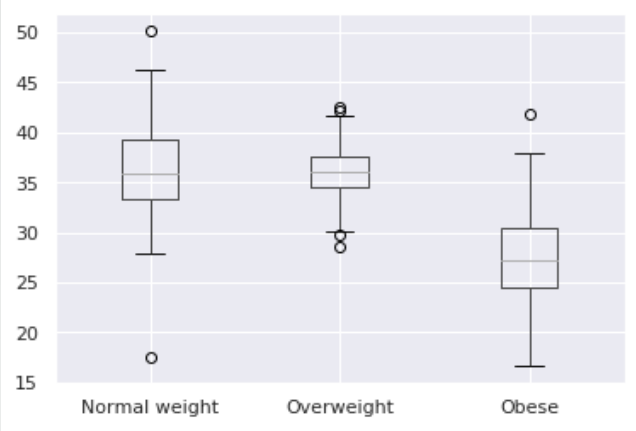
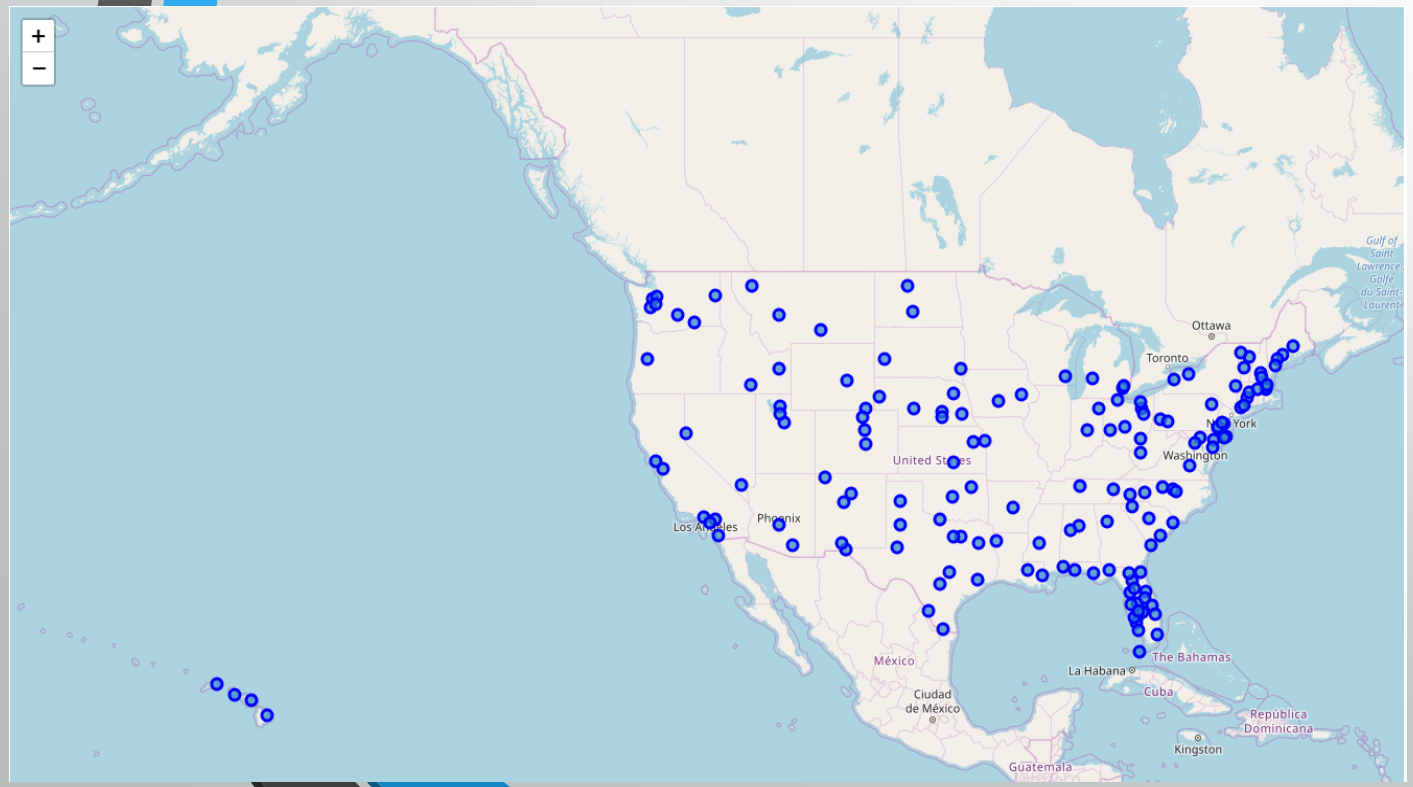
Is there a link between restaurants in a city and the city's obesity rates?

Datasets

Obesity rates & Geographical data





2 Initial datasets…

Obesity rates of 192 American cities
Data follows a normal distribution, with 2 outlier cities

Census data of 28,889 incorporated US cities & towns

…combined into a final dataset

The following data for 160 cities:

- Normal weight, % of population
- Overweight, % of population
- Obese, % of population
- Latitude
- Longitude

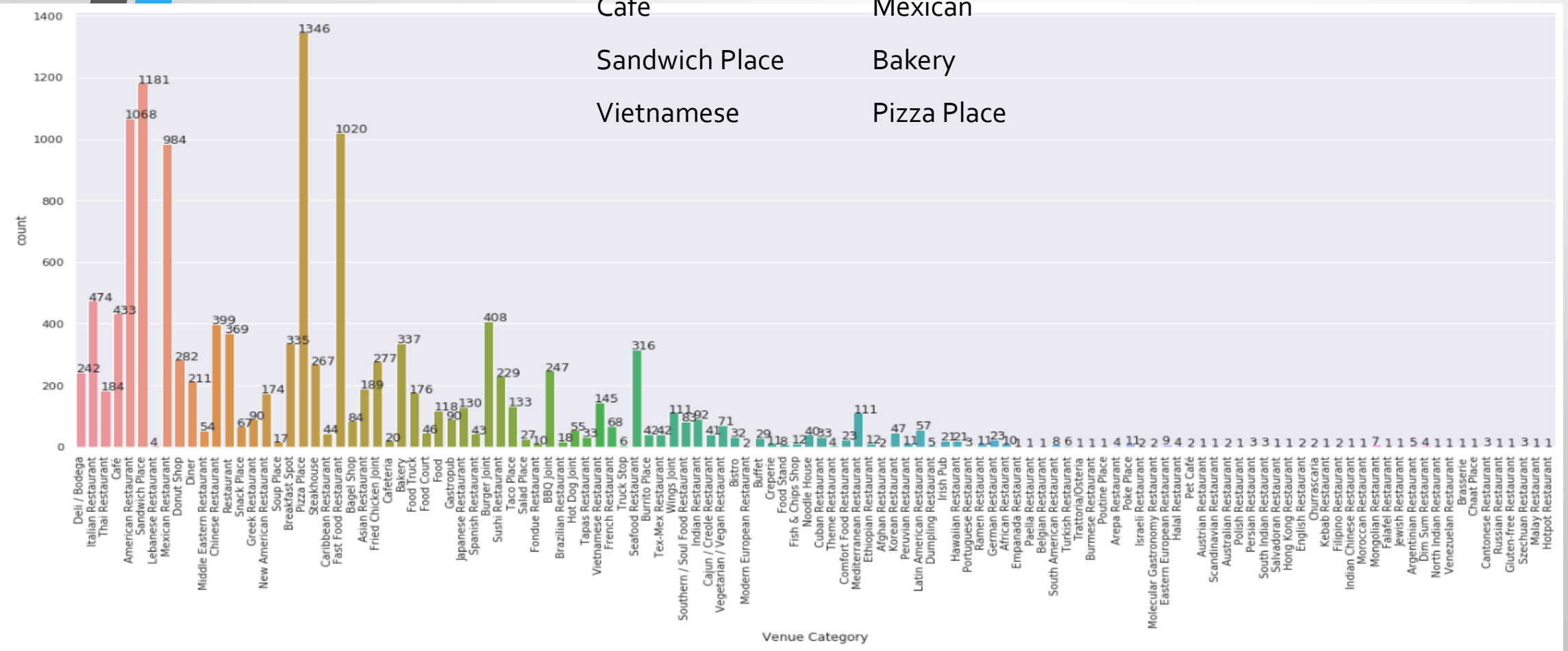
Restaurant dataset

13,568 results across 160 cities

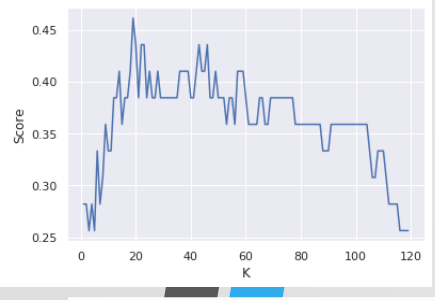
120 different restaurant categories - Most common category is Pizza place

Between the 160 cities, 8 categories come back as the most common ones:

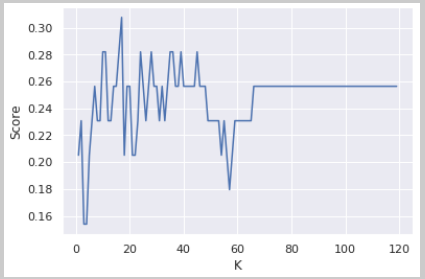
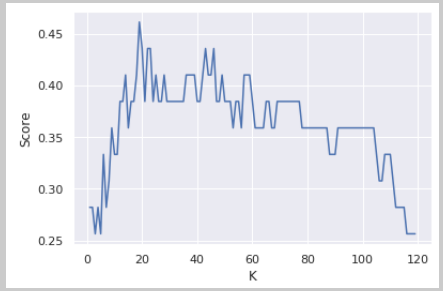
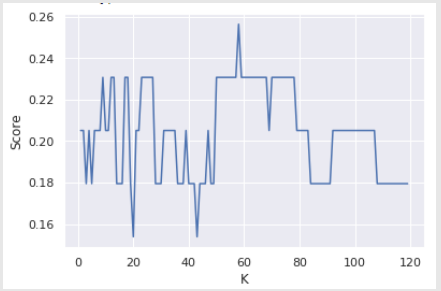
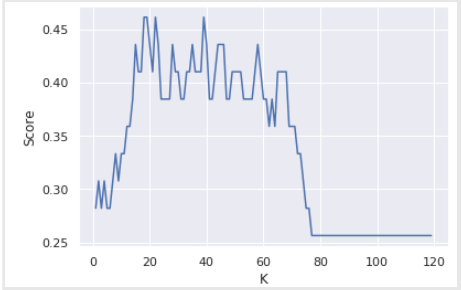
American	Fast Food
Café	Mexican
Sandwich Place	Bakery
Vietnamese	Pizza Place



Analysis



Algorithm evaluation & performance

Algorithm	Score without GPS coordinates as input variables	Score with GPS coordinates as input variables
K-Nearest neighbors, uniform weights	K: 17 Score: 0.308 	K: 18 Score: 0.462 
K-Nearest neighbors, distance weights	K: 58 Score: 0.256 	K: 19 Score: 0. 462 
Naïve Bayes Gaussian	0.308	0.385
Decision Tree	0.256	0.282

K-Nearest neighbors, uniform weights, with K = 18 yields the best results

Final results

F1 score:
4/8 categories have a score of 0
2/4 between 0 and 0.5
2/4 ~0.62

Jaccard index	array([0. , 0. , 0. , 0.28571429, 0.46153846, 0.44444444, 0.22222222, 0.])
F-1 score	array([0., 0. , 0. , 0.44444444, 0.63157895, 0.61538462, 0.36363636, 0.])

Conclusion

Future work

Other input variables to consider

- Physical activity of city's population
- 2nd & 3rd most common restaurant types