



# OPENING A LEBANESE RESTAURANT IN LOS ANGELES, US

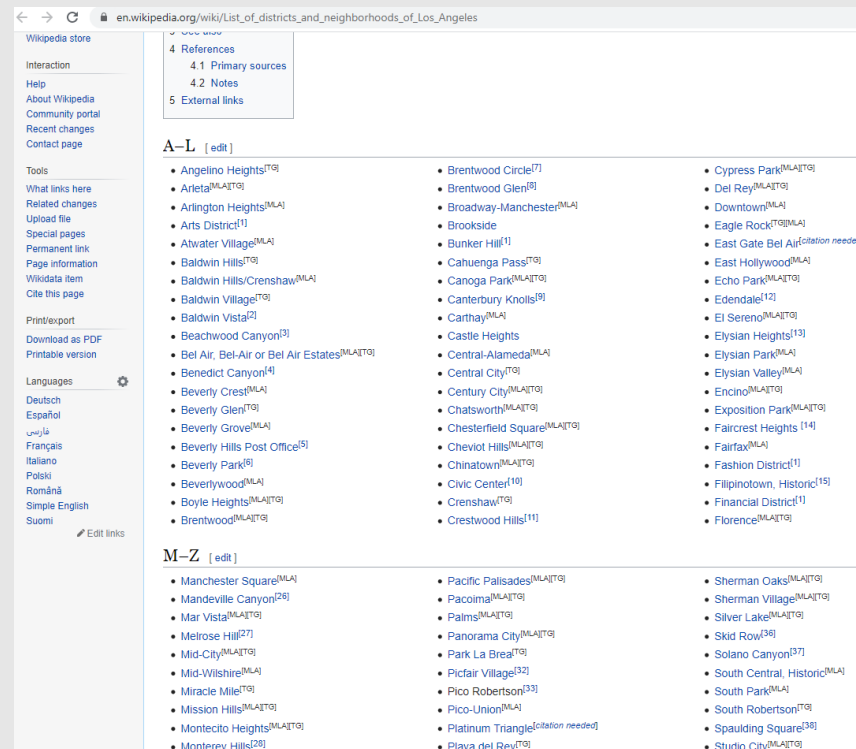
By Abbass Hammoud

# Problem description and background

- Fadi, a Lebanese-American, is aiming to open a new Lebanese restaurant in Los Angeles
- He doesn't know which neighborhood he should pick for the location of his restaurant, in order to maximize his success and eventually his potential revenues
- So he calls our data consulting firm to help him figure out the best neighborhoods, and avoid the less attractive ones.

# Data acquisition

- To get the data for our study we use the Wikipedia page for the list of neighborhoods:
- [https://en.wikipedia.org/wiki/List\\_of\\_districts\\_and\\_neighborhoods\\_of\\_Los\\_Angeles](https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_of_Los_Angeles)



# Data acquisition

- We use this page for the population and area of the neighborhoods:
- [http://wiki.stat.ucla.edu/socr/index.php/SOCR\\_Data\\_LA\\_Neighborhoods\\_Data](http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_LA_Neighborhoods_Data)

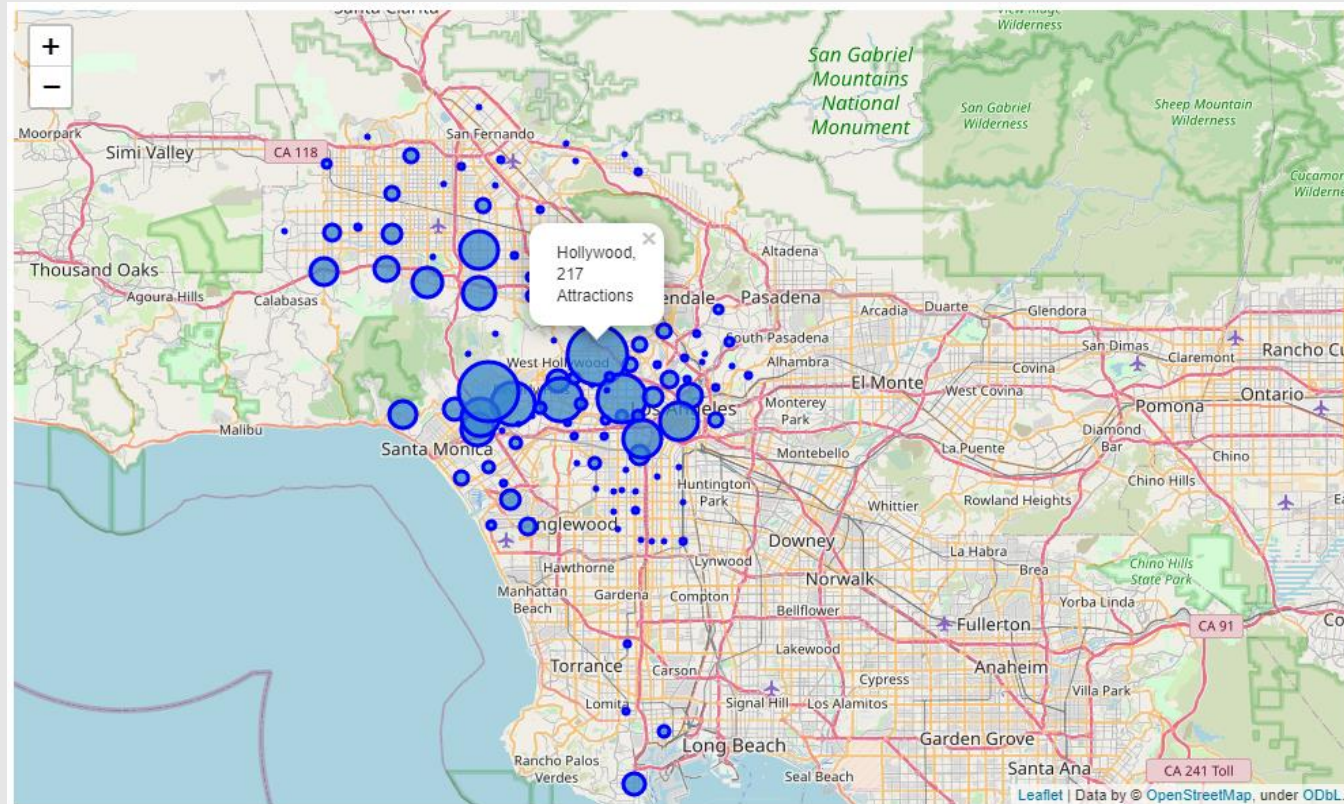
Not secure | [wiki.stat.ucla.edu/socr/index.php/SOCR\\_Data\\_LA\\_Neighborhoods\\_Data](http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_LA_Neighborhoods_Data)

The table below contains the data for the 14 variables observed for all 110 LA neighborhoods: Income, Schools, Diversity, Age, Homes, Vets, Asian, Black, Latino, White, Population, Area, Longitude, Latitude.

LA_Nbhd	Income	Schools	Diversity	Age	Homes	Vets	Asian	Black	Latino	White	Population	Area	Longitude	Latitude
<a href="#">Adams_Normandie</a>	29606	691	0.6	26	0.26	0.05	0.05	0.25	0.62	0.06	31068	0.8	-118.30027	34.03097
<a href="#">Arleta</a>	65649	719	0.4	29	0.29	0.07	0.11	0.02	0.72	0.13	31068	3.1	-118.430015	34.240603
<a href="#">Arlington_Heights</a>	31423	687	0.8	31	0.31	0.05	0.13	0.25	0.57	0.05	22106	1	-118.320109	34.043611
<a href="#">Atwater_Village</a>	53872	762	0.9	34	0.34	0.06	0.2	0.01	0.51	0.22	14888	1.8	-118.265808	34.124908
<a href="#">Baldwin_Hills/Crenshaw</a>	37948	656	0.4	36	0.36	0.1	0.05	0.71	0.17	0.03	30123	3	-118.3667	34.01909
<a href="#">Bel-Air</a>	208861	924	0.2	46	0.46	0.13	0.08	0.01	0.05	0.83	7928	6.6	-118.463558	34.096148
<a href="#">Beverly_Crest</a>	168104	0	0.1	45	0.45	0.1	0.04	0.02	0.03	0.88	10610	7.9	-118.424713	34.112107
<a href="#">Beverly_Grove</a>	63039	791	0.2	38	0.38	0.05	0.05	0.02	0.06	0.82	21417	1.7	-118.372111	34.076009
<a href="#">Beverlywood</a>	105253	872	0.2	39	0.39	0.08	0.07	0.04	0.06	0.8	6080	0.8	-118.395011	34.04161
<a href="#">Boyle_Heights</a>	33235	689	0.1	25	0.25	0.03	0.02	0.01	0.94	0.02	92785	6.5	-118.20537	34.0337
<a href="#">Brentwood</a>	112927	882	0.2	39	0.39	0.08	0.06	0.01	0.05	0.84	31344	15.2	-118.473	34.052
<a href="#">Broadway_Manchester</a>	29897	656	0.5	23	0.23	0.04	0	0.39	0.59	0.01	2347	1.6	-118.27388	33.96153
<a href="#">Canoga_Park</a>	51601	706	0.8	30	0.3	0.06	0.11	0.04	0.51	0.31	53227	4.3	-118.598	34.201
<a href="#">Carthay</a>	71398	762	0.8	37	0.37	0.04	0.09	0.13	0.16	0.58	4866	0.5	-118.369	34.059
<a href="#">Central_Alameda</a>	31559	669	0.1	22	0.22	0.02	0.01	0.13	0.85	0.01	40947	2.2	-118.249	33.998
<a href="#">Century_City</a>	95135	0	0.2	46	0.46	0.12	0.09	0.01	0.04	0.83	5513	0.7	-118.416	34.054
<a href="#">Chatsworth</a>	84549	762	0.6	40	0.4	0.11	0.14	0.02	0.14	0.65	33829	15	-118.602	34.258
<a href="#">Chesterfield_Square</a>	37737	544	0.5	31	0.31	0.07	0.01	0.59	0.37	0.02	6062	0.6	-118.31135	33.99325
<a href="#">Cheviot_Hills</a>	109980	0	0.3	42	0.42	0.11	0.09	0.01	0.09	0.79	5776	1.3	-118.408511	34.04061
<a href="#">Chinatown</a>	22837	784	1	34	0.34	0.04	0.35	0.21	0.32	0.11	20476	0.9	-118.235806	34.063711
<a href="#">Cypress_Park</a>	42615	697	0.2	27	0.27	0.04	0.11	0.01	0.82	0.05	9764	0.7	-118.224306	34.09201
<a href="#">Del_Rey</a>	63317	783	1	35	0.35	0.08	0.14	0.04	0.42	0.36	29895	2.6	-118.423111	33.988811
<a href="#">Downtown</a>	15003	724	1	39	0.39	0.1	0.21	0.22	0.37	0.16	27860	5.8	-118.246707	34.040812

# Data acquisition

- And foursquare library for location data



# Data acquisition

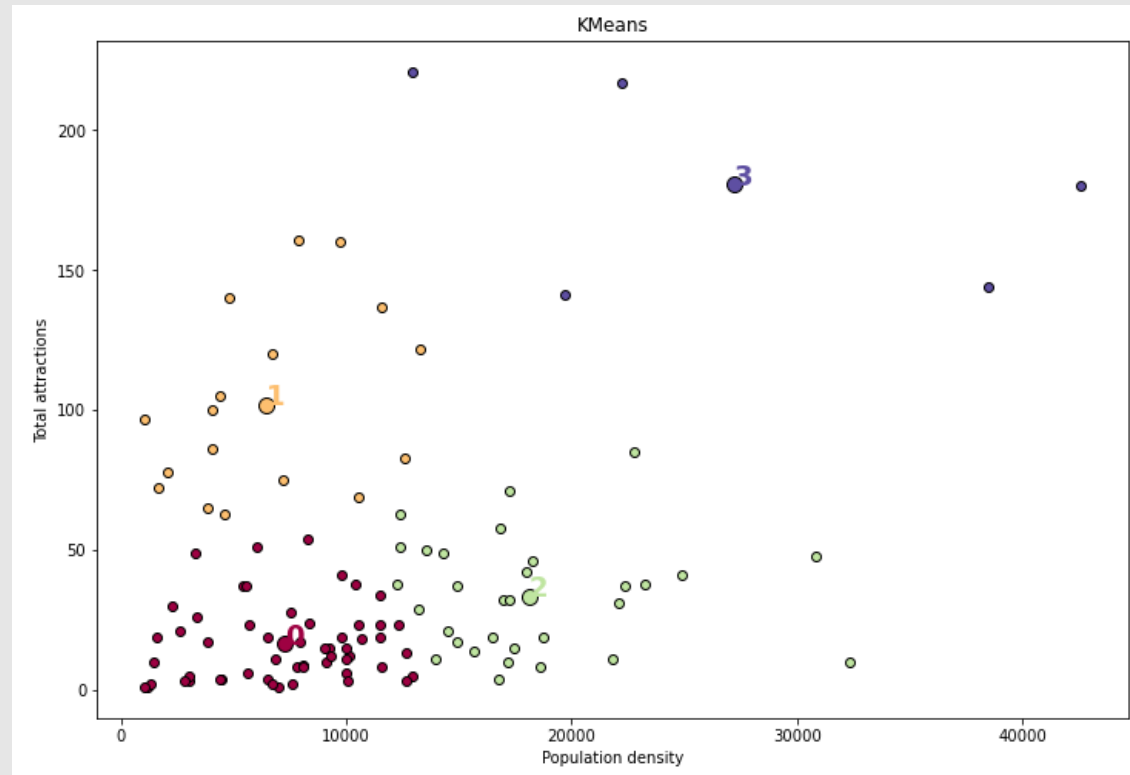
- And here is the dataframe for processing

	Neighborhood	Latitude	Longitude	Population	Area	Population density	Attractions entertainment	Attractions college	Attractions events	Attractions outdoors	Attractions professional	Attractions travel	Total attractions
0	Arlota	34.241327	-118.432205	31068.0	3.1	10022.0	2	1	0	0	3	0	6
1	Arlington Heights	34.043494	-118.321374	22106.0	1.0	22106.0	8	1	1	1	14	6	31
2	Atwater Village	34.118698	-118.262392	14888.0	1.8	8271.0	14	3	0	6	27	4	54
3	Baldwin Hills/Crenshaw	34.007568	-118.350596	30123.0	3.0	10041.0	0	0	0	3	0	0	3
4	Bel Air	34.098883	-118.459881	7928.0	6.6	1201.0	0	0	0	0	1	0	1
5	Beverly Crest	34.116770	-118.432261	10610.0	7.9	1343.0	0	0	0	1	0	1	2
6	Beverly Grove	34.076034	-118.360072	21417.0	1.7	12598.0	2	1	0	17	56	6	82

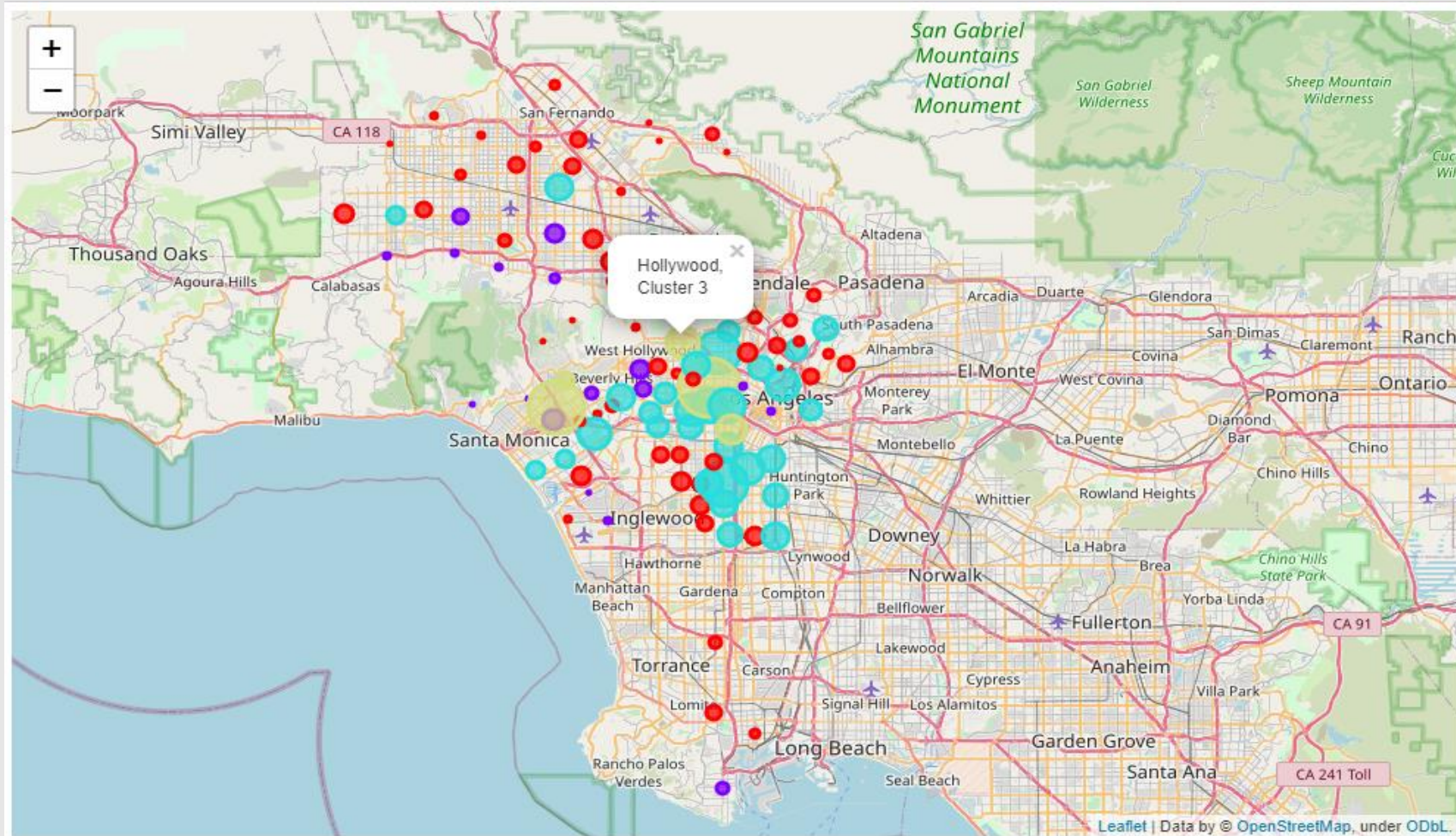


# Machine learning methodology

- We use k-means for clustering the neighborhoods
- Number of clusters = 4



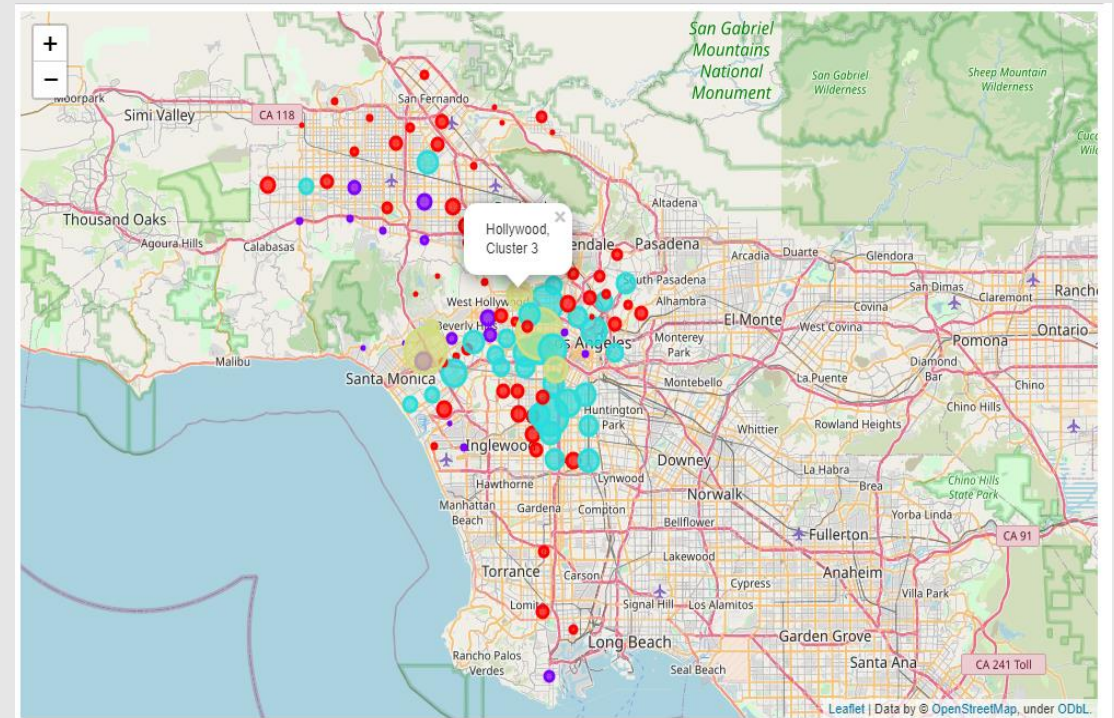
# Neighborhoods clustering





# Neighborhoods clustering

- Cluster 0: contains the neighborhoods with low population density and low number of attractions
- Cluster 1: contains the neighborhoods with high number of attractions but relatively low population density
- Cluster 2: contains the neighborhoods with high population density but relatively low number of attractions
- Cluster 3: contains the neighborhoods with high population density and high number of attractions



# Best neighborhoods

	Neighborhood	Cluster label	Population density	Total attractions	Restaurants count	Lebanese restaurants count
42	Hollywood	3	22234.0	217	84	1
47	Koreatown	3	42619.0	180	65	0
88	University Park	3	19663.0	141	27	0
102	West Los Angeles	3	38459.0	144	25	2
103	Westwood	3	12950.0	221	14	0

# Runners-up

	Neighborhood	Cluster label	Population density	Total attractions	Restaurants count	Lebanese restaurants count
6	Beverly Grove	1	12598.0	83	42	1
9	Brentwood	1	2062.0	78	6	0
12	Carthay	1	9732.0	160	23	0
14	Century City	1	7876.0	161	68	2
21	Downtown	1	4803.0	140	37	0
28	Encino	1	4411.0	105	16	1
64	Pacific Palisades	1	1055.0	97	32	0

	Neighborhood	Cluster label	Population density	Total attractions	Restaurants count	Lebanese restaurants count
1	Arlington Heights	2	22106.0	31	12	0
8	Boyle Heights	2	14275.0	49	26	0
11	Canoga Park	2	12378.0	63	14	0
13	Central Alameda	2	18612.0	8	4	0
18	Chinatown	2	22751.0	85	49	0
19	Cypress Park	2	13949.0	11	4	0
23	East Hollywood	2	30820.0	48	21	2
24	Echo Park	2	16856.0	58	39	0
29	Exposition Park	2	17257.0	71	20	0

# Least attractive neighborhoods

	Neighborhood	Cluster label	Population density	Total attractions	Restaurants count	Lebanese restaurants count
0	Arleta	0	10022.0	6	2	0
2	Atwater Village	0	8271.0	54	27	0
3	Baldwin Hills/Crenshaw	0	10041.0	3	0	0
4	Bel Air	0	1201.0	1	0	0
5	Beverly Crest	0	1343.0	2	0	0
7	Beverlywood	0	7600.0	2	2	0
10	Broadway Manchester	0	1467.0	10	3	0