Major attractions nearby and Data Science jobs (in MA).

Date: 4 April 2019

Author: Nasir Ahmad

Publication: This article is purely based on personal research for non commercial usages. The services used by third party are credited without any intention of copy right violation.

Table of Contents

Table of Figures

Introduction

Data science jobs are available around the globe yet the hunt for job is not easy when you have too many places to choose from. In this article I will focus on the types of locations where Data Science jobs are available in Massachusetts US.

There are two approaches to get any job:

1. Look for a new job around your current location.
2. Search with keyword and then go through each job description.

Both of above methods fail to include one basic question: What kind of neighborhood is the company located in? Everyone knows the high-tech jobs are available in Silicon Valley, Seattle, New York, and Boston and so on. But what if you are not a big fan of living in a populous city, what if you just want a peaceful country side to live and code for living!

This article answers the following question:

“What are the major attractions nearby locations where Data Science jobs are being offered in Massachusetts?”

Data Section

Foursquare (F) is a platform which provides information about places in a given neighborhood. In this article Foursquare‘s listing is used to get popular sites in a given area. It provides developers with a good API which provides result in the form of JSON file, which developer can comprehend as per requirements.

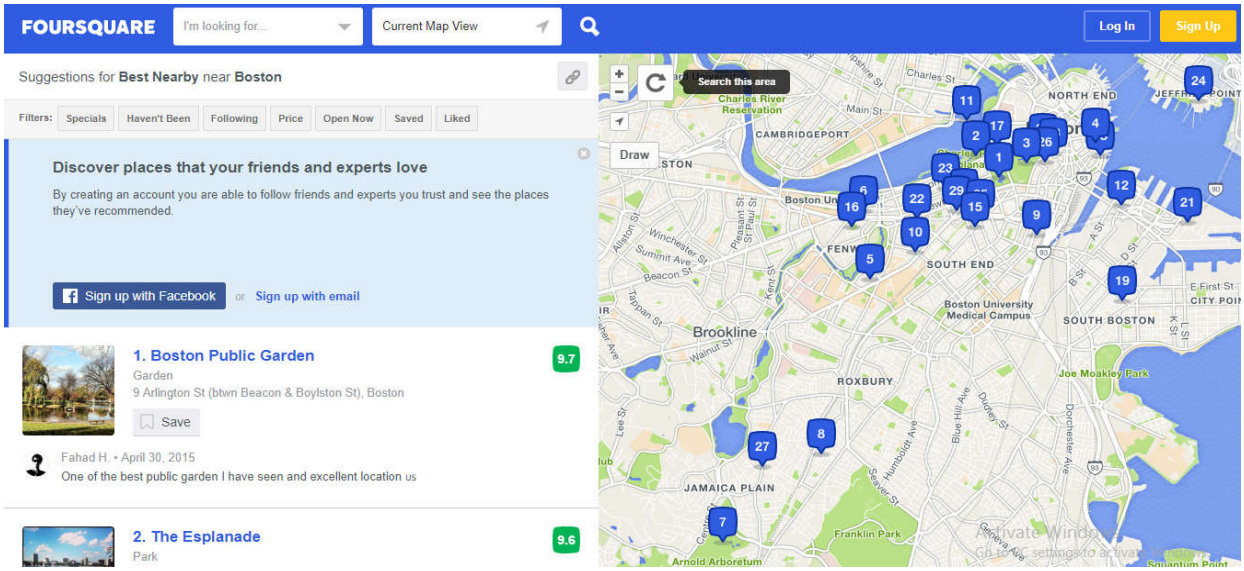

Figure 1 : Foursquare page of Boston city MA.

Adzuna (A) is a website which provides jobs listing for any given location in addition to other useful services related to jobs. This article uses Adzuna developer API for information related to Data Science jobs in Massachusetts.

Figure 2: Adzuna developer API user interface page.

Massachusetts State is divided into 14 counties (W) as shown in below image.

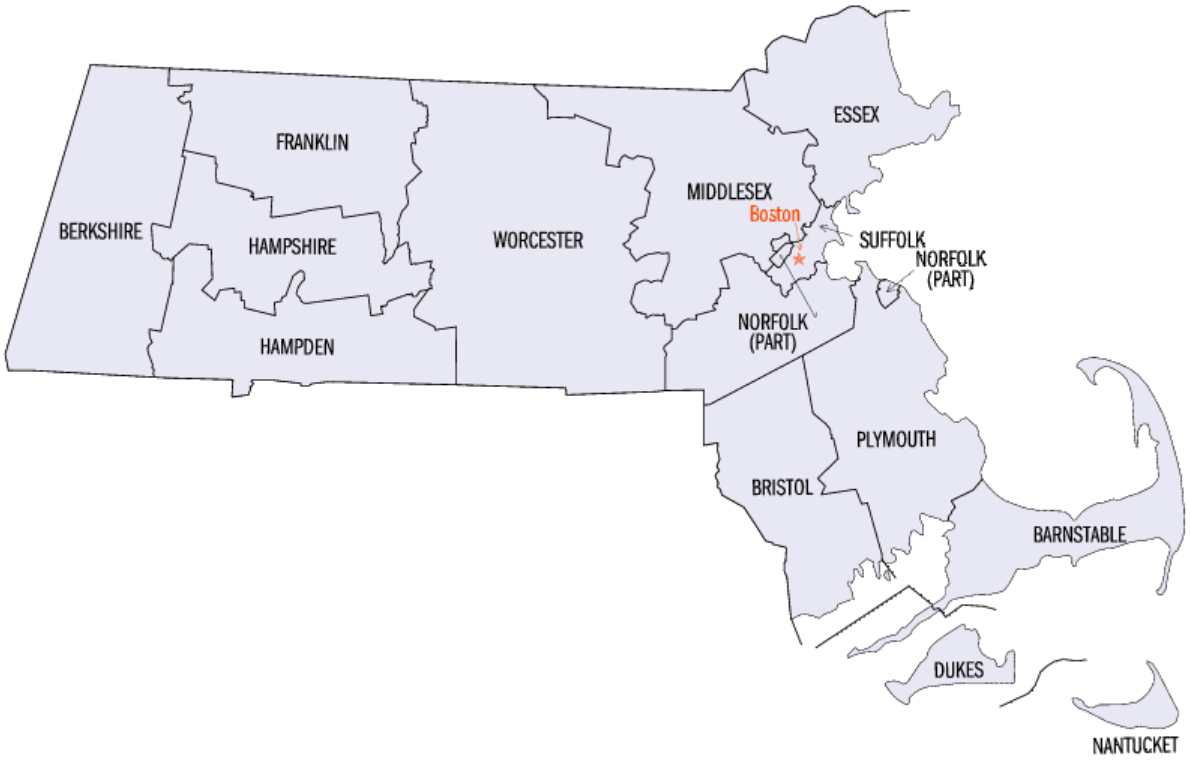

Figure 3 : Massachusetts state county list.

The first step is to find out Data science jobs available in each county so let's explore Adzuna API to get the list of jobs in Massachusetts. Below is part of JSON returned by Adzuna when input argument was set to Data Science jobs with location of Massachusetts.


```
{
  "locations": [
    {
      "location": {
        "__CLASS__": "Adzuna::API::Response::Location",
        "display_name": "Middlesex County, Massachusetts",
        "area": [
          "US",
          "Massachusetts",
          "Middlesex County"
        ]
      },
      "__CLASS__": "Adzuna::API::Response::LocationJobs",
      "count": 545S
    },
    {
      "count": 412,
      "__CLASS__": "Adzuna::API::Response::LocationJobs",
      "location": {
        "__CLASS__": "Adzuna::API::Response::Location",
```

```
        "area": [
          "US",
          "Massachusetts",
          "Suffolk County"
        ],
        "display_name": "Suffolk County, Massachusetts"
      }
    },

  "__CLASS__": "Adzuna::API::Response::JobGeoData",
  "location": {
    "__CLASS__": "Adzuna::API::Response::Location",
    "area": [
      "US",
      "Massachusetts"
    ],
    "display_name": "Massachusetts, US"
  }
}
```


Table 1: Adunza JSON file for Job Geo data

Transform the JSON file data into dataframe which looks like below:

[4]:

	__CLASS__	count	location.__CLASS__	location.area	location.display_name
0	Adzuna::API::Response::LocationJobs	463	Adzuna::API::Response::Location	[US, Massachusetts, Middlesex County]	Middlesex County, Massachusetts
1	Adzuna::API::Response::LocationJobs	356	Adzuna::API::Response::Location	[US, Massachusetts, Suffolk County]	Suffolk County, Massachusetts
2	Adzuna::API::Response::LocationJobs	54	Adzuna::API::Response::Location	[US, Massachusetts, Worcester County]	Worcester County, Massachusetts
3	Adzuna::API::Response::LocationJobs	39	Adzuna::API::Response::Location	[US, Massachusetts, Norfolk County]	Norfolk County, Massachusetts
4	Adzuna::API::Response::LocationJobs	13	Adzuna::API::Response::Location	[US, Massachusetts, Essex County]	Essex County, Massachusetts
5	Adzuna::API::Response::LocationJobs	7	Adzuna::API::Response::Location	[US, Massachusetts, Hampden County]	Hampden County, Massachusetts
6	Adzuna::API::Response::LocationJobs	7	Adzuna::API::Response::Location	[US, Massachusetts, Hampshire County]	Hampshire County, Massachusetts
7	Adzuna::API::Response::LocationJobs	5	Adzuna::API::Response::Location	[US, Massachusetts, Bristol County]	Bristol County, Massachusetts

Figure 4 : Dataframe of JSON file from Adzuna

[73]:

	job_count	area	Neighborhood	latitude	longitude
0	457	[US, Massachusetts, Middlesex County]	Middlesex County	42.485452	-71.396826
1	359	[US, Massachusetts, Suffolk County]	Suffolk County	42.354445	-70.978877
2	52	[US, Massachusetts, Worcester County]	Worcester County	42.365013	-71.958455
3	38	[US, Massachusetts, Norfolk County]	Norfolk County	42.153861	-71.182801
4	12	[US, Massachusetts, Essex County]	Essex County	42.629142	-70.866495
5	7	[US, Massachusetts, Hampden County]	Hampden County	42.172589	-72.629525
6	7	[US, Massachusetts, Hampshire County]	Hampshire County	42.369013	-72.713946
7	5	[US, Massachusetts, Bristol County]	Bristol County	41.742554	-71.085655
8	4	[US, Massachusetts, Plymouth County]	Plymouth County	41.942666	-70.761859
9	2	[US, Massachusetts, Franklin County]	Franklin County	42.518933	-72.561820
10	2	[US, Massachusetts, Nantucket]	Nantucket	41.265718	-70.068314

Figure 5 : DataFrame after cleaning data of Adzuna

Foursquare Data from API is fetched in the form of a JSON file, which for a single location is shown below:


```
[
 {
  'reasons':{
   'count':0,
   'items':[
    {
     'summary':'This spot is popular',
     'type':'general',
     'reasonName':'globalInteractionReason'
    }
   ]
  },
  'venue':{
   'id':'4ace1015f964a52025ce20e3',
   'name':'Verrill Farm',
   'location':{
    'address':'11 Wheeler Rd',
    'lat':42.42468135700807,
    'lng':-71.37407157515872,
    'labeledLatLngs':[
     {
      'label':'display',
      'lat':42.42468135700807,
      'lng':-71.37407157515872
     }
    ],
    'distance':2790,
    'postalCode':'01742',
    'cc':'US',
    'city':'Concord',
    'state':'MA',
    'country':'United States',
    'formattedAddress':[
     '11 Wheeler Rd',
     'Concord, MA 01742',
     'United States'
    ]
   },
   'categories':[
    {
     'id':'4bf58dd8d48988d15b941735',
     'name':'Farm',
     'pluralName':'Farms',
```

```
            'shortName':'Farm',
            'icon':{
              'prefix':'https://ss3.4sqi.net/img/categories_v2/parks_outdoors/farm_',
              'suffix':'.png'
            },
            'primary':True
          }
        ],
        'photos':{
          'count':0,
          'groups':[

          ]
        }
      },
      'referralId':'e-0-4ace1015f964a52025ce20e3-0'
    }
]
```


Table 2 : Foursquare JSON file

Cleaning the data from Foursquare gives us Venue detail as shown in below data frame:

[92]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Middlesex County	42.485452	-71.396826	Trader Joe's	42.482846	-71.414835	Grocery Store
1	Middlesex County	42.485452	-71.396826	Colonial Spirits of Acton	42.478297	-71.411557	Liquor Store
2	Middlesex County	42.485452	-71.396826	Nashoba Brook Bakery	42.458521	-71.396562	Bakery
3	Middlesex County	42.485452	-71.396826	Reasons To Be Cheerful	42.457451	-71.395757	Ice Cream Shop
4	Middlesex County	42.485452	-71.396826	Woods Hill Table	42.456593	-71.393105	New American Restaurant

Figure 6 : Foursquare Venue DataFrame

Methodology

Given the geographic data our objective is to find out major attraction points nearby Data Science jobs so we need to know which county are similar to each other. With this approach we can then make it easy to figure out the lifestyle at a given county for Data Scientist.

K-means (KM) clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K* groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the *K*-means clustering algorithm are:

1. The centroids of the *K* clusters, which can be used to label new data
2. Labels for the training data (each data point is assigned to a single cluster)

Rather than defining groups before looking at the data, clustering allows you to find and analyze the groups that have formed organically. The "Choosing K" section below describes how the number of groups can be determined.

Thus this article will be using K means to find the clustering in between the various county which offer Data Science jobs.

Let's start with some **exploratory data analysis**:

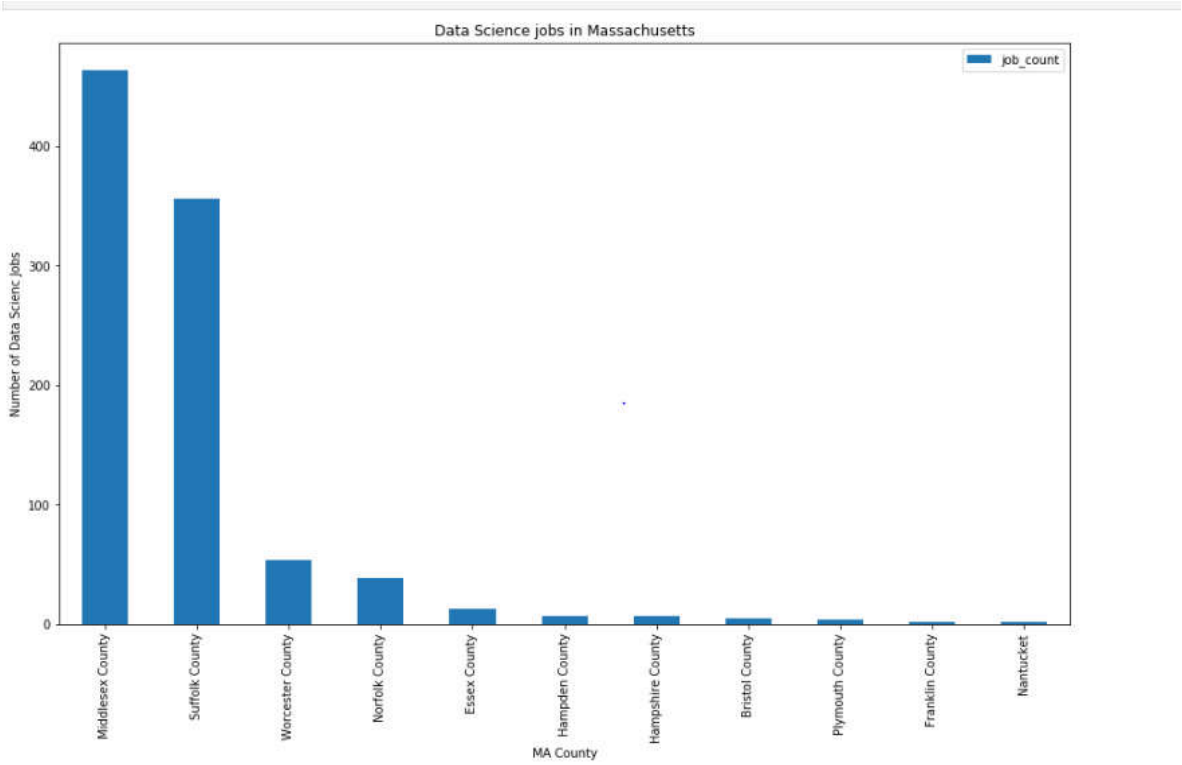

Figure 7 : Data Science Jobs in Massachusetts State with respect to County

The above figure shows that major Data Science jobs are being offered in two states (Middlesex and Suffolk). It is a bar chart with county on the X-axis and number of jobs on the Y-axis. The rest of county offer few Data Science jobs but let's visualize their part:

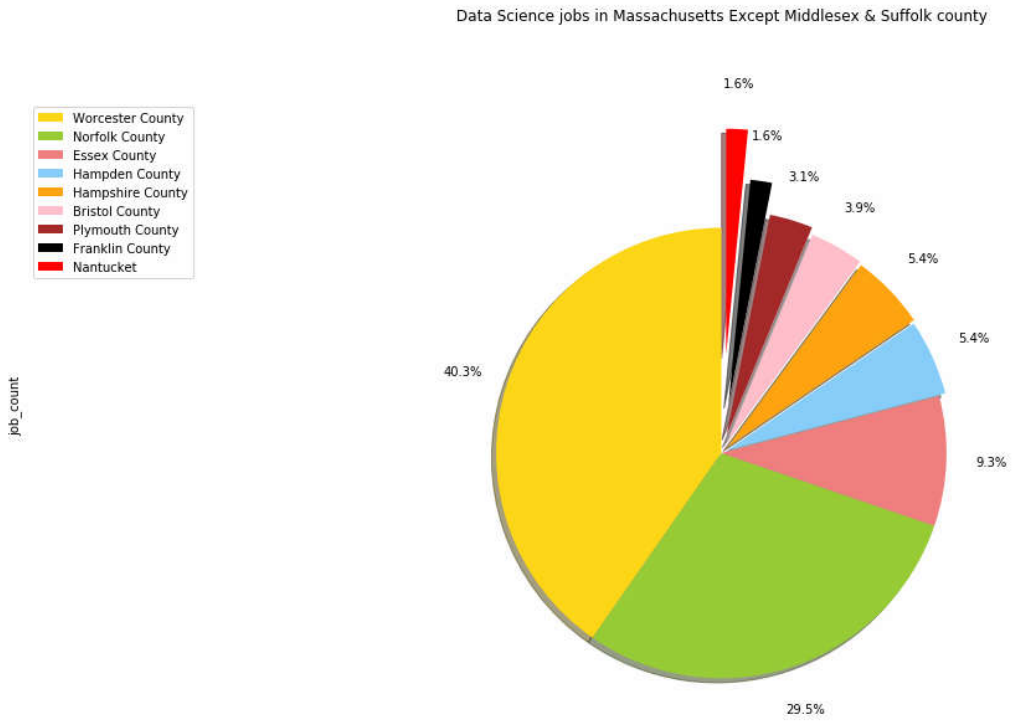

Figure 8 : Pie Chart of Data Science jobs excluding Middlesex and Suffolk

The above pie chart shows that in Data science jobs count Worcester and Norfolk county take major portion of the jobs. Although all these county data science jobs combined are still not competitive to Middlesex or Suffolk.

With K = 5 the map shows like this

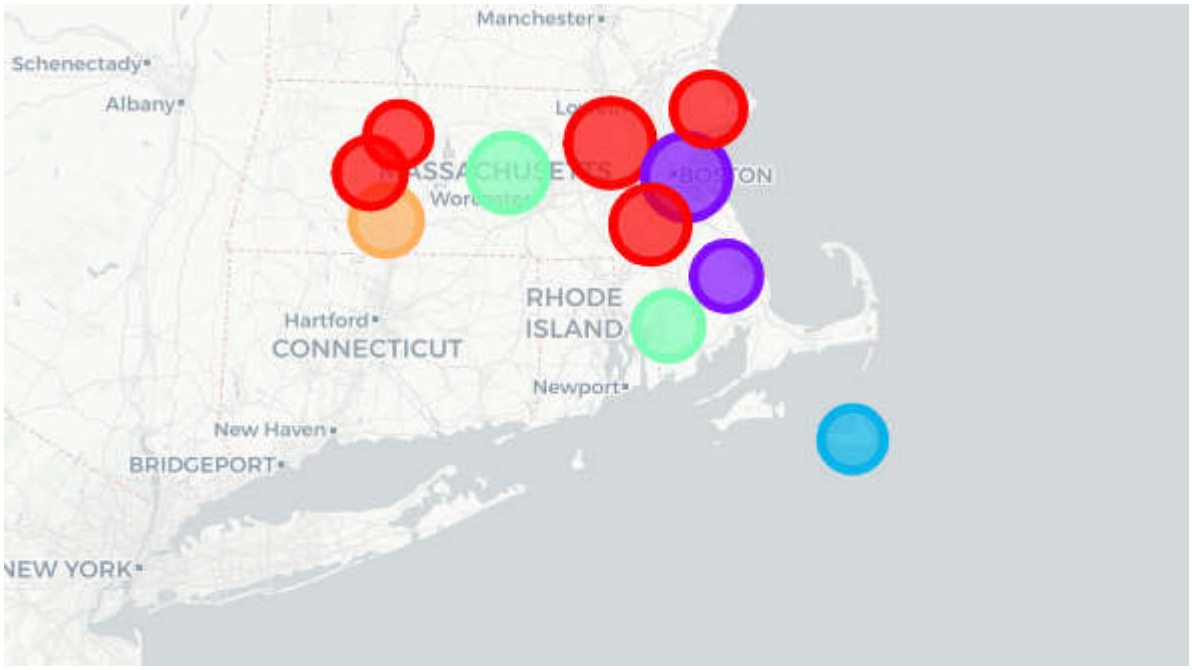

Figure 9 : K =5 clusters on the map

The major attractions are listed below

[29]:

	job_count	area	Neighborhood	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	458	[US, Massachusetts, Middlesex County]	Middlesex County	42.485452	-71.396826	0	Ice Cream Shop	Chinese Restaurant	Grocery Store	American Restaurant	History Museum	Sandwich Place	Café	Historic Site	Coffee Shop	Park
1	365	[US, Massachusetts, Suffolk County]	Suffolk County	42.354445	-70.978877	1	Coffee Shop	Donut Shop	Seafood Restaurant	Airport Lounge	Italian Restaurant	Café	Park	Electronics Store	Pizza Place	American Restaurant
2	52	[US, Massachusetts, Worcester County]	Worcester County	42.365013	-71.958455	3	Donut Shop	American Restaurant	Pizza Place	Trail	Diner	Campground	Pharmacy	Golf Course	Restaurant	Food & Drink Shop
3	38	[US, Massachusetts, Norfolk County]	Norfolk County	42.153861	-71.182801	0	Coffee Shop	Chinese Restaurant	Breakfast Spot	Gym	Mexican Restaurant	Italian Restaurant	Café	Thai Restaurant	Donut Shop	Japanese Restaurant
4	12	[US, Massachusetts, Essex County]	Essex County	42.629142	-70.866495	0	American Restaurant	Italian Restaurant	Seafood Restaurant	Chinese Restaurant	Ice Cream Shop	Sandwich Place	Coffee Shop	Farm	Golf Course	Restaurant
5	7	[US, Massachusetts, Hampden County]	Hampden County	42.172589	-72.629525	4	American Restaurant	Clothing Store	Coffee Shop	Pizza Place	Pharmacy	Furniture / Home Store	Ice Cream Shop	Department Store	Bakery	Cosmetics Shop

Figure 10 : K = 5 clusters top common attraction points

As the figure shows that major attraction points are similar so we need to reduce the number of clusters so that we can get county that have much more differences.

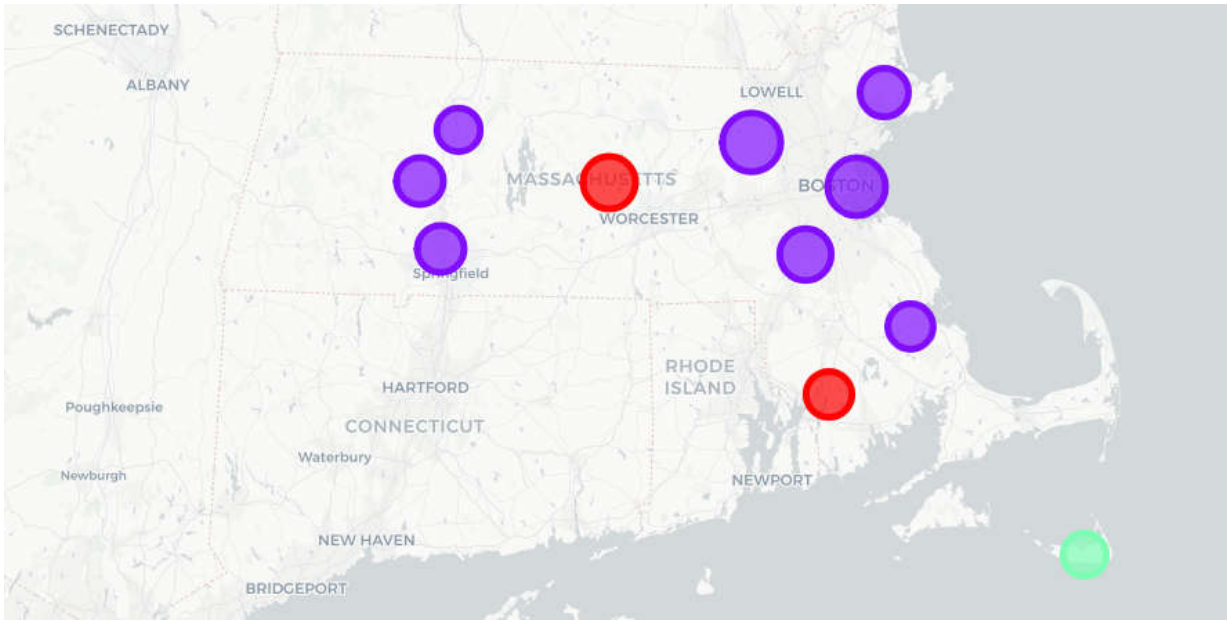

Figure 11 : K=3 clusters on the map

[61]:

	job_count	Neighborhood	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	458	Middlesex County	42.485452	-71.396826	1	Ice Cream Shop	Chinese Restaurant	Grocery Store	American Restaurant	History Museum	Sandwich Place	Café	Historic Site	Coffee Shop	Park
1	365	Suffolk County	42.354445	-70.978877	1	Coffee Shop	Donut Shop	Seafood Restaurant	Airport Lounge	Italian Restaurant	Café	Park	Electronics Store	Pizza Place	American Restaurant
2	52	Worcester County	42.365013	-71.958455	0	Donut Shop	American Restaurant	Pizza Place	Trail	Diner	Campground	Pharmacy	Golf Course	Restaurant	Food & Drink Shop
3	38	Norfolk County	42.153861	-71.182801	1	Coffee Shop	Chinese Restaurant	Breakfast Spot	Gym	Mexican Restaurant	Italian Restaurant	Café	Thai Restaurant	Donut Shop	Japanese Restaurant
4	12	Essex County	42.629142	-70.866495	1	American Restaurant	Italian Restaurant	Seafood Restaurant	Chinese Restaurant	Ice Cream Shop	Sandwich Place	Coffee Shop	Farm	Golf Course	Restaurant
5	7	Hampden County	42.172589	-72.629525	1	American Restaurant	Clothing Store	Coffee Shop	Pizza Place	Pharmacy	Furniture / Home Store	Ice Cream Shop	Department Store	Bakery	Cosmetics Shop
6	7	Hampshire County	42.369013	-72.713946	1	Pizza Place	American Restaurant	Vegetarian / Vegan Restaurant	Bar	Breakfast Spot	Ice Cream Shop	Grocery Store	Brewery	Café	Italian Restaurant
7	5	Bristol County	41.742554	-71.085655	0	Donut Shop	Restaurant	Breakfast Spot	American Restaurant	Bar	Ice Cream Shop	Bakery	Seafood Restaurant	Pizza Place	Sandwich Place
8	4	Plymouth County	41.942666	-70.761859	1	Italian Restaurant	Seafood Restaurant	Coffee Shop	Sandwich Place	Donut Shop	Convenience Store	Pub	Bar	Pizza Place	American Restaurant
9	2	Franklin County	42.518933	-72.561820	1	Pizza Place	American Restaurant	Convenience Store	Sandwich Place	Gift Shop	Bar	Brewery	Donut Shop	Discount Store	Diner
10	2	Nantucket	41.265718	-70.068314	2	Beach	American Restaurant	Seafood Restaurant	Hotel	Golf Course	Italian Restaurant	French Restaurant	Bookstore	Café	Lighthouse

Activate Windows
Go to PC settings to activate Windows.

Figure 12 : K = 3 clusters top common things

Results:

Looking at above data we can draw few lines between each cluster. Let's explore each cluster in detail

Red Cluster (number 0): If you are a fan of Donuts then definitely this is your place to jump in, but hold on! One can't eat sweets all the times. This cluster offers good American restaurants and Pizza places as answer to this question.

Purple Cluster (number 1): This cluster has listed its top attraction points as Italian restaurants, Chinese restaurants, coffee shops, shopping places.

Green Cluster (number 2): This cluster offers Beach as primary site seeing location which implies people like outdoor activities in this place. Other places include restaurants, hotels and golf course.

Discussion:

This article discussed the major Data Science jobs locations in Massachusetts State by looking at the current job listing (March-April 2019). It is important to mention that there are so many parameters to choose from and multiple sources of data available online so this article has limited the scope for learning purposes.

Data science is an emerging field and is slowly trending to new locations, which is very important for small towns, med size businesses and untouched horizons.

Conclusion:

In the state of Massachusetts the major attraction points are restaurants, pizza place, coffee/donut shops, shopping malls, parks, museums, hotels, ice cream shops, beaches, golf course, bar, trail. The concentration of each venue is different in each county and jobs are advertised all over these locations. I have provided few stats and figures to help you out with choosing your place of work.

One of the reason I performed this analysis is that personally I believe in work-home balance, so don't ignore one for the other.

Works Cited

(n.d.). Retrieved from https://foursquare.com/city-guide

(n.d.). Retrieved from https://www.adzuna.com/

(n.d.). Retrieved from https://en.wikipedia.org/wiki/List_of_counties_in_Massachusetts

(n.d.). Retrieved from https://www.datascience.com/blog/k-means-clustering