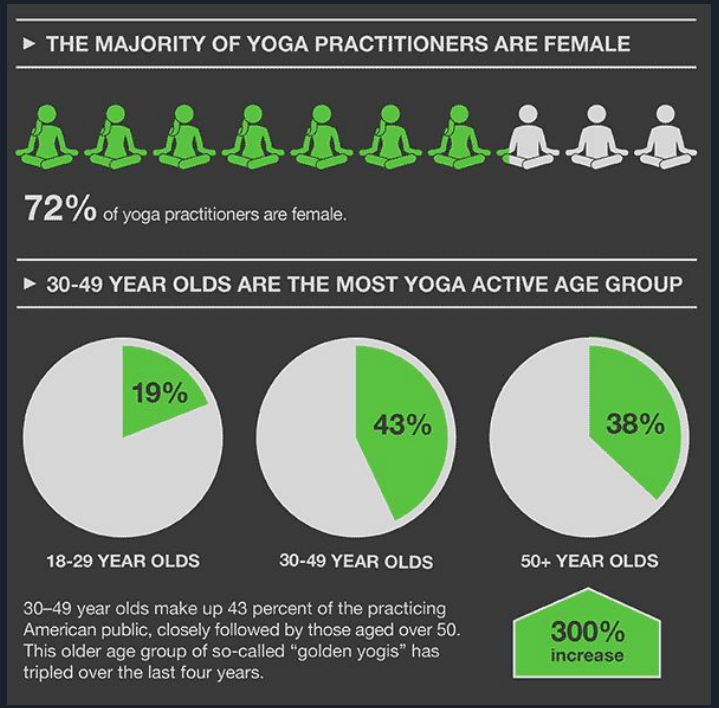
New yoga class in Toronto

Base input


► THE MAJORITY OF YOGA PRACTITIONERS ARE FEMALE
72% of yoga practitioners are female.
► 30-49 YEAR OLDS ARE THE MOST YOGA ACTIVE AGE GROUP
19%
43%
38%
18-29 YEAR OLDS
30-49 YEAR OLDS
50+ YEAR OLDS
30–49 year olds make up 43 percent of the practicing American public, closely followed by those aged over 50. This older age group of so-called "golden yogis" has tripled over the last four years.
300% increase


► MOST YOGA PRACTITIONERS ALSO TAKE OTHER EXERCISE
75% of yoga practitioners also take part in other kinds of exercise including running, cycling and group sports.

Task

Need to find a good place for creating new yoga class in Toronto. Based on different open data source we should find on which attributes of Neighbourhoods have more influence on yoga studios and where is the best place for it, depending of base surroundings. So our definition of done will be list of neighbourhoods of Toronto, which have lack of yoga studios, calculated by some model.

Selectable data

Locations & Mapping – Data Catalogue

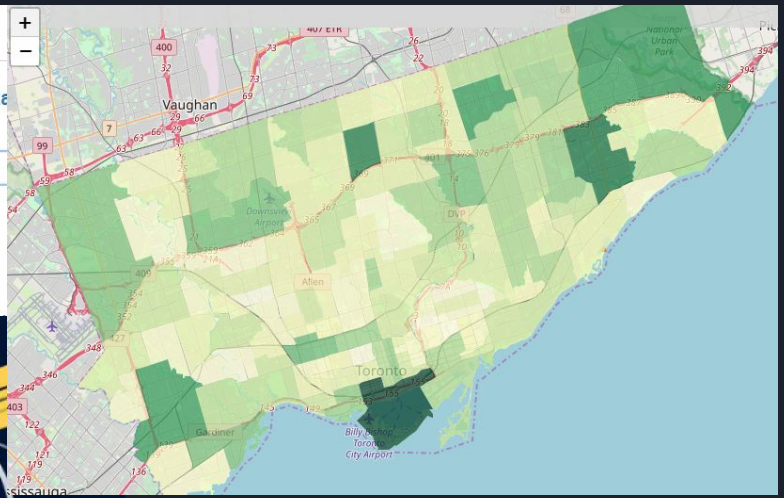
You are now viewing datasets under the **Locations & Mapping** category. Return to the main page of the Open Data Catalogue.

Neighbourhood Profiles

Owner	Social Development, Finance & Administration
Currency	May 2016
Format	csv
Refresh Rate	As Available
Website	https://www.toronto.ca/city-government/data-r
Contact	Open Data
	opendata@toronto.ca

toronto.ca/city-government

Foursquare.com



Toronto geodata

Data acquisition and cleaning

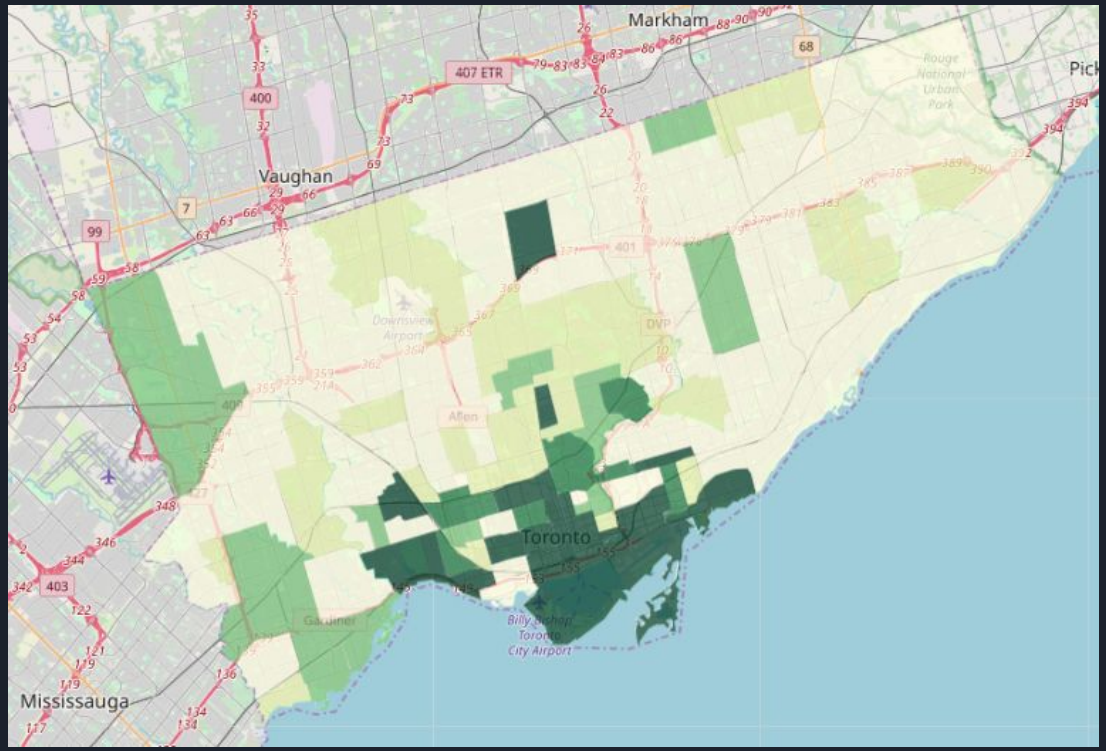
Demographical data downloaded from official site and were combined into one table. There were a lot unnecessary information, so I left there only info about how many:

- male persons are splitted on ages
- female person are splitted on ages
- married persons
- not married persons
- person have no certificate\diploma\degree
- person have secondary certificate\diploma\degree
- person have postgraduate certificate\diploma\degree
- employed persons
- unemployed

Also I got Toronto coordinates and get from foursquare.com all yoga classes across the city and it's coordinates. This was not enough, cause I had to split this yogas across different neighbourhoods. I got coordinates of each neighbourhoods Toronto from open source geojson file and create polygon object. After that I checked each yoga studio from my list and count it in certain neighbourhood, if its coordinated had been included into neighbourhood poligon. So that's how I got data about **count of yoga classes** across different neighbourhoods.

Then I got all nearest places from foursquare.com for each yoga studio and grouped them by each neighbourhood. That's how I got all **nearest objects to yoga classes** which could be used as surroundings.

Existing yoga classes across Toronto



Feature selection

Parameters checked for correlation with existing yoga studios

Latin American Restaurant	0.644522
Hotel Bar	0.574975
Arepa Restaurant	0.547231
Other Nightlife	0.528123
Lake	0.528123
Basketball Stadium	0.528123
Baseball Stadium	0.528123
Aquarium	0.528123
Roof Deck	0.528123
Street Art	0.519939

From this list selected 'Aquarium','Basketball Stadium','Lake','Baseball Stadium', as which were most nearest to exercises and other sport activities.

Male: 30 to 34 years	0.647936
Female: 25 to 29 years	0.630664
Female: 30 to 34 years	0.625783
Male: 25 to 29 years	0.611233
Male: 35 to 39 years	0.606367
Postsecondary diploma	0.562332
Employed	0.526583
Female: 35 to 39 years	0.494748
Male: 40 to 44 years	0.488849

From that list selected 'Male: 30 to 34 years','Female: 25 to 29 years','Female: 30 to 34 years','Male: 25 to 29 years','Male: 35 to 39 years' as most correlated

Methodology

My purpose was to find out, which neighbourhoods has lack of yoga studios. So first of all I had to to calculate model, which will count it. Because I had a lot of parameters, I use multi linear regression model.

Multilinear regression

In [56]:
```
X = ComposedData[['Aquarium','Basketball Stadium','Lake','Baseball Stadium','Male: 30 to 34 years','Female: 25 to 29 years','Female: 30 to 34 years','Male: 25 to 29 years','Male: 35 to 39 years']]
Y = ComposedData['VenueCount']
```

In [57]:
```
regr = linear_model.LinearRegression()
regr.fit(X, Y)
```

Out[57]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)
```

In [58]:
```
print('Intercept: \n', regr.intercept_)
print('Coefficients: \n', regr.coef_)
```

```
Intercept:
 0.0016202586839977147
Coefficients:
 [ 1.09156096  0.27289024  0.27289024  1.09156096  5.06191523  3.67004876
 -3.66992319 -4.43284991  0.13961876]
```

Results

```
ComposedData['ModelValues']=regr.predict(X)
ComposedData['res']=ComposedData['ModelValues']-ComposedData['VenueCount']
```

Let's select 2 Neighbourhoods, where difference between model value and real quantity of yoga centres is the biggest

```
ComposedData.sort_values(by=(['res']), ascending=False).head(2)
```

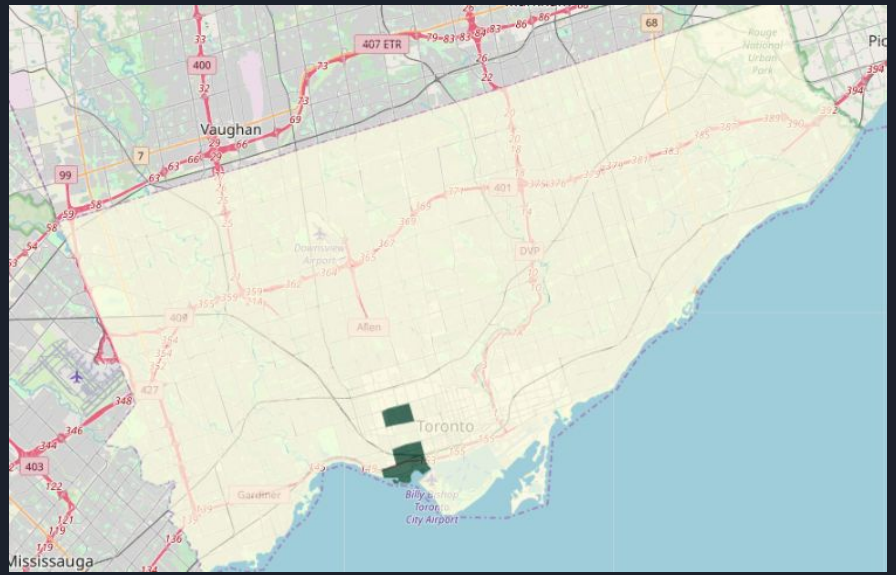
]:

	Aquarium	Basketball Stadium	Lake	Baseball Stadium	Male: 30 to 34 years	Female: 25 to 29 years	Female: 30 to 34 years	Male: 25 to 29 years	Male: 35 to 39 years	VenueCount	ModelValues	res
082	0.0	0.0	0.0	0.0	0.031223	0.030326	0.029220	0.028664	0.023334	0.0	0.026323	0.026323
080	0.0	0.0	0.0	0.0	0.008357	0.011089	0.008169	0.009130	0.006272	0.0	0.015126	0.015126

```
They are 82 and 80 neighbourhoods
```

Results at map

As we can see above, we got 2 neighbourhoods of Toronto, which have lack of yoga classes. It’s Niagara and Palmerston-Little Italy



Discussion

Of course, process of selecting new place for yoga studio can't be such simple. There are tons of additional parameters, such as:

- nearest transport hubs (ex. metro)
- nearest office buildings (which are absent in foursquare)
- rent price

But this research compare basic parameters to find out which of them are better and gain direction to find out main answer on questions, which this work based on.

Conclusion

In this study, I analyzed the relationship between demographic and nearest neighbours of certain yoga classes to amount of classes itself (depending of each neighbourhood of Toronto). Worked a lot with geo data and data manipulation. Of course, for big real projects this amount of work is not enough, cause in real - there are much more parameters and dependencies, but even this amount can show the way for appropriate decision process of opening new yoga class.