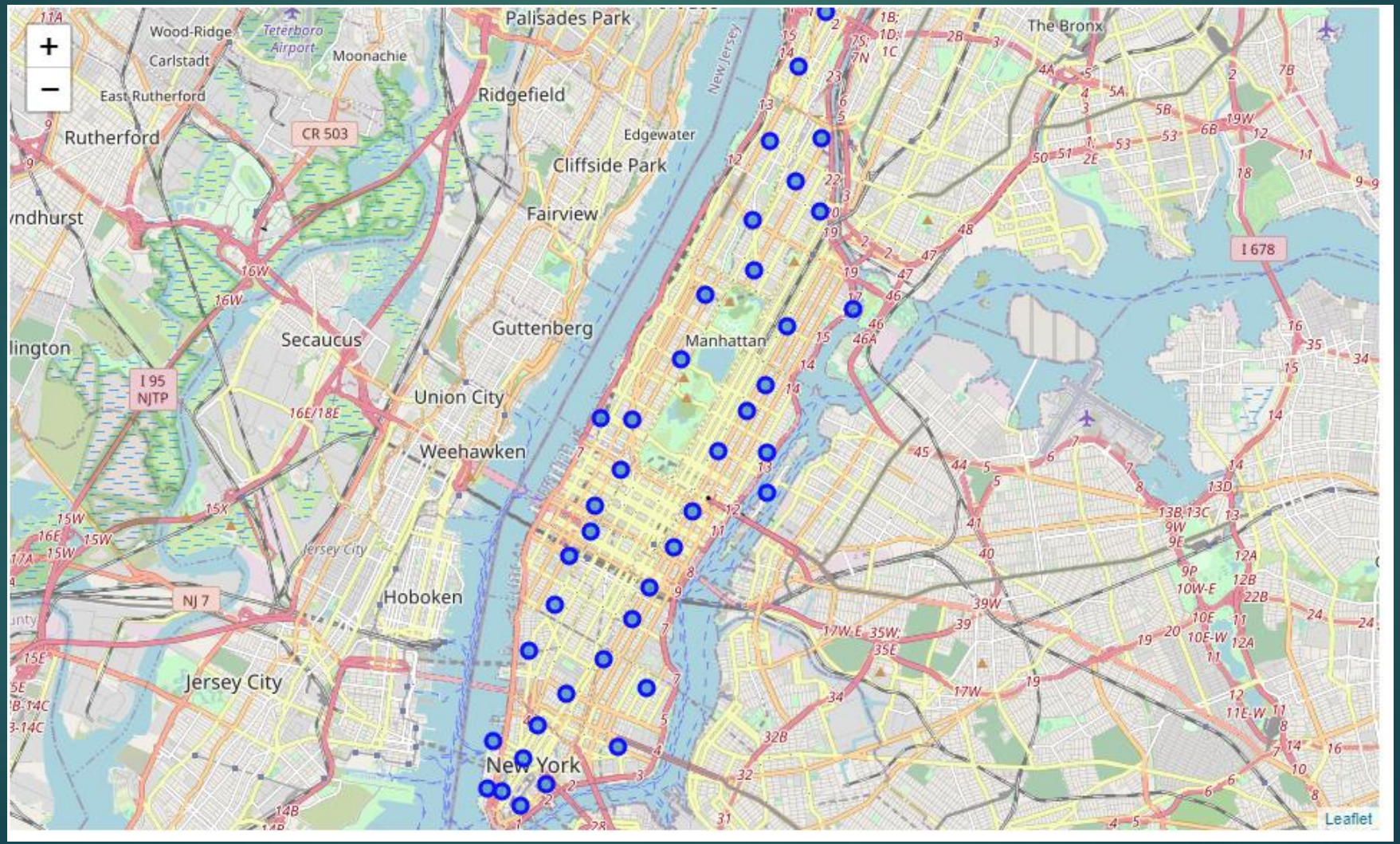
The Battle of Neighborhoods

ALEX RANJBAR

Gather Median Incomes by Zip Code

#	Zip Code	Location	City	Population	Avg. Income/H/hold	National Rank
1.	10007	40.713973, -74.008336	New York, New York	3,522	**$112,947.00**	#132
2.	10280	40.708501, -74.016828	New York, New York	6,614	**$108,536.00**	#157
3.	10162	40.769298, -73.949942	New York, New York	1,726	**$108,416.00**	#158
4.	10004	40.691870, -74.014109	New York, New York	1,225	**$101,868.00**	#238
5.	10069	40.775551, -73.989689	New York, New York	1,403	**$88,091.00**	#507
6.	10282	40.717017, -74.015566	New York, New York	1,574	**$86,066.00**	#591
7.	10006	40.707970, -74.013610	New York, New York	1,447	**$81,334.00**	#766
8.	10022	40.758615, -73.967752	New York, New York	30,642	**$80,406.00**	#820
9.	10005	40.705390, -74.008988	New York, New York	884	**$79,517.00**	#860
10.	10024	40.786307, -73.970694	New York, New York	61,414	**$78,066.00**	#935
11.	10028	40.776865, -73.954721	New York, New York	44,987	**$77,565.00**	#955
12.	10021	40.769742, -73.961653	New York, New York	102,078	**$75,472.00**	#1,082
13.	10023	40.775394, -73.982197	New York, New York	62,206	**$72,424.00**	#1,297
14.	10128	40.781656, -73.950281	New York, New York	59,856	**$70,031.00**	#1,547
15.	10017	40.752292, -73.972393	New York, New York	16,201	**$69,273.00**	#1,611
16.	10014	40.733471, -74.007098	New York, New York	32,667	**$66,601.00**	#1,931
17.	10016	40.744983, -73.978262	New York, New York	51,217	**$66,342.00**	#1,954
18.	10010	40.739122, -73.982363	New York, New York	26,408	**$62,467.00**	#2,471
19.	10011	40.741844, -74.000711	New York, New York	46,669	**$61,986.00**	#2,569
20.	10003	40.731856, -73.989145	New York, New York	53,673	**$60,891.00**	#2,796
21.	10012	40.725639, -73.998090	New York, New York	26,000	**$58,313.00**	#3,310
22.	10019	40.766331, -73.985013	New York, New York	36,012	**$55,869.00**	#3,901
23.	10044	40.762050, -73.949933	New York, New York	9,520	**$49,976.00**	#5,953
24.	10025	40.797958, -73.964917	New York, New York	97,086	**$49,733.00**	#6,044
25.	10018	40.754969, -73.992126	New York, New York	4,255	**$48,705.00**	#6,447
26.	10036	40.759614, -73.991148	New York, New York	18,751	**$41,002.00**	#11,152
27.	10001	40.750722, -73.997276	New York, New York	17,310	**$40,932.00**	#11,228
28.	10009	40.726622, -73.978955	New York, New York	58,595	**$40,176.00**	#11,980
29.	10013	40.720067, -74.004829	New York, New York	25,042	**$38,304.00**	#13,643
30.	10033	40.849124, -73.935956	New York, New York	58,259	**$31,348.00**	#21,836

Map Zip Codes and Income



Gather all gyms in each zip code

In [18]:
```
#Foursquare has two categories for gyms, gather a complete list of both
gymlist = ['Gym', 'Gym / Fitness Center']
gyms = ny_venues[ny_venues['Venue Category'].isin(gymlist)]
gyms
```

Out[18]:

	Zip Code	Latitude	Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
2	10007	40.713973	-74.008336	The Class by Taryn Toomey	58a253c830ecc66c9e5b40a0	40.712753	-74.008734	Gym / Fitness Center
7	10007	40.713973	-74.008336	Equinox Tribeca	4a6e331af964a52031d41fe3	40.714099	-74.009686	Gym
43	10007	40.713973	-74.008336	Equinox Brookfield Place	541a2341498e51ce1037963a	40.712704	-74.014995	Gym
49	10007	40.713973	-74.008336	Y7 Studio - Tribeca	5a817f4859c4231a14ba3998	40.717831	-74.006204	Gym / Fitness Center
59	10007	40.713973	-74.008336	Crunch - FiDi	5a697271ad178907f79be157	40.708614	-74.010013	Gym / Fitness Center
...	...	...	...	...	...	...	...	...
3978	10029	40.792245	-73.945230	92nd Street Y May Center	4c9a873ad4b1b1f79dffd035	40.783108	-73.952758	Gym / Fitness Center
3990	10029	40.792245	-73.945230	ChaiseFitness	52b9c263498ebbff8b240197	40.783065	-73.952648	Gym
4018	10030	40.818492	-73.943283	Harlem YMCA	4a6b5879f964a52082ce1fe3	40.814790	-73.942910	Gym / Fitness Center
4069	10030	40.818492	-73.943283	Blink Fitness	51b35586498e4fcfa012f235	40.810539	-73.951460	Gym
4232	10035	40.795487	-73.929501	Women's World of Boxing	4e4e91ca18505e3814e804c6	40.793549	-73.940495	Gym

162 rows × 8 columns

Find Ratings and Likes for all gyms

```
In [32]: gymratings
```

Out[32]:

	Venue ID	Rating	Likes
0	42af6f80f964a5205a251fe3	9.2	259
1	45938647f964a52045401fe3	8.8	181
2	46b78e2cf964a520bc491fe3	8.7	108
3	49c9439bf964a5204e581fe3	8.9	637
4	49c943d2f964a5204f581fe3	8.4	364
...	...	...	...
72	5a6ca12395d986525fa14684	8.5	18
73	5a817f4859c4231a14ba3998	8.4	20
74	5b60d31767a9fe002c34b374	8.7	14
75	5bccb80067f62b002c693885	7.9	7
76	5c47a642db2aeb002c7a3e73	8.7	16

77 rows × 3 columns

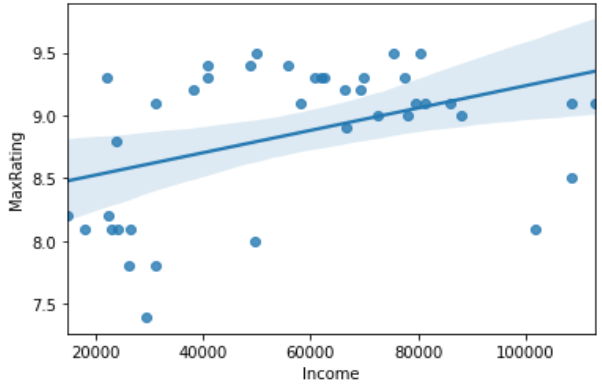
Create DataFrame with Income, Max Rating, Average Rating, and Number of Likes

Out[60]:

Zip Code	Income	MaxRating	AvgRating	SumLikes
10007	112947.0	9.1	8.800000	794
10280	108536.0	9.1	8.880000	774
10162	108416.0	8.5	8.500000	262
10004	101868.0	8.1	8.100000	80
10069	88091.0	9.0	8.775000	1169
10282	86066.0	9.1	8.825000	463
10006	81334.0	9.1	8.880000	774
10022	80406.0	9.5	9.100000	997
10005	79517.0	9.1	8.900000	524
10024	78066.0	9.0	8.566667	460
10028	77565.0	9.3	8.640000	316
10021	75472.0	9.5	8.860000	527
10023	72424.0	9.0	8.742857	986
10128	70031.0	9.3	8.820000	448
10017	69273.0	9.2	8.916667	901
10014	66601.0	8.9	8.900000	637
10016	66342.0	9.2	8.857143	1191
10010	62467.0	9.3	9.114286	753
10011	61986.0	9.3	8.966667	1538
10003	60891.0	9.3	9.150000	564
10012	58313.0	9.1	8.933333	333
10019	55869.0	9.4	9.100000	851
10044	49976.0	9.5	8.725000	504
10025	49733.0	8.0	7.900000	32

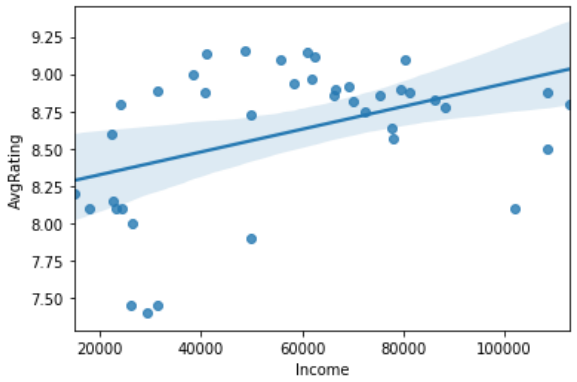
See linear regression R squared and correlation between income and Foursquare data

```
In [70]: sns.regplot(x="Income", y="MaxRating", data=finaldf)
Out[70]: <matplotlib.axes._subplots.AxesSubplot at 0x1ec7fbb45f8>
```



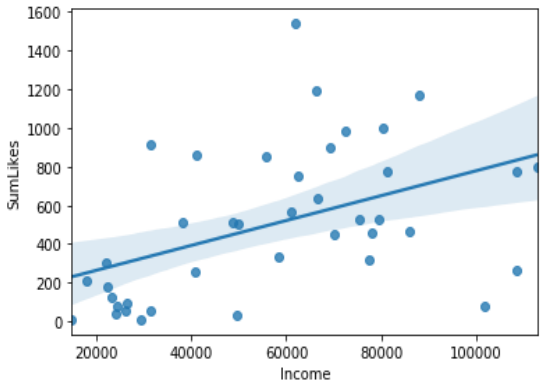
R squared: .1780

```
In [71]: sns.regplot(x="Income", y="AvgRating", data=finaldf)
Out[71]: <matplotlib.axes._subplots.AxesSubplot at 0x1ec0035c2e8>
```



R squared: .1853

```
In [72]: sns.regplot(x="Income", y="SumLikes", data=finaldf)
Out[72]: <matplotlib.axes._subplots.AxesSubplot at 0x1ec003d1d68>
```



R squared: .2145

```
Correlation with Income
MaxRating    0.421953
AvgRating    0.430417
SumLikes     0.463172
```

Create Final DataFrame with GymScore and Sort Descending

In [66]:
```
#Sort by GymScore to find most and least desirable zip codes
finaldfnorm.sort_values(by = 'GymScore', ascending=False, inplace = True)
finaldfnorm
```

Out[66]:

	Income	MaxRating	AvgRating	SumLikes	GymScore
10004	0.887008	0.333333	0.397727	0.045812	3.279022
10162	0.953789	0.523810	0.625000	0.164921	2.766485
10007	1.000000	0.809524	0.795455	0.513089	1.559320
10280	0.955013	0.809524	0.840909	0.500000	1.315109
10025	0.355295	0.285714	0.284091	0.014398	0.892158
10034	0.148729	0.000000	0.000000	0.000000	0.743644
10282	0.725847	0.809524	0.809659	0.296466	0.607595
10024	0.644257	0.761905	0.662879	0.294503	0.445589
10033	0.167790	0.190476	0.028409	0.029450	0.370689
10028	0.639147	0.904762	0.704545	0.200262	0.281142
10005	0.659055	0.809524	0.852273	0.336387	0.151180
10032	0.115664	0.190476	0.028409	0.029450	0.110059
10006	0.677586	0.809524	0.840909	0.500000	-0.072026
10069	0.746499	0.761905	0.781250	0.758508	-0.089579
10128	0.562309	0.904762	0.806818	0.286649	-0.378093
10021	0.617801	1.000000	0.829545	0.338351	-0.417242
10014	0.527328	0.714286	0.852273	0.410340	-0.464887
10037	0.118969	0.333333	0.340909	0.055628	-0.523989
10023	0.586715	0.761905	0.762987	0.638743	-0.630708
10031	0.095756	0.333333	0.397727	0.047120	-0.679853
10012	0.442800	0.809524	0.871212	0.211387	-0.699034