

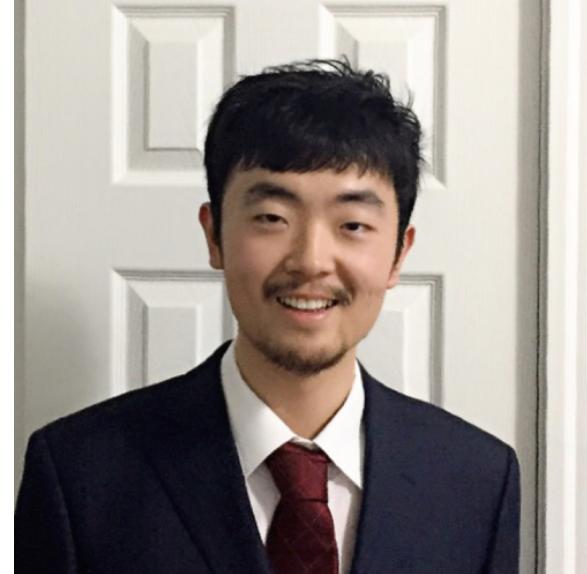
Final Project Report

Boston Real Estate Rents Analysis

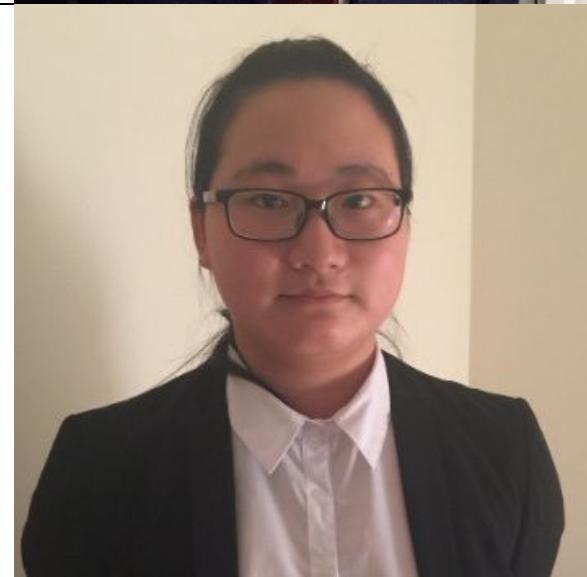
INFO7390 Advances Data Science/Architecture

Team 4

Jianxing Lu



Qiaomin Ling



Wenjin Cao



CATALOG

Project Description -----	4
Three Kinds of Dataset Sources -----	4
Deploying Web Services in Azure Studio -----	6
Using Cloud Storage Service in Azure Portal -----	11
Visualizing Monthly Rents in Power BI -----	13
Demo -----	14

1. Description & Business Case

In our project, we can predict rents in Boston based on different arguments such as number of bedrooms/bathrooms, building type, location (zip code, public transportation), etc.

We also did a vivid and interactive map-based visualization and analysis of Boston rents in accordance with specific neighborhoods and dates.

2. Data Resources

(1) Zillow Data

<http://www.zillow.com/research/data/#rental-data>

We used the monthly rental data of different neighborhoods in Boston from 2013/01 to 2016/06.

Parameters: Median rent (per neighborhood)

RegionName	X2015.01	X2015.02	X2015.03	X2015.04	X2015.05	X2015.06	X2015.07	X2015.08	X2015.09	X2015.10	X2015.11	X2015.12
Jamaica Plain	1325.0	1337.5	1197.5	1275.0	1262.5	1275	1300.0	1400	1425.0	1475.0	1625.0	1547.5
Brighton	1375.0	1397.5	1395.0	1395.0	1400.0	1395	1450.0	1425	1425.0	1400.0	1395.0	1410.0
South Boston	2250.0	2400.0	2299.5	2321.0	2348.0	2325	2405.5	2400	2375.0	2375.0	2302.5	2278.0
South End	1862.5	1950.0	1750.0	1750.0	1725.0	1675	1750.0	1800	1950.0	2000.0	2000.0	1995.0
Allston	1395.0	1400.0	1395.0	1395.0	1400.0	1400	1400.0	1400	1395.0	1395.0	1395.0	1395.0
Fenway	1650.0	1625.0	1650.0	1650.0	1650.0	1650	1675.0	1655	1650.0	1625.0	1645.0	1645.0
Back Bay	1850.0	1800.0	1725.0	1750.0	1700.0	1745	1750.0	1795	1800.0	1800.0	1800.0	1750.0
Kenmore	1550.0	1550.0	1600.0	1600.0	1625.0	1625	1625.0	1610	1595.0	1575.0	1600.0	1600.0
Mission Hill	1300.0	1412.5	1450.0	1484.5	1495.0	1400	1400.0	1450	1510.0	1550.0	1400.0	1450.0
North End	1850.0	1850.0	1800.0	1775.0	1787.5	1700	1750.0	1800	1700.0	1799.0	1900.0	1900.0
Beacon Hill	1750.0	1750.0	1750.0	1750.0	1750.0	1750	1700.0	1700	1725.0	1797.5	1800.0	1800.0
West End	2000.0	2100.0	2100.0	2047.5	2100.0	2050	2200.0	2240	2267.5	2275.0	2327.5	2245.0

Those values are median rents for each Boston neighborhoods (21) of different building types, in this case, different number of bedrooms (studio, 1bed, 2bed, 3bed, 4bed, 5bed).

(2) Rental Data from an Individual Third Party

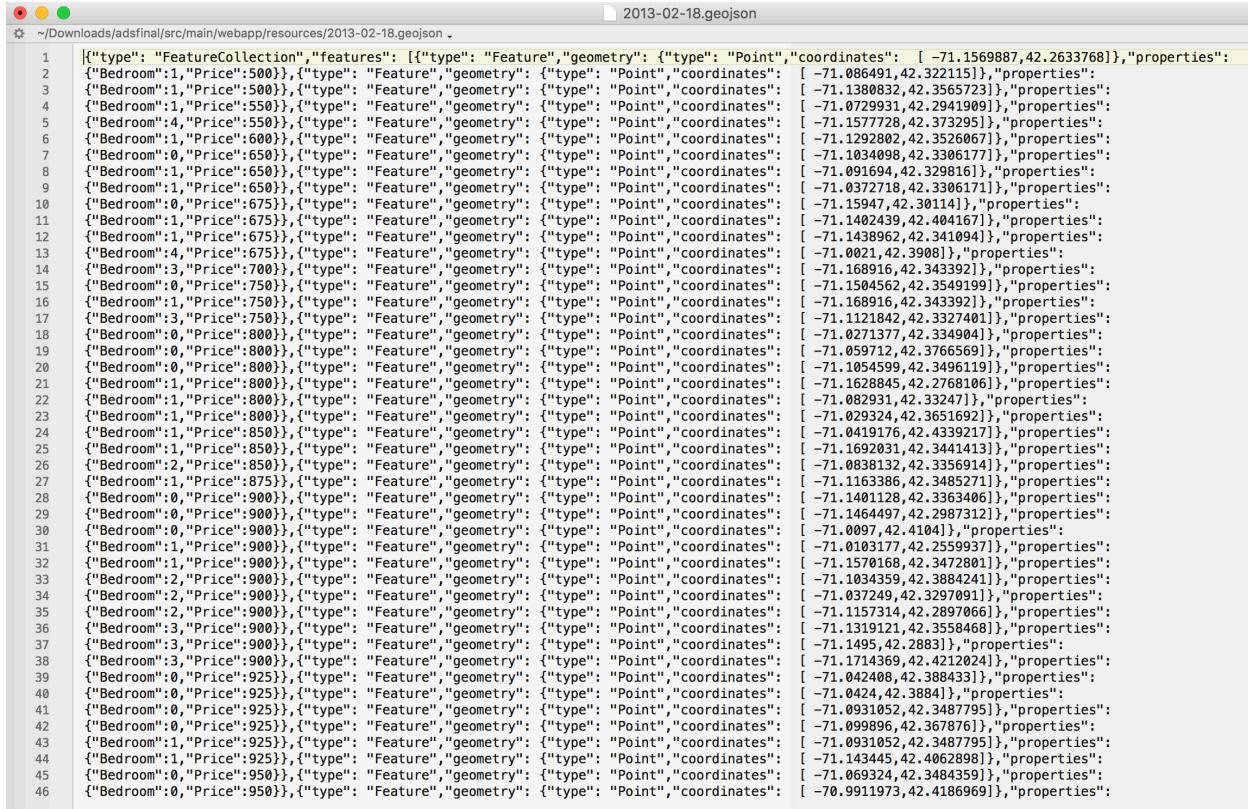
http://www.jefftk.com/apartment_prices/data-listing

(Credit to this cool guy: <http://www.jefftk.com>)

We used the monthly rental data of different houses/apartments in Boston from 2013/02 to 2016/07.

Parameters: Longitude and latitude of the house, number of bedrooms, rent (per house/apt).

We transferred the .csv files to .geojson files and stored them using Azure Storage service.



```

1  [{"type": "FeatureCollection", "features": [{"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1569887, 42.2633768]}, "properties": 2  {"Bedroom":1,"Price":500}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.086491, 42.322115]}, "properties": 3  {"Bedroom":1,"Price":500}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1380832, 42.3565723]}, "properties": 4  {"Bedroom":1,"Price":550}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0729931, 42.2941909]}, "properties": 5  {"Bedroom":4,"Price":550}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1577728, 42.3732951]}, "properties": 6  {"Bedroom":1,"Price":600}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1292802, 42.3526667]}, "properties": 7  {"Bedroom":0,"Price":650}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1034098, 42.3306177]}, "properties": 8  {"Bedroom":1,"Price":650}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.091694, 42.3298161]}, "properties": 9  {"Bedroom":1,"Price":650}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0372718, 42.3306171]}, "properties": 10 {"Bedroom":0,"Price":675}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.15947, 42.30114]}, "properties": 11 {"Bedroom":1,"Price":675}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1402439, 42.4041671]}, "properties": 12 {"Bedroom":1,"Price":675}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1438962, 42.3410941]}, "properties": 13 {"Bedroom":4,"Price":675}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0021, 42.39081]}, "properties": 14 {"Bedroom":3,"Price":700}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.168916, 42.343392]}, "properties": 15 {"Bedroom":0,"Price":700}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1504562, 42.3549199]}, "properties": 16 {"Bedroom":1,"Price":750}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.168916, 42.343392]}, "properties": 17 {"Bedroom":3,"Price":750}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1121842, 42.3327401]}, "properties": 18 {"Bedroom":0,"Price":800}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0271377, 42.3349841]}, "properties": 19 {"Bedroom":0,"Price":800}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.059712, 42.3766569]}, "properties": 20 {"Bedroom":4,"Price":800}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1054599, 42.3496119]}, "properties": 21 {"Bedroom":1,"Price":800}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1628845, 42.2768106]}, "properties": 22 {"Bedroom":1,"Price":800}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.082931, 42.332471]}, "properties": 23 {"Bedroom":1,"Price":800}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.029324, 42.3651692]}, "properties": 24 {"Bedroom":1,"Price":850}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0419176, 42.4339217]}, "properties": 25 {"Bedroom":1,"Price":850}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1692031, 42.3441413]}, "properties": 26 {"Bedroom":2,"Price":850}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0838132, 42.3356914]}, "properties": 27 {"Bedroom":1,"Price":875}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1163386, 42.3485271]}, "properties": 28 {"Bedroom":0,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1401128, 42.3363406]}, "properties": 29 {"Bedroom":0,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1464497, 42.2987312]}, "properties": 30 {"Bedroom":0,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0897, 42.41041]}, "properties": 31 {"Bedroom":1,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0103177, 42.2559937]}, "properties": 32 {"Bedroom":1,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1570168, 42.3472801]}, "properties": 33 {"Bedroom":2,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1034359, 42.3884241]}, "properties": 34 {"Bedroom":2,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.037249, 42.3297091]}, "properties": 35 {"Bedroom":2,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1157314, 42.2897066]}, "properties": 36 {"Bedroom":3,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1319121, 42.3558468]}, "properties": 37 {"Bedroom":3,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1495, 42.2883]}, "properties": 38 {"Bedroom":3,"Price":900}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.1714369, 42.4212024]}, "properties": 39 {"Bedroom":0,"Price":925}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.042408, 42.388433]}, "properties": 40 {"Bedroom":0,"Price":925}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0424, 42.3884]}, "properties": 41 {"Bedroom":0,"Price":925}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0931052, 42.3487795]}, "properties": 42 {"Bedroom":0,"Price":925}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.099896, 42.367876]}, "properties": 43 {"Bedroom":1,"Price":925}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.0931052, 42.3487795]}, "properties": 44 {"Bedroom":1,"Price":925}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.143445, 42.4062898]}, "properties": 45 {"Bedroom":0,"Price":950}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-71.069324, 42.3484359]}, "properties": 46 {"Bedroom":0,"Price":950}, {"type": "Feature", "geometry": {"type": "Point", "coordinates": [-70.9911973, 42.4186969]}, "properties": 

```

(3) Manually Collected Data

In order to get first-hand detailed rental information for training models, we collected data of different areas in Boston from <http://www.apartments.com> and <http://www.zillow.com>.

Parameters: Zip-code, number of bedroom, number of bathroom, building type(house/apartment/condo/townhome), 24-emergency/maintenance(yes/no), GYM, AC, free-water, free-electricity, free-heat, dishwasher, microwave, washer-dryer, public-transportation (walking time to nearest transit/subway/commuter rail).

Dataset size: Over 600 rows.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
zipcode	bedroom	bathroom	type	parking	24h-emergenc	GYM	AC	free-water	free-electric	free-heat	dishwasher	microwave	washer-drye	public-transf	price
2108	1	1	2	1	1	1	1	1	0	1	1	1	1	1	2200
2108	2	1	4	0	0	0	0	1	0	1	1	1	2	1	1400
2108	2	2	2	2	1	1	1	1	0	1	1	1	1	2	5500
2108	2	2.5	2	2	1	1	1	1	0	1	1	1	1	2	7000
2108	2	2	2	2	0	0	1	1	0	1	1	1	2	1	1030
2108	2	2	2	2	1	0	1	1	0	1	1	1	1	1	6600
2108	1	1	3	2	0	0	0	1	0	1	0	0	2	1	1450
2108	1	1	2	0	0	0	0	1	0	1	1	0	1	1	3050
2108	1	1	2	0	0	0	0	1	0	1	1	0	1	1	3100
2108	1	1.5	2	1	1	1	1	1	0	1	1	1	1	1	4000
2108	2	2.5	2	1	1	1	1	1	0	1	1	1	1	1	6800
2108	1	1	2	1	1	1	1	1	0	1	1	1	2	1	2200
2108	2	1	4	0	0	0	0	1	0	1	0	1	1	1	1400
2108	2	3	3	2	0	0	0	1	0	1	1	0	1	1	6200
2108	1	1	2	2	0	0	0	1	0	1	1	0	2	1	2600
2108	2	1	3	0	0	0	0	1	0	1	1	1	2	1	2900
2108	1	1.5	3	1	1	1	1	1	1	1	1	1	1	1	4500
2108	1	1	1	0	0	0	1	1	0	1	1	0	2	1	2900
2109	1	1	2	1	1	1	1	1	0	1	1	1	1	1	3025
2109	2	2	2	1	1	1	1	1	0	1	1	1	1	1	4420
2109	1	1	2	1	1	1	1	1	0	1	1	1	1	1	3800
2109	2	1	2	1	1	1	1	1	0	1	1	1	1	1	4000
2109	1	1	2	1	1	1	1	1	0	1	1	1	1	1	2900
2109	1	1	2	1	1	1	1	1	0	1	1	1	1	1	3200
2109	2	2	2	1	1	1	1	1	0	1	1	1	1	1	3700
2109	2	1.5	2	2	0	1	1	1	0	1	1	1	1	2	3600
2109	2	1.5	2	2	0	1	1	1	0	1	1	1	1	2	3400
2109	1	1	2	0	0	0	0	1	0	1	0	1	2	1	1800
2109	1	1	2	2	0	0	1	1	0	1	0	1	1	1	3600
2109	2	2	2	0	0	0	1	1	0	1	1	1	1	1	4200
2109	2	1.5	2	0	0	0	1	1	0	1	1	0	2	1	2900
2109	2	1	2	2	0	0	0	1	0	1	1	0	2	1	2700
2109	1	1	3	2	0	0	0	1	0	1	1	1	2	1	2650
2109	2	1	2	2	0	0	0	1	0	1	1	1	1	1	3523

3. Azure

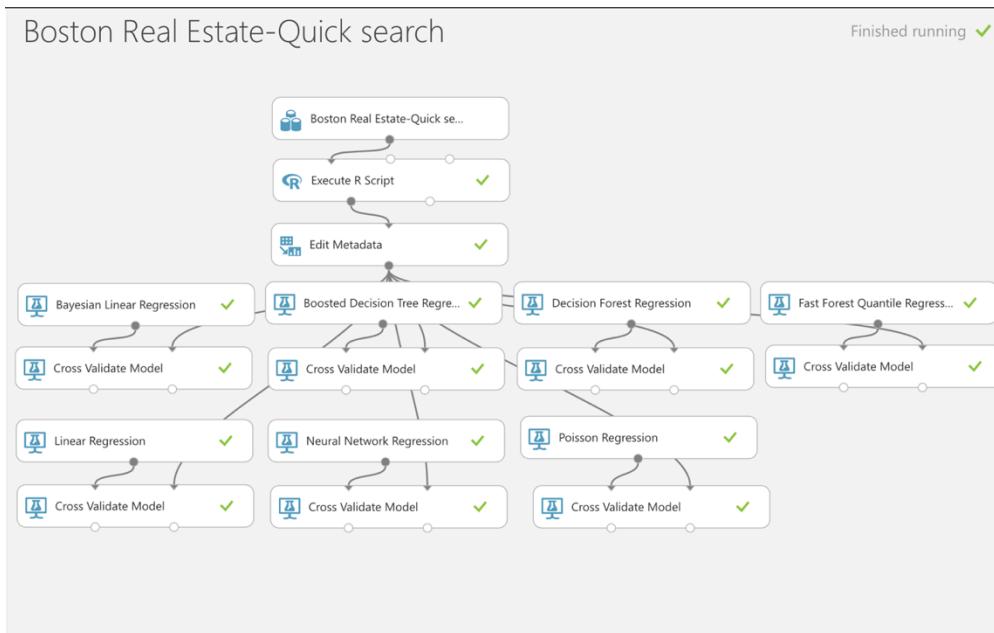
(1) Azure Studio

We deployed 2 web services.

(a) For “quick prediction” function

Arguments: Zip-code, number of bedroom, number of bathroom, building type

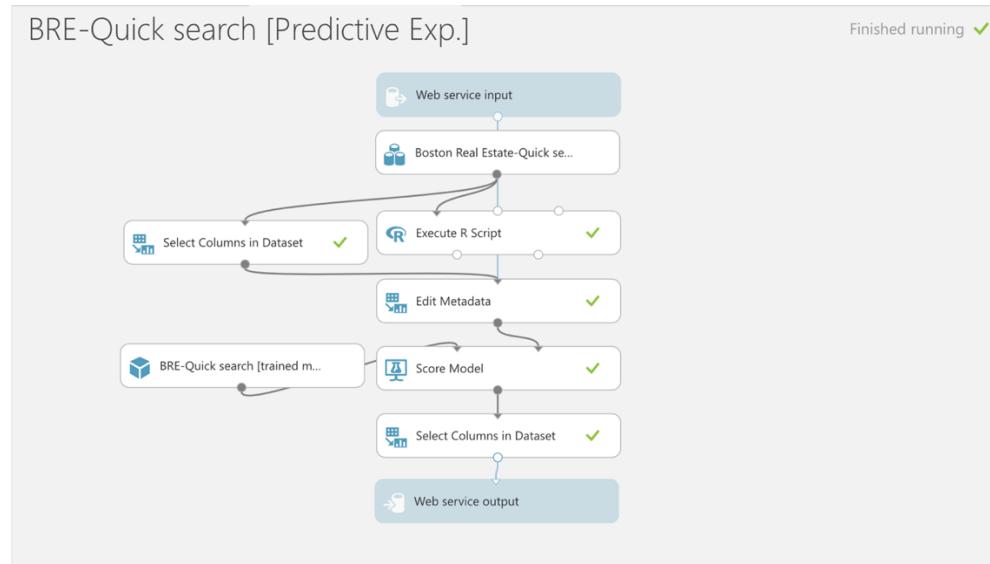
Models we considered:



Performance Metrics :

Price ~ zip + bedroom + bathroom + type					
	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of determination(R^2)
Boosted Decision Tree Regression	575	888	0.59	0.45	0.55
Bayesian Linear Regression	616	899	0.63	0.45	0.55
Decision Forest Regression	619	967	0.63	0.52	0.47
Linear Regression	613	891	0.62	0.44	0.56
Neural Network Regression	1777	2061	1.84	2.55	-1.55
Poisson Regression	769	1080	0.78	0.63	0.37
Fast Forest Quantile Regression	Average Quantile Loss: 318				

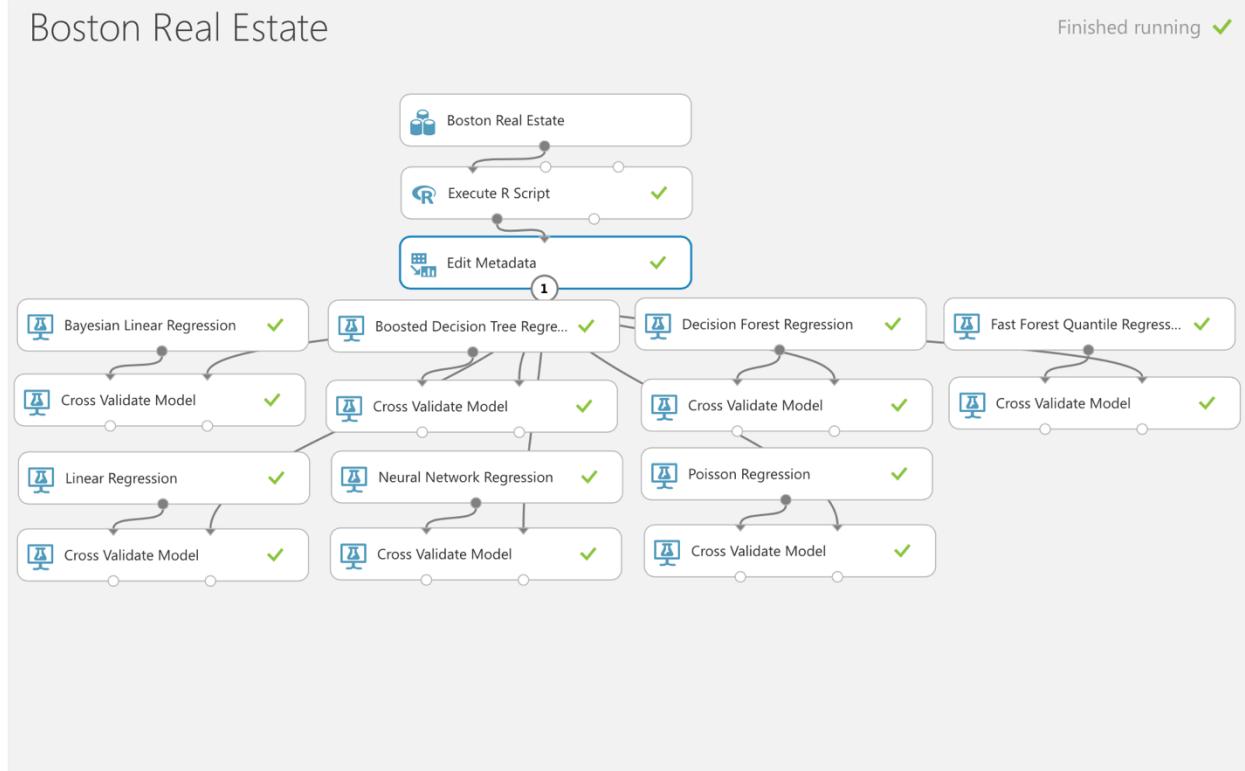
The one we used in web service: Boosted Decision Tree Regression



(b) For “advanced estimate” function

Arguments: Zip-code, number of bedroom, number of bathroom, building type, 24-emergency, GYM, AC, free-water, free-electricity, free-heat, dishwasher, microwave, washer-dryer, public-transportation

Models we considered:



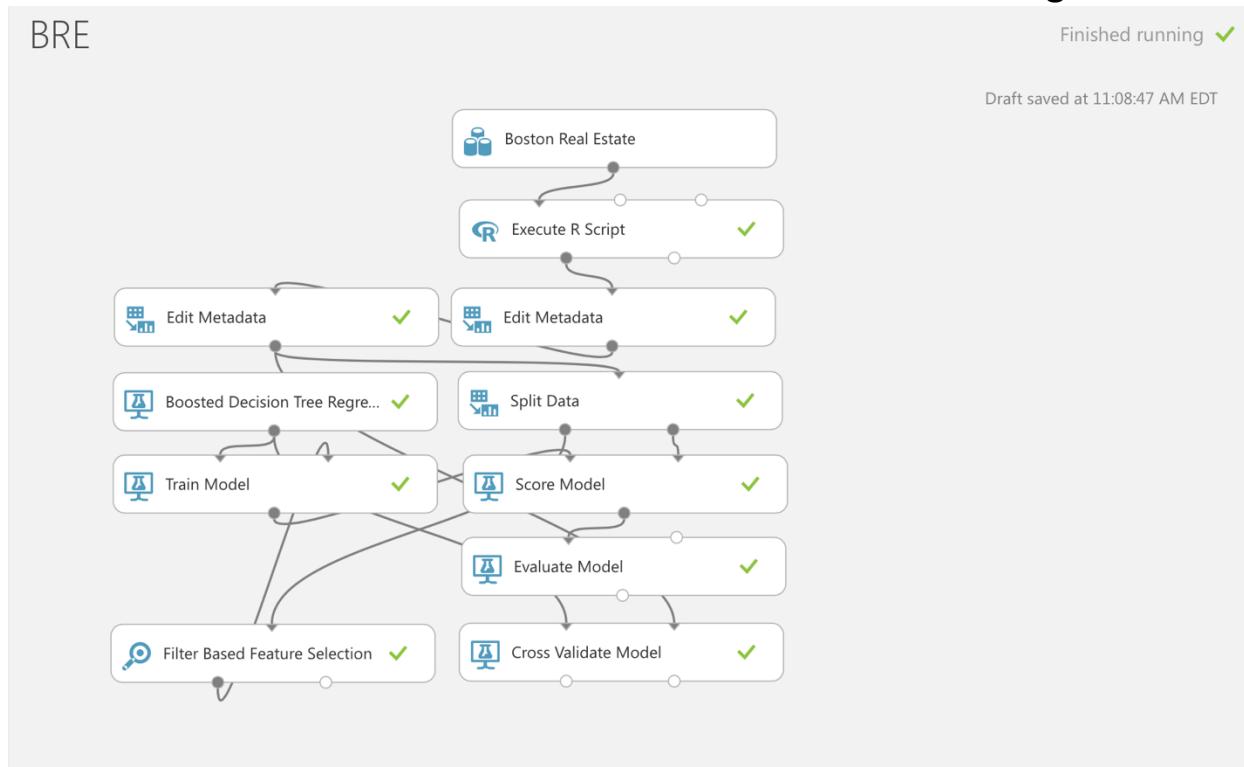
Performance Metrics:

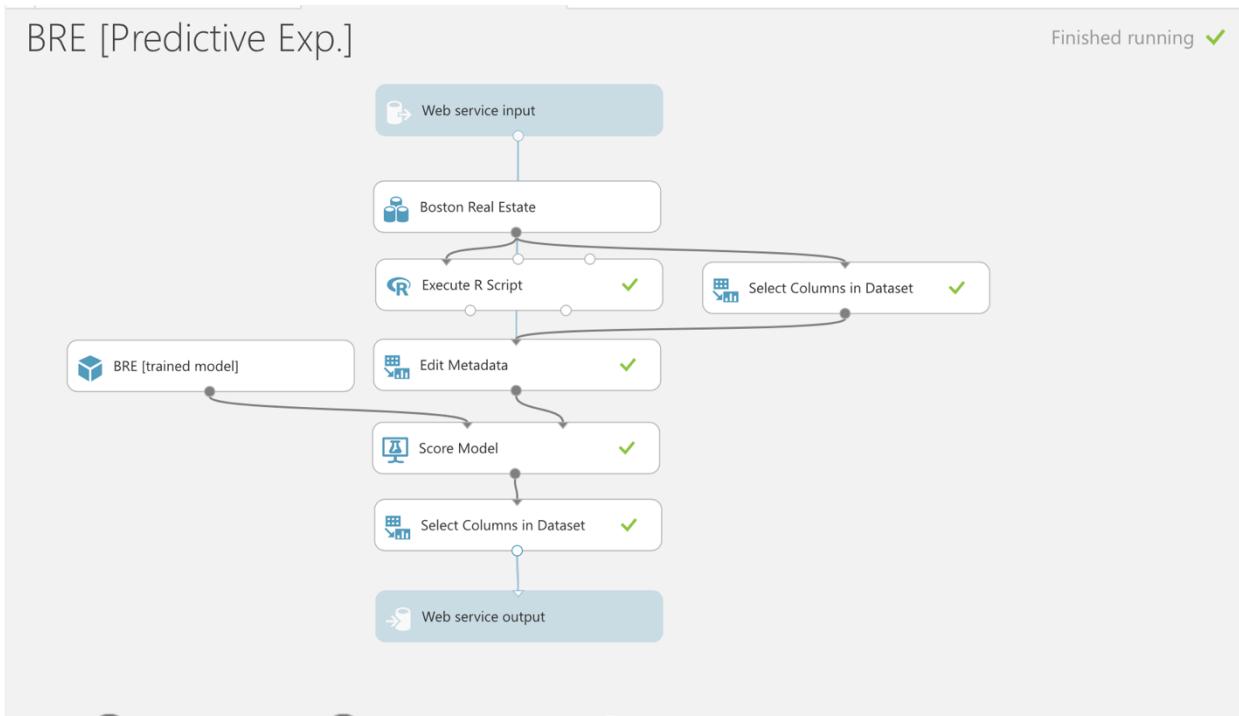
Price ~ zip + bedroom + bathroom + type + 24-emergency/maintenance + GYM + AC + free-water + free-electricity + free-heat + dishwasher + microwave + washer-dryer + public-transportation

	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of determination(R^2)
Boosted Decision Tree Regression	556	841	0.57	0.40	0.60
Bayesian Linear Regression	591	866	0.60	0.42	0.58
Decision Forest Regression	602	918	0.61	0.46	0.54

Linear Regression	498	878	0.61	0.43	0.57
Neural Network Regression	1770	2054	1.83	2.54	-1.54
Poisson Regression	657	960	0.67	0.51	0.49
Fast Forest Quantile Regression	Average Quantile Loss: 286				

The model we used in web service: Boosted decision tree regression





(c) Boosted Decision Tree Regression

It builds each regression tree in a step-wise fashion, using a predefined loss function to measure the error in each step and correct for it in the next. Thus the prediction model is actually an ensemble of weaker prediction models.

In regression problems, boosting builds a series of trees in a step-wise fashion, and then selects the optimal tree using an arbitrary differentiable loss function.

Trained model:

Boosted Decision Tree Regression

Create trainer mode
Single Parameter

Maximum number of leaves per tree
20

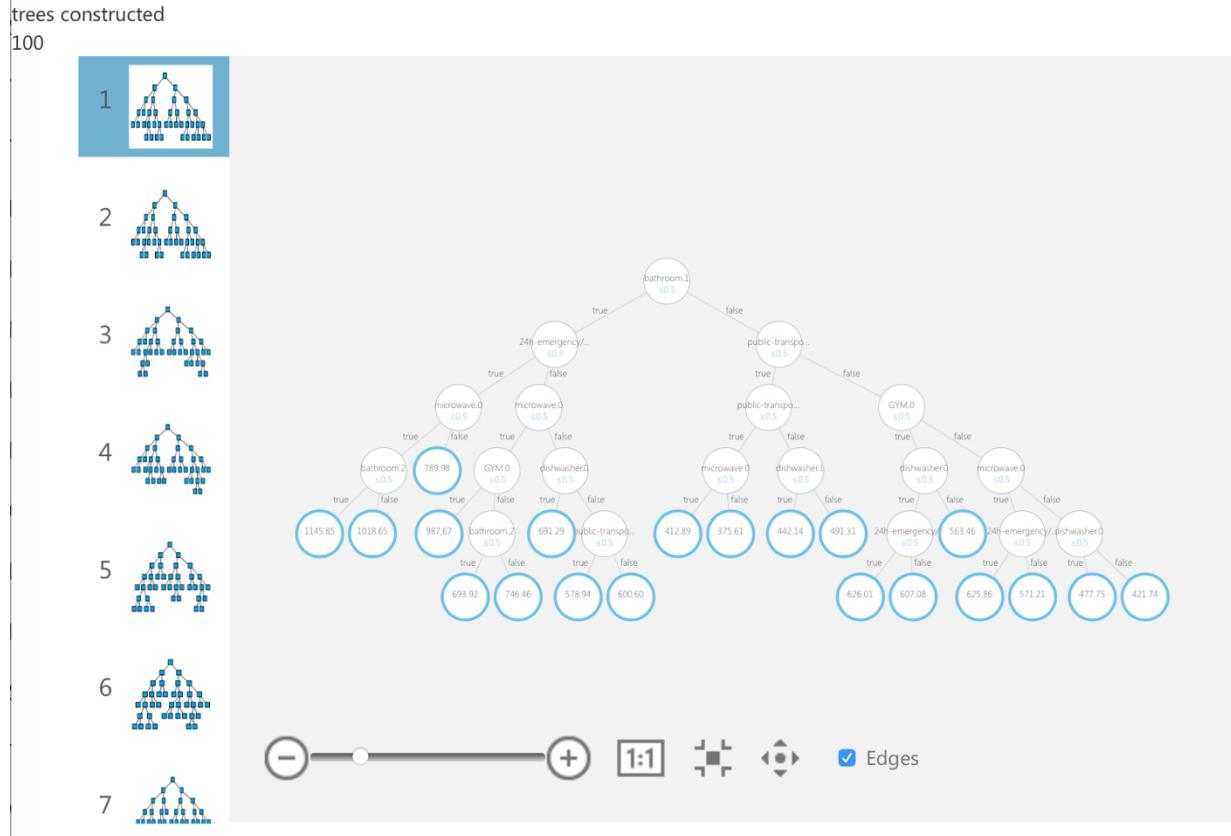
Minimum number of samples per leaf node
10

Learning rate
0.2

Total number of trees constructed
100

Random number seed
[empty]

Allow unknown categorical levels



(2) Azure Portal

Azure File storage offers shared storage for applications using the standard SMB 2.1 or SMB 3.0 protocol. Microsoft Azure virtual machines and cloud services can share file data across application components via mounted shares, and on-premises applications can access file data in a share via the File storage API.

Files we stored: .geojson files with monthly rental data. They will be displayed in the map in our web application.

Microsoft Azure finalprojbostonrent

finalprojbostonrent Storage account - General

Settings Delete

Essentials

Resource group: finalprojresourcegroup

Status: Primary: Available, Secondary: Available

Location: Central US, East US 2

Subscription name: 免费试用

Subscription ID: 70a7d39e-34b9-47b5-a59b-2fa34cccdce8f

Performance: Standard

Replication: Read-access geo-redundant storage (RA-G...)

SETTINGS

- Access keys
- Configuration
- Custom domain
- Encryption
- Shared access signature

Blobs

Files

Tables

Queues

Monitoring

Total requests

This screenshot shows the Azure Storage account overview for 'finalprojbostonrent'. It includes sections for 'Essentials' (resource group, status, location, subscription info, and performance), 'SETTINGS' (access keys, configuration, custom domain, encryption, shared access signature), and monitoring tiles for Blobs, Files, Tables, and Queues. A 'Monitoring' section displays a chart for 'Total requests' with a sharp peak around 6 PM on Aug 18.

finalprojbostonrent Storage account - General

Settings Delete

Monitoring

Add tiles +

Total requests

2.5K
2K
1.5K
1K
0.5K
0K

Aug 18 6 AM 12 PM 6 PM

BLOB 4 TABLE 938 QUEUE 4 FILE 2.46 k

Essentials

File service

File service (finalprojbostonrent)

Settings File share

NAME

asd

MONITORING

Total egress

Average E2E latency

Success percentage

8MB
6MB
4MB
2MB
0MB

0
20
10
0

100%
50%
0%

Aug 18 6 AM 12 PM 6 PM

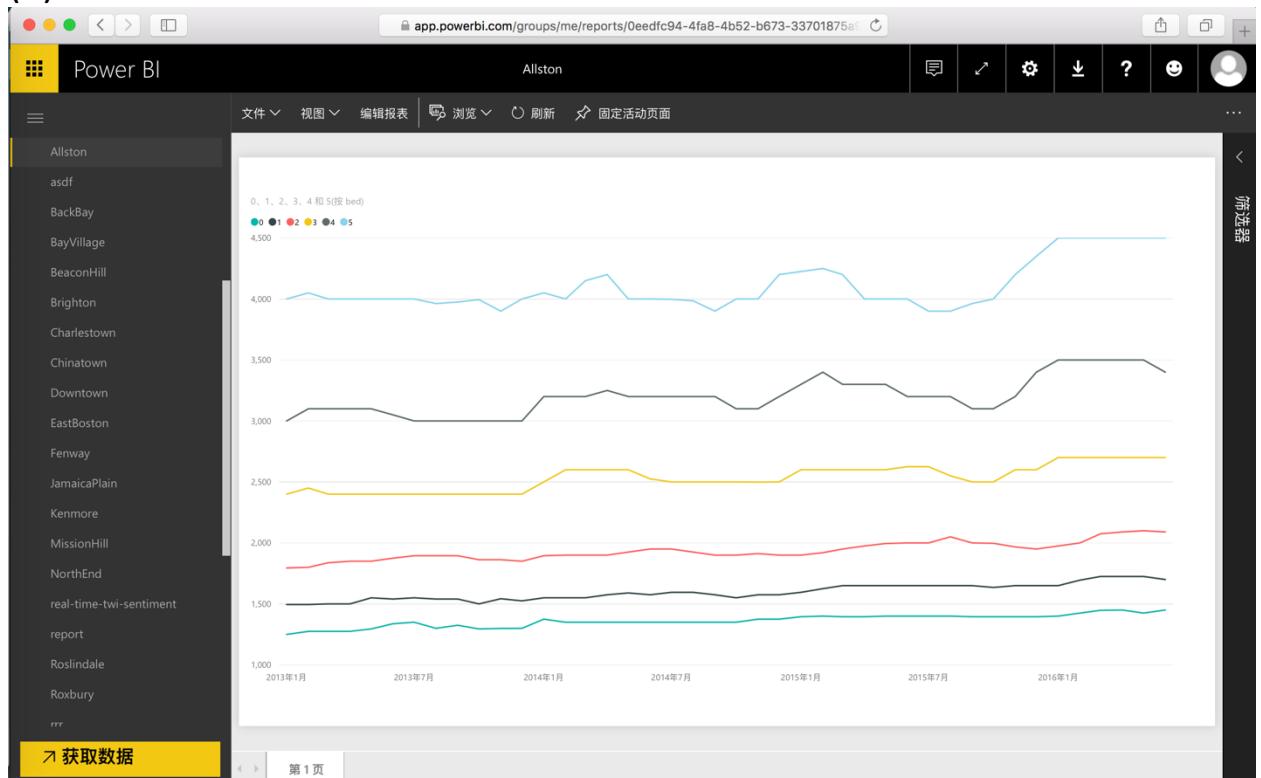
Add a section +

This screenshot shows the monitoring section of the Azure Storage account. It displays four main metrics: 'Total egress' (8MB to 0MB), 'Average E2E latency' (0 to 20 seconds), and 'Success percentage' (100% to 0%). Below these are three charts for 'BLOB', 'TABLE', and 'QUEUE' with specific values: 4, 938, and 4 respectively, and a total file count of 2.46k. A search bar for file shares by prefix 'asd' is also present.

The screenshot shows the Microsoft Azure File service interface. On the left, there's a navigation sidebar with options like Resource groups, All resources, Recent, App Services, Virtual machines (classic), Virtual machines, SQL databases, Cloud services (classic), Security Center, Subscriptions, and More Services. The main area is titled 'File service' and shows a list of files in a share named 'asd'. The list includes files named 2013-02-18.geojson through 20132.geojson, all listed as 'File' type. The interface has a dark theme with light-colored cards for each file entry.

4. Visualization

(1) Power BI



Visualized data: Monthly median rents of houses/aps with different bedrooms (0/1/2/3/4/5) for each neighborhood in Boston from 2013 to 2016.

Users can view these online graphs by clicking dots in different areas in the map.

(2) Map

a. Data Visualization

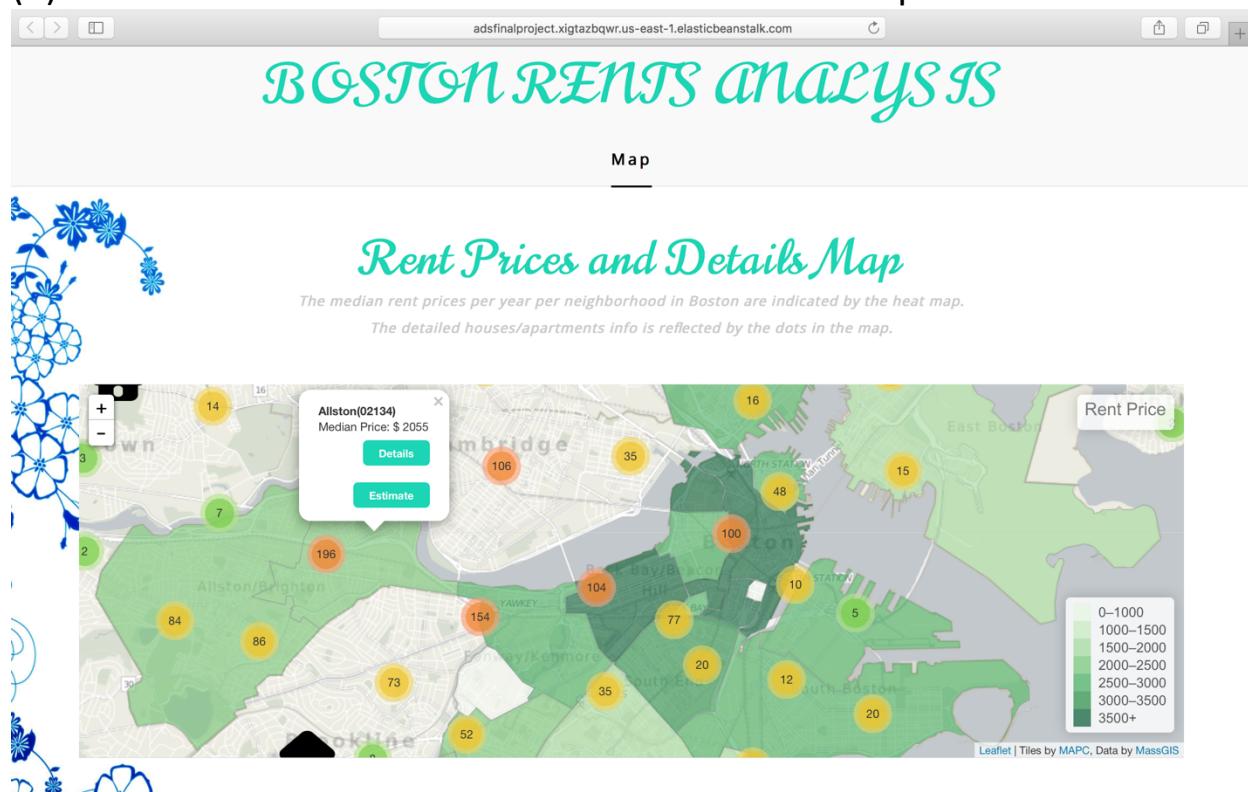
Users can select month in recent years to check for detailed rental info per house/apartment per neighborhood. Or select a year to view median rents per neighborhood reflected in the heat map.

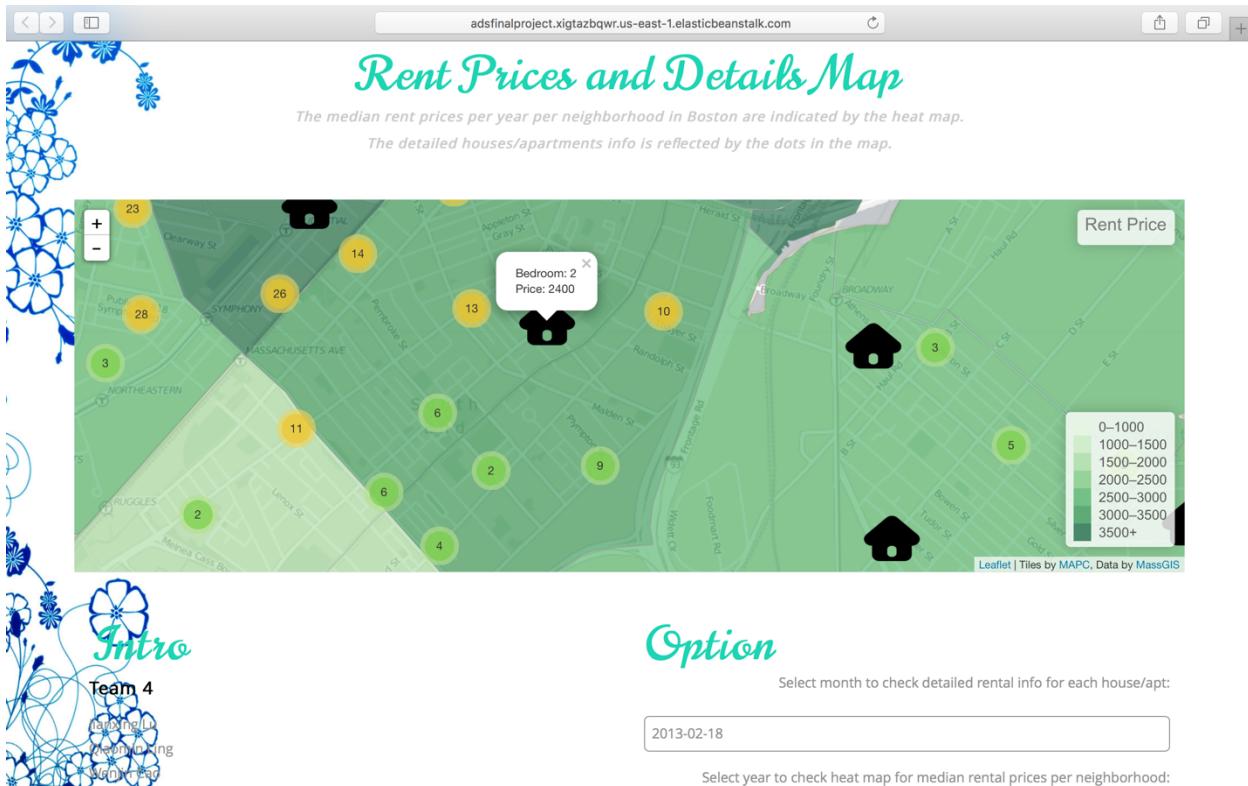
b. Rent Estimate

Users can click on an area and go to the prediction page to get an estimated rent.

5. Demo

(1) Interactive visualization of rental data in the map





(2) Estimate rent

This is
Detail Estimate
Estimate Result:
2269.38

a. Quick Estimate

Quick Estimate

Zipcode:	02134
Housing Category:	<input checked="" type="radio"/> Apartment <input type="radio"/> House <input type="radio"/> Condos <input type="radio"/> Townhouse
Room Type:	<input checked="" type="radio"/> Studio <input type="radio"/> One Bed One Bath <input type="radio"/> Two Bed One Bath <input type="radio"/> Two Bed Two Bath <input type="radio"/> Three Bed One Bath <input type="radio"/> Three Bed Two Bath <input type="radio"/> Four Bed Two Bath

b. Advanced(detailed) Estimate

adsfinalproject.xigtzbqwr.us-east-1.elasticbeanstalk.com/Estimate.htm

AC:	<input checked="" type="radio"/> No <input type="radio"/> Yes
Water-Fee Included:	<input checked="" type="radio"/> No <input type="radio"/> Yes
Electronic-Fee Included:	<input checked="" type="radio"/> No <input type="radio"/> Yes
Heat-Fee Included:	<input checked="" type="radio"/> No <input type="radio"/> Yes
Dishwasher Included:	<input checked="" type="radio"/> No <input type="radio"/> Yes
Microwave Included:	<input checked="" type="radio"/> No <input type="radio"/> Yes
Laundry:	<input checked="" type="radio"/> No
	<input type="radio"/> In Unit
	<input type="radio"/> In Building
Walk To Public Transportation:	<input checked="" type="radio"/> Within 10 Mins
	<input type="radio"/> 10-20 Mins
	<input type="radio"/> More Than 20 Mins