Best Neighborhoods for Housing - Massachusetts, USA

SURESH KUMAR RANGASAMY

Introduction

Massachusetts has 14 Counties, 39 Cities and 312 Towns. Mass is well endowed with neighborhoods that are active, vibrant, progressive, thoughtful, diverse, fun, friendly, filled with natural beauty and historic charm. This abundance in choice increases the complexity for home buyers in their decision-making process. While buying a home, the neighborhood matters the most, sometimes even higher than the home itself. With the help of Data Science, we are attempting to help "Young Home Buyers with Kids" to choose a neighborhood that supports an active lifestyle, surrounded by natural beauty, good education, safety, well connected public transport system based on the prevailing Median Pricing. The home buyer can then shop for homes within the neighborhood.

Target Audience & Objective

Our Target Audience:

Home Buyers - Young Family with Toddlers/Kids

Our objective is to perform a detailed exploratory analysis and recommend a neighborhood by finding answers to the below:

1. Which Neighborhood has the most Parks & Recreation avenues for a young family with toddlers/kids?

2. Which Neighborhood has the most Top-Rated Schools?

3. Which Neighborhood has the best transit score for easy commute?

4. Which Neighborhood is safest with the least Crime Rate?

5. Suggest Top 50 Best Neighborhoods in Massachusetts for a "Young Family with Kids" to own a home.

Why it is important?:

Massachusetts has multiple culturally rich, diverse, vibrant neighborhoods with varying home prices, often overwhelming and complicating the home buyer's decision making process. Our system attempts to simplify the process by recommending the Best 50 Neighborhoods in MA for a "Young Family with Kids" to own a home.

Data Sourcing

1. We would need a list of Median Home Sale Prices of MA nieghborhoods (City/Zip Code level). Data can be sourced from RedFin, a National Real Estate Brokerage, https://www.redfin.com/blog/data-center/
Data Points - Zip Code, City, County, Median Sale Price.
For E.g.: 02148, Malden, 500000

2. We would need neighborhood specific information like Supermarkets, Trails, Parks, Water Fronts, Restaurants, Coffee Shops, Gyms etc. Data can be sourced with help of FOURSQUARE Developer Apps, https://foursquare.com/developers/apps
Data Points - Zip Code, City, Lattitue, Longitude, Venues, Venue Categories
Using Get_Venue API we can source the list of venues and venue category information around 5 KM radius of each Neighborhood.
For E.g.: 01906, Saugus, -71.0110, 42.4651, Breakheart Reservation, Trail

3. We would need list of public schools in a neighborhood and their progress and performance index information. Data can be sourced for the year 2017 from https://www.kaggle.com/ndalziel/massachusetts-public-schools-data
Data Points - Zip Code, City, County, School, School PPI etc.
For E.g.: 02478, Belmont, Middlesex, Belmont High, 100

4. We would need transit scores, walkability scores, bikeability scores of a neighborhood
Data can be sourced from https://www.walkscore.com/MA
Data Points - Zip Code, City, Walk Score, Transit Score, Bike Score etc.
For E.g.: 02138, Cambridge, 88, 74, 96

Data Sourcing…

5. We need crime statistics for Mass by city or zip code
Data can be sourced from https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/downloads/download-printable-files
Data Points - City, Population, No. of Violent Crimes, No. of Property Crimes etc.
For E.g.: Hoyolke, 40470, 191, 1640

6. We would need list of neighborhoods in MA with 1 set of latitude and longitude coordinates
We can source it from: https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=MA
Data Points - Zip Code, City, Latitude and Longitude
For E.g.: 01701, Framingham, -71.4162, 42.2793

7. We would need Zip Codes to City to County crossreference mapping
Data can be sourced from https://www.unitedstateszipcodes.org/ma/#zips-list
Data Points - Zip Code, City, County
For E.g.: 01020, Springfield, Hampden

8. We would need GeoJSON GeoSpatial Coordinates data of City boundaries for maps, choropleths. Data can be sourced from http://maps-massgis.opendata.arcgis.com/datasets/43664de869ca4b06a322c429473c65e5_0.geojson
Data Points - City, GeoSpatial Lat, Long coordinates define polygonal shape boundaries of cities
For E.g.: QUINCY, (-70.987520657600399, 42.304533090819092), (-70.987515724246094, 42.304588167507383), (-70.98749032019883, 42.304654946893386), ..., (-70.960227444070043, 42.29578425180248)

Methodology – Foursquare API

Source Venue & Category data around a Neighborhood. Derive Counts.

*** Derive below Data Points and WRITE into a Data Frame.**

Recreation Count = Total Number of Venues with Venue Category containing text like Trail or River or Beach or Waterfront or Park

Fitness Count = Total Number of Venues with Venue Category containing text like Gym or Fitness

Eateries Count = Total Number of Venues with Venue Category containing text like Restaurant or Bakery or Coffee

Groceries Count = Total Number of Venues with Venue Category containing text like Supermarket or Grocer

Hospitals Count = Total Number of Venues with Venue Category containing text like Hospital

[86]:
```
import re
recreation_count=0
fitness_count=0
hospital_count=0
eateries_count=0
groceries_count=0
column_names=['Zip','Neighborhood', 'recreation_count','fitness_count','hospital_count','eateries_count','groceries_count']
venue_details=pd.DataFrame(columns=column_names)

for row in Geospatial_Coordinates.values.tolist():
    Zip, City, State, Latitude, Longitude=row
    venues = get_venues(Zip, City, Latitude, Longitude)
    #venue_with_trails=venues[venues['Category']=='Trail']
    recreation_count = str(np.sum(venues['Category'].str.contains('Trail|River|Beach|Waterfront|Park',flags=re.IGNORECASE, regex=True)))
    fitness_count = str(np.sum(venues['Category'].str.contains('Gym|Fitness',flags=re.IGNORECASE, regex=True)))
    hospital_count = str(np.sum(venues['Category'].str.contains('Hospital',flags=re.IGNORECASE, regex=False)))
    eateries_count = str(np.sum(venues['Category'].str.contains('Restaurant|Bakery',flags=re.IGNORECASE, regex=True)))
    groceries_count = str(np.sum(venues['Category'].str.contains('Supermarket',flags=re.IGNORECASE, regex=False)))
    venue_details = venue_details.append({'Zip':Zip,'Neighborhood': City,
                                          'recreation_count': recreation_count, 'fitness_count' : fitness_count,
                                          'hospital_count' : hospital_count, 'eateries_count' : eateries_count,
                                          'groceries_count' : groceries_count
                                         }, ignore_index=True)
venue_details
```

[86]:

	Zip	Neighborhood	recreation_count	fitness_count	hospital_count	eateries_count	groceries_count
0	02351	Abington	2	3	0	16	1
1	02018	Accord	6	1	0	22	0
2	01720	Acton	2	5	0	22	2

Methodology – Correlation Matrix

Correlation Matrix.

Create a Correlation Matrix for Data Exploration

```
[9]: venue_details.corr()
```

[9]:

	recreation_count	fitness_count	eateries_count	groceries_count	Latitude	Longitude	Median Sale Price	Walk Score	Transit Score	Bike Score	Total_crimes	TopSchoolCount
recreation_count	1.000000	0.090716	0.380197	-0.108236	-0.098852	0.537474	0.318862	0.093013	0.119780	0.069122	0.194718	0.049518
fitness_count	0.090716	1.000000	0.576378	0.134262	0.263200	0.000356	0.239613	0.265656	0.258772	0.272963	0.241690	0.289822
eateries_count	0.380197	0.576378	1.000000	0.152317	0.125538	0.196289	0.233465	0.333394	0.304205	0.325789	0.353826	0.327294
groceries_count	-0.108236	0.134262	0.152317	1.000000	0.088881	0.000966	-0.148630	0.150775	0.117516	0.137926	-0.127797	0.147417
Latitude	-0.098852	0.263200	0.125538	0.088881	1.000000	-0.468261	0.022505	0.177948	0.134030	0.163914	0.103056	0.223001
Longitude	0.537474	0.000356	0.196289	0.000966	-0.468261	1.000000	0.234077	0.015013	0.032024	0.018465	0.030226	-0.020626
Median Sale Price	0.318862	0.239613	0.233465	-0.148630	0.022505	0.234077	1.000000	0.032230	0.048884	0.052284	0.245400	0.123493
Walk Score	0.093013	0.265656	0.333394	0.150775	0.177948	0.015013	0.032230	1.000000	0.897708	0.977337	-0.008212	0.384302
Transit Score	0.119780	0.258772	0.304205	0.117516	0.134030	0.032024	0.048884	0.897708	1.000000	0.876366	0.025269	0.356251
Bike Score	0.069122	0.272963	0.325789	0.137926	0.163914	0.018465	0.052284	0.977337	0.876366	1.000000	-0.015562	0.393654
Total_crimes	0.194718	0.241690	0.353826	-0.127797	0.103056	0.030226	0.245400	-0.008212	0.025269	-0.015562	1.000000	0.004625
TopSchoolCount	0.049518	0.289822	0.327294	0.147417	0.223001	-0.020626	0.123493	0.384302	0.356251	0.393654	0.004625	1.000000

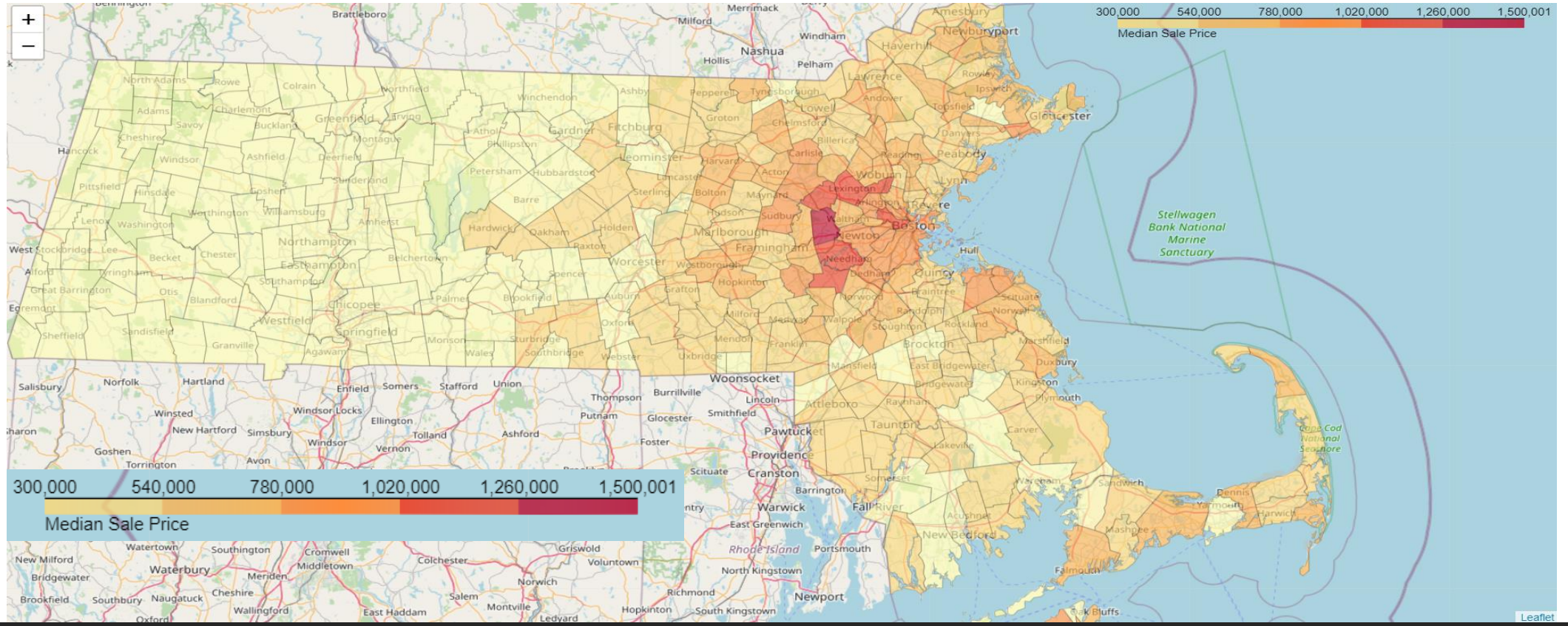
Correlation Observations:

We simply observed the correaltion patterns between all features. As expected, There were no significant coorelations. For E.g. Home Prices, We have ONLY sourced the Median Home Prices rolled up at a Neighborhood level, NOT the individual Home Prices. The individual home prices may depend upon other features like Age of Home, Lot Size, SQFT Area, # of Bed Rooms, # of Bathrooms etc. So we decided to take the below parameters for our Exploratory Analysis and Recommendation.

1. Top School Count within a Neighborhood - Higher the better
2. Recreation Count within a Neighborhood - Higher the better
3. Transit Score within a Neighborhood - Higher the better
4. Crime Stats within a Neighborhood - Lower the better

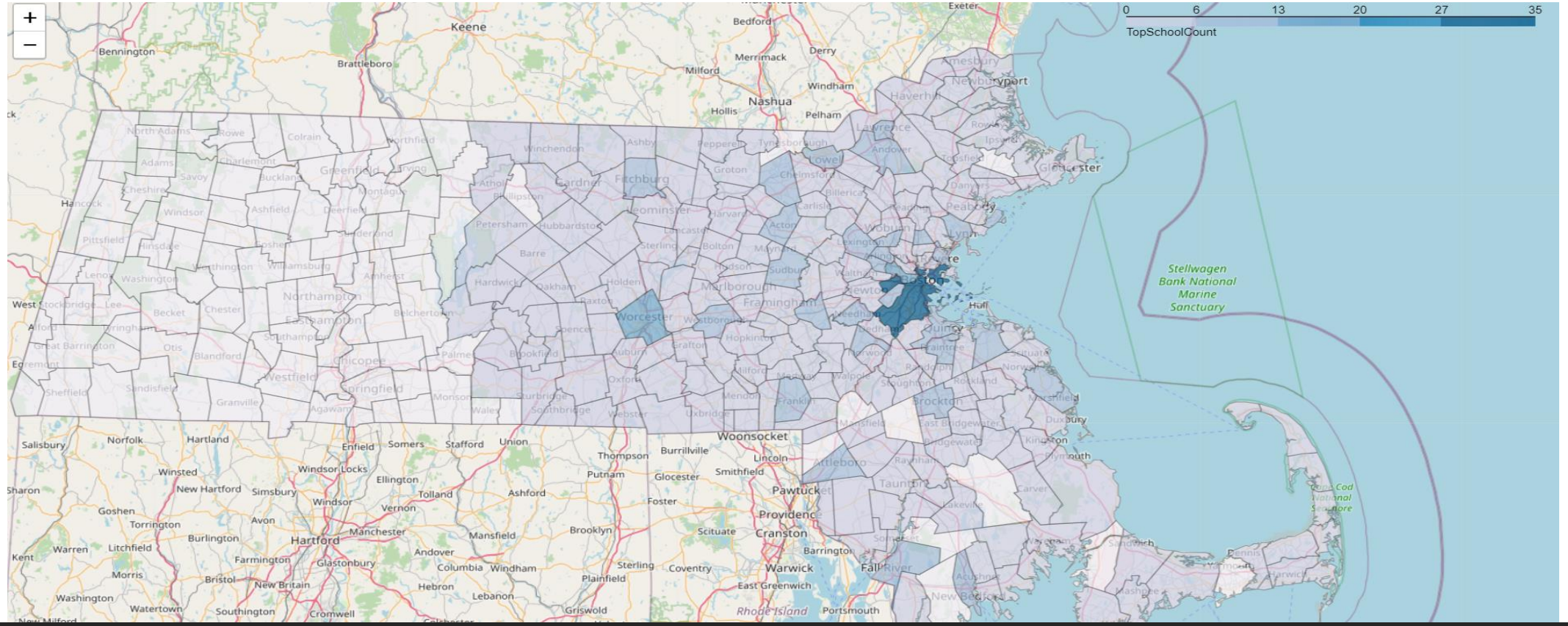
Methodology – Choropleths

Plot Median Home Prices by City



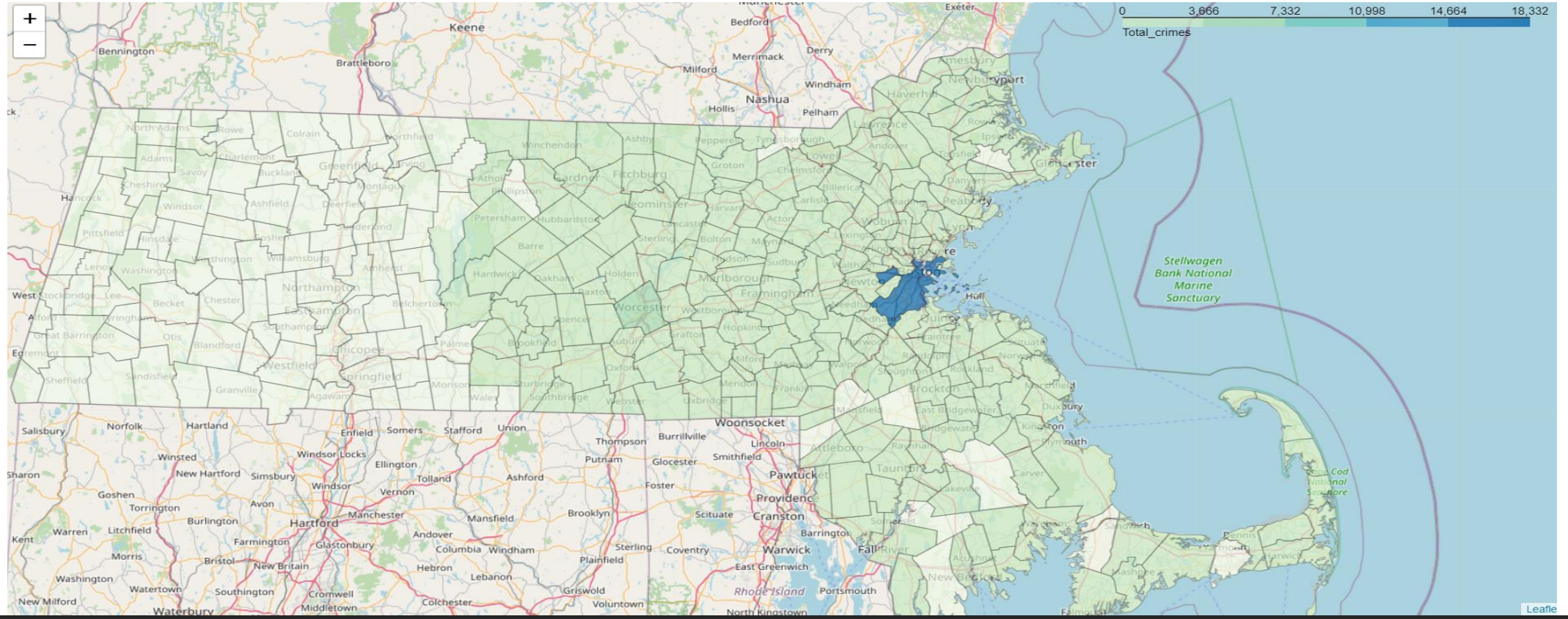
Methodology – Choropleths

Plot Number of Top Rated Schools by City



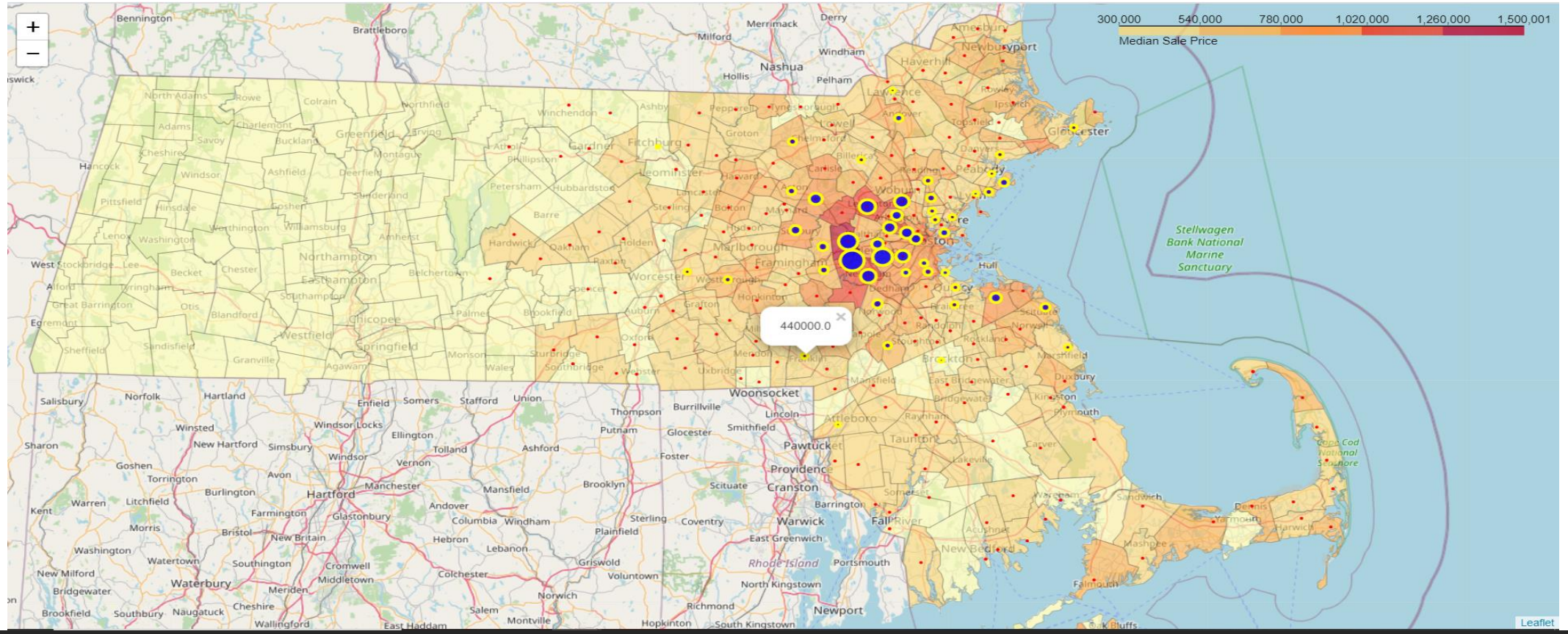
Methodology – Choropleths

Plot Number of Crimes by City



Methodology – Choropleths

Plot the Best 50 Neighborhoods in Massachusetts to own a home for a Young Family with Kids



Discussion & Conclusion

Which Neighborhood has the most Parks & Recreation avenues for a young family with toddlers/kids?

The Top 10 Neighborhoods with most recreation avenues:

Nahant
Brewster
Lynn
Orleans
Prides Crossing
Swampscott
North Eastham
West Dennis
North Chatham
Marblehead

Discussion & Conclusion…

Which Neighborhood has the most Best Public Schools?

The Top 10 Neighborhoods with best public schools:

Braintree
Boston
Franklin
Natick
Winchester
Attleboro
Andover
Westford
Westwood
Revere

Discussion & Conclusion…

Which Neighborhood has the best transit score for easy commute?

The Top 10 Neighborhoods with best transit scores:

Cambridge

Boston

Brookline

Somerville

Chelsea

Revere

Malden

Medford

Everett

Quincy

Discussion & Conclusion…

Which Neighborhood is safest with the least Crime Rate?

The Top 10 Neighborhoods with least crime rates:

Wayland
New Braintree
Princeton
Chilmark
Dover
Nahant
Oakham
Boylston
Wenham
Paxton

Discussion & Conclusion…

Suggest Top 50 Best Neighborhoods in Massachusetts for a "Young Family with Kids" to own a home.

The Top 50 Neighborhoods in MA for "Young Family with Kids" to own a Home are:

Braintree, Boston, Natick, Franklin, Winchester, Attleboro, Westwood, Andover, Westford, Revere, Newton Center, Belmont, Melrose, Fitchburg, Marblehead, Hingham, Marshfield, Lexington, Wellesley Hills, Acton, Sudbury, Westborough, Everett, Lynn, Lawrence, Brockton, Salem, Gloucester, Wakefield, Quincy, Newtonville, Cambridge, Concord, Weston, Scituate, Sharon, Billerica, Shrewsbury, Cambridge, Boston, Malden, Quincy, Arlington, Needham, Swampscott, Brookline, Boston, Beverly, Roslindale & Wayland.