provide the average number of venue types by airport and stadium
- Coffee Shops lead the venues at Airports with 1.5 per location
- Bars/Sports Bars lead the categories for Stadiums.
- There are overlapping categories however like Coffee Shops, American Restaurants, and Fast Food Restaurants.

----Airport----

	venue	freq
0	Coffee Shop	1.50
1	American Restaurant	0.66
2	Fast Food Restaurant	0.45
3	Snack Place	0.42
4	Gift Shop	0.40
5	Burger Joint	0.34
6	Sandwich Place	0.34
7	Electronics Store	0.32
8	Pizza Place	0.31
9	Mexican Restaurant	0.31

----Stadium----

	venue	freq
0	Bar	1.00
1	Sports Bar	0.94
2	Hotel	0.88
3	American Restaurant	0.84
4	Coffee Shop	0.62
5	Sporting Goods Shop	0.59
6	History Museum	0.53
7	Park	0.50
8	Fast Food Restaurant	0.44
9	Sports Club	0.44

Machine Learning

- Multiple algorithms were applied to a test data set.
- The most successful in identifying a Stadium versus an Airport were KNN and Decision Tree

	Algorithm	Jaccard	1-score	LogLoss
0	KNN	0.79	0.76	NA
1	Decision Tree	0.74	0.73	NA
2	SVM	0.68	0.60	NA
3	LogLoss	0.68	0.60	0.535326

Results

	Algorithm	KNN	Decision Tree
0	Accuracy	0.96	0.89
1	Misclassification	0.04	0.11
2	Sensitivity	0.88	0.75
3	False Positive Rate	0.00	0.03
4	Specificity	1.00	0.97
5	False Positive Rate	0.00	0.03

		KNN Actual	
		Stadium	Airport
KNN Predicted	Stadium	28	0
	Airport	4	62

		Decision Tree Actual	
		Stadium	Airport
Decision Tree Predicted	Stadium	24	2
	Airport	8	60

KNN performed slightly better. It was able to identify all stadiums correctly and identified 28 of the 32 stadiums correctly. 4 stadiums were misidentified as airports.

Conclusion

- The algorithm worked well however it could be improved by better categorizing the surrounding venues. The categories provided by foursquare are specific (i.e. type or restaurant, bar). It might help to generalize the categories. The radius could also be revised to include/exclude venues.
- This exercise was limited to specific locations. It would be interesting to expand the analysis to use general locations to see if it can be determined whether a stadium or airport exists there.
- It seems that, for these examples, foursquare data can be used to identify if a type of venue is in the vicinity.