Introduction / Business Problem

Vegetarian and vegan restaurants are becoming more and more popular, especially in high population cities like New York City, Los Angeles, and Chicago. Because there are fewer vegetarians in smaller cities, opening vegetarian/vegan restaurants in smaller cities can still be a risky proposition.

We can decrease this risk by identifying neighborhoods that are most similar to neighborhoods in big cities with high numbers of vegetarian/vegan restaurants.

Many types of people would be interested in this type of analysis. These include people interested in starting a restaurant, investors, developers, and even city officials looking to improve different parts of their city in order to attract more people.

Data

For this project, I plan to train a machine learning model to identify neighborhoods that are likely to support a new vegetarian / vegan restaurant.

To train this model, I will use neighborhood data from NYC, LA, and Chicago, which all have many vegetarian vegan restaurants. I think it's important to use training data from a few different cities to control for any particular demographics or characteristics unique to one city.

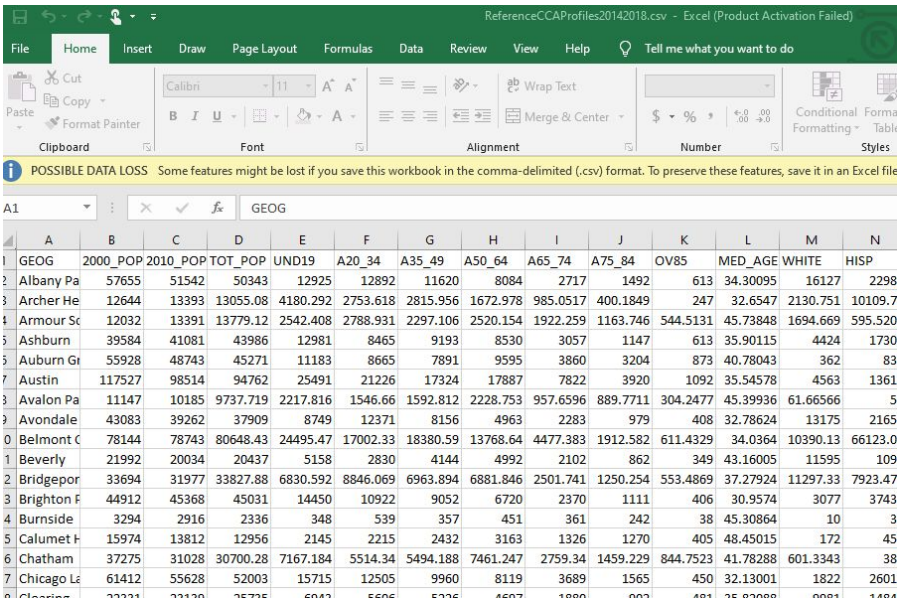
This neighborhood data will consist of:

- Venue data from Foursquare
- Income and demographic data from the links below. Screenshots of some of the data are included below each link.
    - NYC: http://app.coredata.nyc
        - Example income data:

Sub-Borough Area	2000	2005	2006	2007	2008	2009	2010	2011	2012
Astoria	$57,619	$46,660	$53,078	$56,937	$61,424	$55,715	$53,730	$52,458	$57,107
Bay Ridge	$67,577	$60,180	$69,438	$60,340	$64,696	$62,182	$55,857	$63,490	$57,710
Bayside/Little Neck	$88,792	$81,950	$85,040	$85,921	$87,299	$84,651	$81,411	$86,495	$79,404
Bedford Stuyvesant	$35,831	$38,140	$34,784	$38,432	$45,537	$35,423	$42,712	$36,308	$41,960
Bensonhurst	$53,433	$49,670	$48,794	$50,860	$45,604	$45,307	$44,166	$49,929	$55,084
Borough Park	$49,818	$40,450	$50,465	$46,605	$47,353	$48,129	$41,737	$38,599	$41,644
Brooklyn Heights/Fort Greene	$64,591	$63,290	$69,261	$78,502	$73,358	$77,704	$82,910	$69,671	$83,410
Brownsville/Ocean Hill	$33,474	$27,480	$28,957	$31,006	$32,260	$30,354	$30,300	$28,933	$31,233

    - Chicago: https://datahub.cmap.illinois.gov/dataset/community-data-snapshots-raw-data

- Example demographic data

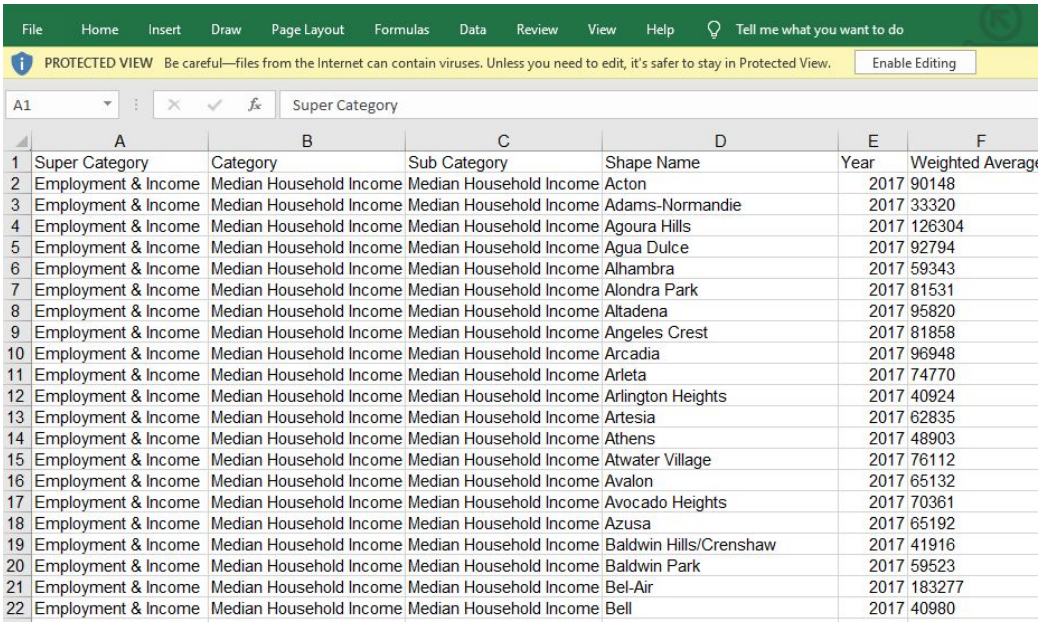


ReferenceCCAProfiles20142018.csv - Excel (Product Activation Failed)

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	GEOG	2000_POP	2010_POP	TOT_POP	UND19	A20_34	A35_49	A50_64	A65_74	A75_84	OV85	MED_AGE	WHITE	HISP
2	Albany Pa	57655	51542	50343	12925	12892	11620	8084	2717	1492	613	34.30095	16127	2298
3	Archer He	12644	13393	13055.08	4180.292	2753.618	2815.956	1672.978	985.0517	400.1849	247	32.6547	2130.751	10109.7
4	Armour Sc	12032	13391	13779.12	2542.408	2788.931	2297.106	2520.154	1922.259	1163.746	544.5131	45.73848	1694.669	595.520
5	Ashburn	39584	41081	43986	12981	8465	9193	8530	3057	1147	613	35.90115	4424	1730
6	Auburn Gr	55928	48743	45271	11183	8665	7891	9595	3860	3204	873	40.78043	362	83
7	Austin	117527	98514	94762	25491	21226	17324	17887	7822	3920	1092	35.54578	4563	1361
8	Avalon Pa	11147	10185	9737.719	2217.816	1546.66	1592.812	2228.753	957.6596	889.7711	304.2477	45.39936	61.66566	5
9	Avondale	43083	39262	37909	8749	12371	8156	4963	2283	979	408	32.78624	13175	2165
10	Belmont C	78144	78743	80648.43	24495.47	17002.33	18380.59	13768.64	4477.383	1912.582	611.4329	34.0364	10390.13	66123.0
11	Beverly	21992	20034	20437	5158	2830	4144	4992	2102	862	349	43.16005	11595	109
12	Bridgepor	33694	31977	33827.88	6830.592	8846.069	6963.894	6881.846	2501.741	1250.254	553.4869	37.27924	11297.33	7923.47
13	Brighton P	44912	45368	45031	14450	10922	9052	6720	2370	1111	406	30.9574	3077	3743
14	Burnside	3294	2916	2336	348	539	357	451	361	242	38	45.30864	10	3
15	Calumet H	15974	13812	12956	2145	2215	2432	3163	1326	1270	405	48.45015	172	45
16	Chatham	37275	31028	30700.28	7167.184	5514.34	5494.188	7461.247	2759.34	1459.229	844.7523	41.78288	601.3343	38
17	Chicago La	61412	55628	52003	15715	12505	9960	8119	3689	1565	450	32.13001	1822	2601

- LA: http://la.myneighborhooddata.org/data/
  - Sample income data



PROTECTED VIEW Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View. Enable Editing

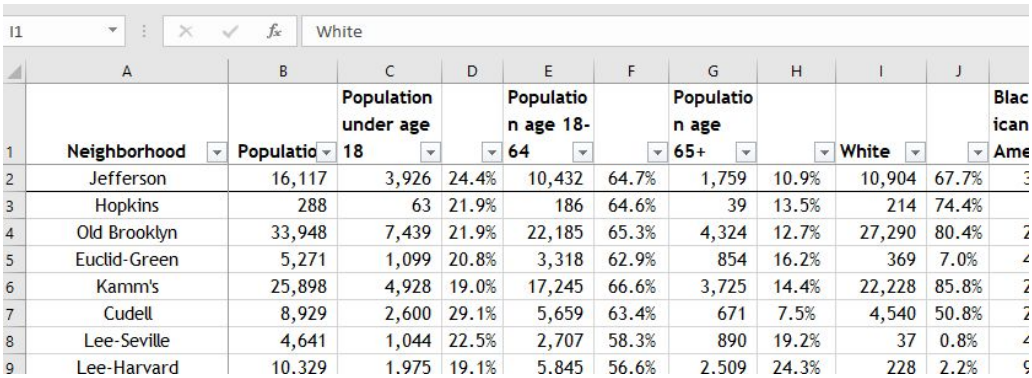
	A	B	C	D	E	F
1	Super Category	Category	Sub Category	Shape Name	Year	Weighted Average
2	Employment & Income	Median Household Income	Median Household Income	Acton	2017	90148
3	Employment & Income	Median Household Income	Median Household Income	Adams-Normandie	2017	33320
4	Employment & Income	Median Household Income	Median Household Income	Agoura Hills	2017	126304
5	Employment & Income	Median Household Income	Median Household Income	Agua Dulce	2017	92794
6	Employment & Income	Median Household Income	Median Household Income	Alhambra	2017	59343
7	Employment & Income	Median Household Income	Median Household Income	Alondra Park	2017	81531
8	Employment & Income	Median Household Income	Median Household Income	Altadena	2017	95820
9	Employment & Income	Median Household Income	Median Household Income	Angeles Crest	2017	81858
10	Employment & Income	Median Household Income	Median Household Income	Arcadia	2017	96948
11	Employment & Income	Median Household Income	Median Household Income	Arleta	2017	74770
12	Employment & Income	Median Household Income	Median Household Income	Arlington Heights	2017	40924
13	Employment & Income	Median Household Income	Median Household Income	Artesia	2017	62835
14	Employment & Income	Median Household Income	Median Household Income	Athens	2017	48903
15	Employment & Income	Median Household Income	Median Household Income	Atwater Village	2017	76112
16	Employment & Income	Median Household Income	Median Household Income	Avalon	2017	65132
17	Employment & Income	Median Household Income	Median Household Income	Avocado Heights	2017	70361
18	Employment & Income	Median Household Income	Median Household Income	Azusa	2017	65192
19	Employment & Income	Median Household Income	Median Household Income	Baldwin Hills/Crenshaw	2017	41916
20	Employment & Income	Median Household Income	Median Household Income	Baldwin Park	2017	59523
21	Employment & Income	Median Household Income	Median Household Income	Bel-Air	2017	183277
22	Employment & Income	Median Household Income	Median Household Income	Bell	2017	40980

Then, I will use data from a smaller city (Cleveland, OH) to identify new neighborhoods where a new vegetarian / vegan restaurant is likely to do well. I will verify the model's accuracy by determining whether it predicted neighborhoods that already contain vegan / vegetarian restaurants. Once validated, this model could in theory be applied to any city.

The neighborhood data I will use can be found at:

- Venue data from Foursquare

- Income and demographic data from: https://www.communitysolutions.com/resources/community-fact-sheets/cleveland-neighborhoods-and-wards/
    - Sample demographics data



I1 | White

	A	B	C	D	E	F	G	H	I	J	
1	Neighborhood	Populatio	Population under age 18		Populatio n age 18-64		Populatio n age 65+		White		Blac ican Ame
2	Jefferson	16,117	3,926	24.4%	10,432	64.7%	1,759	10.9%	10,904	67.7%	3
3	Hopkins	288	63	21.9%	186	64.6%	39	13.5%	214	74.4%	
4	Old Brooklyn	33,948	7,439	21.9%	22,185	65.3%	4,324	12.7%	27,290	80.4%	2
5	Euclid-Green	5,271	1,099	20.8%	3,318	62.9%	854	16.2%	369	7.0%	4
6	Kamm's	25,898	4,928	19.0%	17,245	66.6%	3,725	14.4%	22,228	85.8%	2
7	Cudell	8,929	2,600	29.1%	5,659	63.4%	671	7.5%	4,540	50.8%	2
8	Lee-Seville	4,641	1,044	22.5%	2,707	58.3%	890	19.2%	37	0.8%	4
9	Lee-Harvard	10,329	1,975	19.1%	5,845	56.6%	2,509	24.3%	228	2.2%	9