Boston Real Estate Rents Analysis

Team 4

Wenjin Cao
Jianxing Lu
Qiaomin Ling

Project Description

- 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.

3 Kinds of Dataset Sources

- (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	X201
Jamaica Plain	1325.0	1337.5	1197.5	1275.0	1262.5	1275	1300.0	1400	1425.0	1475.0	1625.0	
Brighton	1375.0	1397.5	1395.0	1395.0	1400.0	1395	1450.0	1425	1425.0	1400.0	1395.0	
South Boston	2250.0	2400.0	2299.5	2321.0	2348.0	2325	2405.5	2400	2375.0	2375.0	2302.5	
South End	1862.5	1950.0	1750.0	1750.0	1725.0	1675	1750.0	1800	1950.0	2000.0	2000.0	
Allston	1395.0	1400.0	1395.0	1395.0	1400.0	1400	1400.0	1400	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	
Back Bay	1850.0	1800.0	1725.0	1750.0	1700.0	1745	1750.0	1795	1800.0	1800.0	1800.0	
Kenmore	1550.0	1550.0	1600.0	1600.0	1625.0	1625	1625.0	1610	1595.0	1575.0	1600.0	
Mission Hill	1300.0	1412.5	1450.0	1484.5	1495.0	1400	1400.0	1450	1510.0	1550.0	1400.0	
North End	1850.0	1850.0	1800.0	1775.0	1787.5	1700	1750.0	1800	1700.0	1799.0	1900.0	
Beacon Hill	1750.0	1750.0	1750.0	1750.0	1750.0	1750	1700.0	1700	1725.0	1797.5	1800.0	
West End	2000.0	2100.0	2100.0	2047.5	2100.0	2050	2200.0	2240	2267.5	2275.0	2327.5	

3 Kinds of Dataset Sources

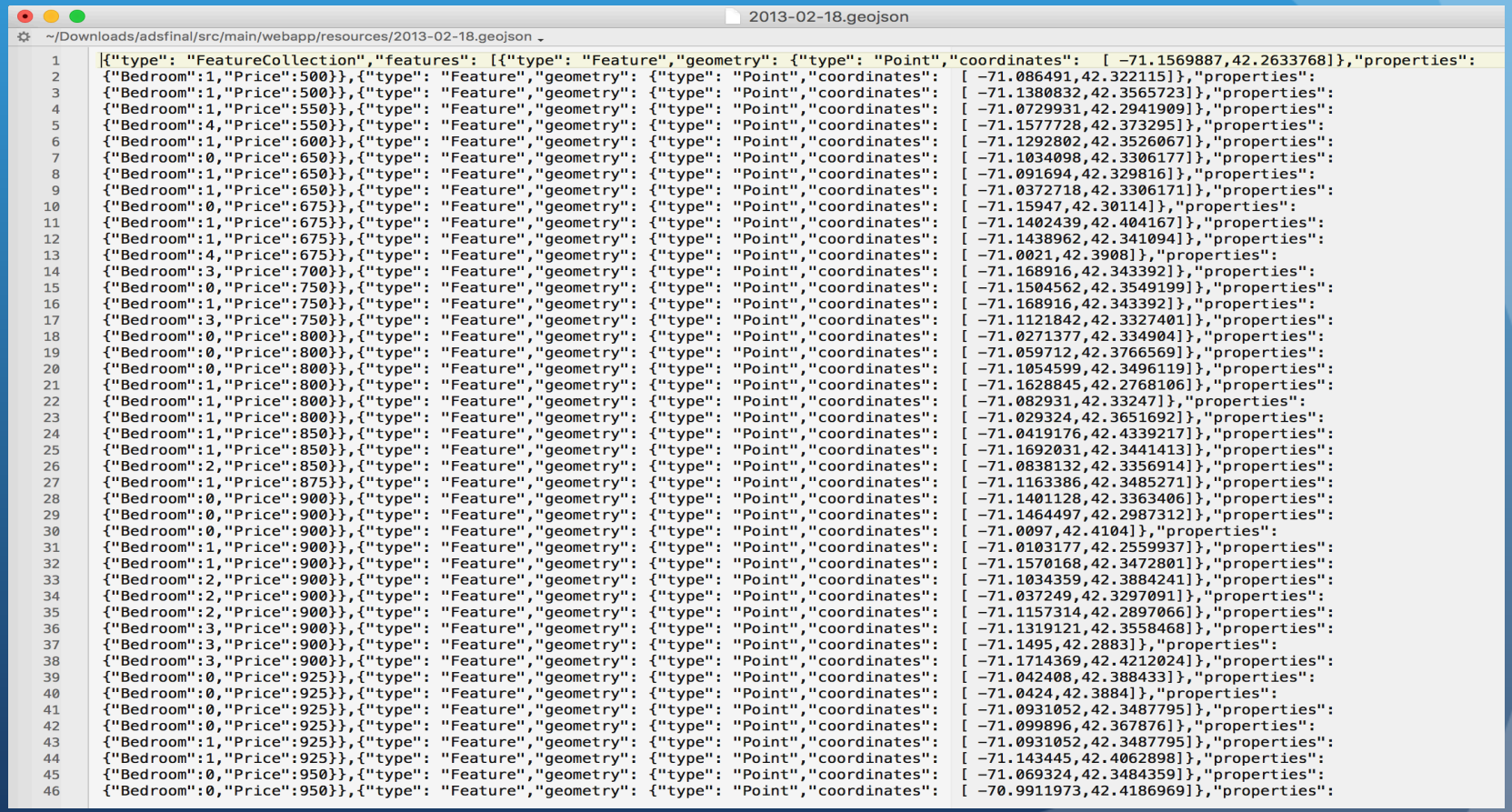
- (2) Rental Data from Individual Third Party
  http://www.jefftk.com/apartment_prices/data-listing

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

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

3 Kinds of Dataset Sources

- (2) Rental Data from Individual Third Party
  http://www.jefftk.com/apartment_prices/data-listing



3 Kinds of Dataset Sources

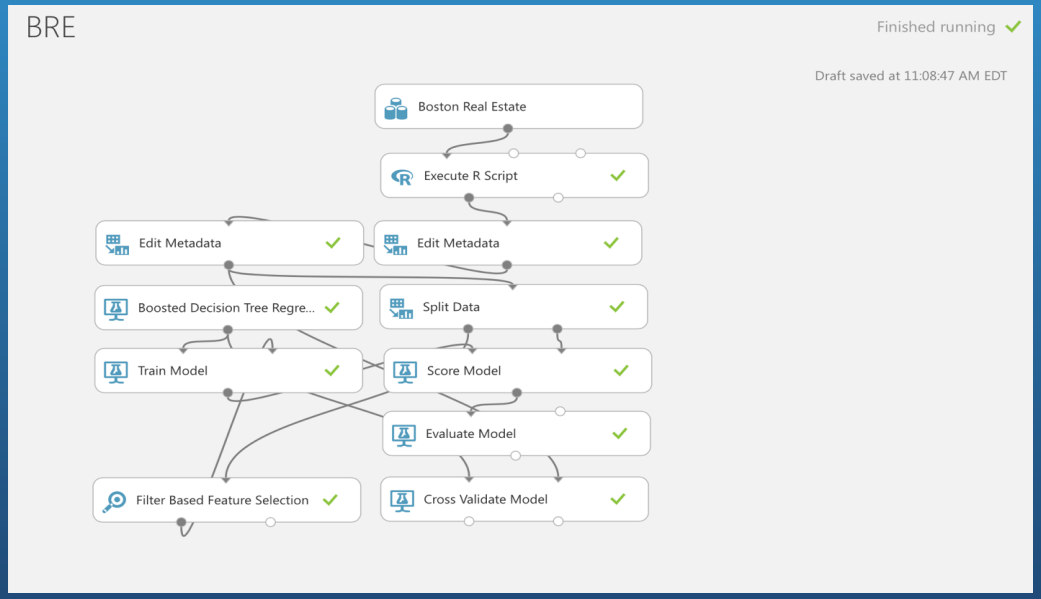
- (3) Manually Collected Data
  In order to get first-hand detailed rental information for training models, we collected data of different areas in Boston from http://www.apartments.com and http://www.zillow.com.
  Parameters: Zip-code, number of bedroom, number of bathroom, building type(house/apartment/condo/townhome), 24-emergency/maintenance(yes/no), GYM, AC, free-water, free
  was
  nea

- Data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
zipcode	bedroom	bathroom	type	parking	24h-emerger	GYM	AC	free-water	free-electrici	free-heat	dishwasher	microwave	washer-drye	public-transp	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

3 Tools We Used:

- (1) Azure Studio
  We compared different regression models and deployed 2 web service. One is for quick prediction in our web application, the other is for advanced prediction.



3 Tools We Used:

- Azure Studio

s deployed 2
n our web

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			
Bayesian Linear Regression	616	899			
Decision Forest Regression	619	967			
Linear Regression	613	891			
Neural Network Regression	1777	2061			
Poisson Regression	769	1080			
Fast Forest Quantile Regression	Average Quantile Loss: 318				

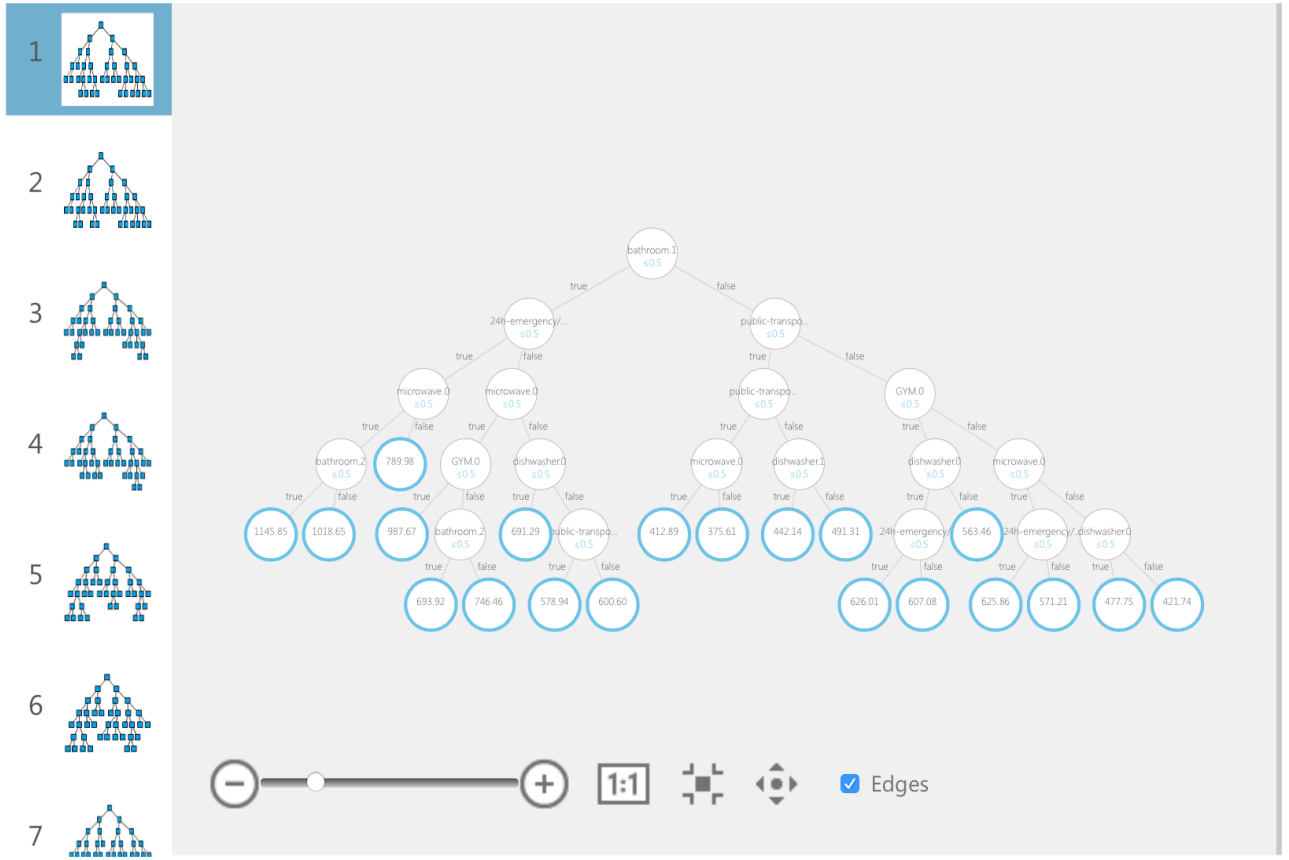
Price ~ zip + bedroom + bathroom + type + 24-emergency/maintenance + GYM + AC + free-water + free-electricity + free-heat + dishwasher + microwave + washer-dryer + public-transportation					
	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of determination(R^2)
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				

3 Tools We Used:

- Az

BRE - Features › Train Model › Trained model

trees constructed
100

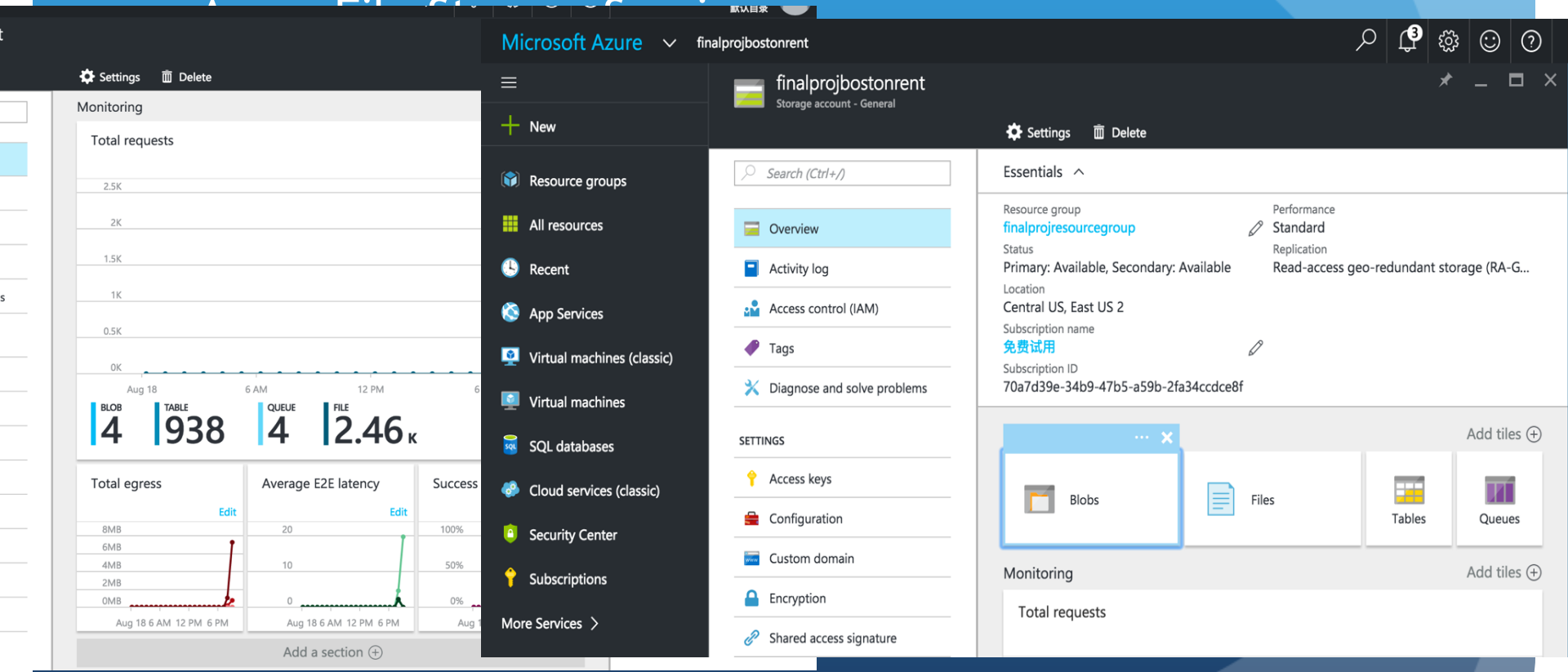


Price ~ z
Boosted Decision Tree Regressi
Bayesian Linear Regressi
Decision Forest Regressi
Linear Regressi
Neural Network Regressi
Poisson Regressi
Fast For Quantile Regressi

Regression					
Fast Forest Quantile	Average Quantile Loss: 286				

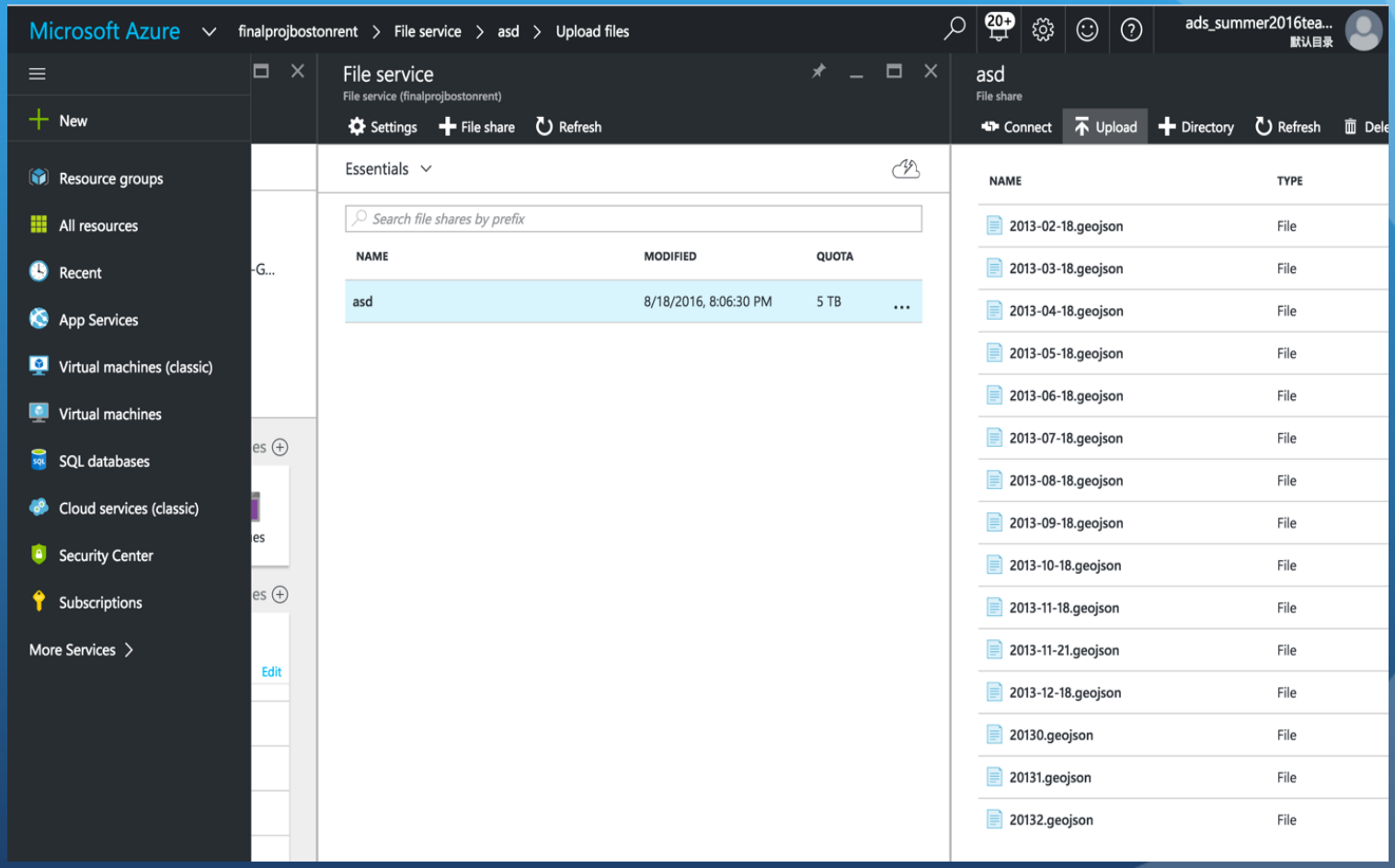
3 Tools We Used:

- (2) Azure Portal
  We stored the geojson files we mentioned before using



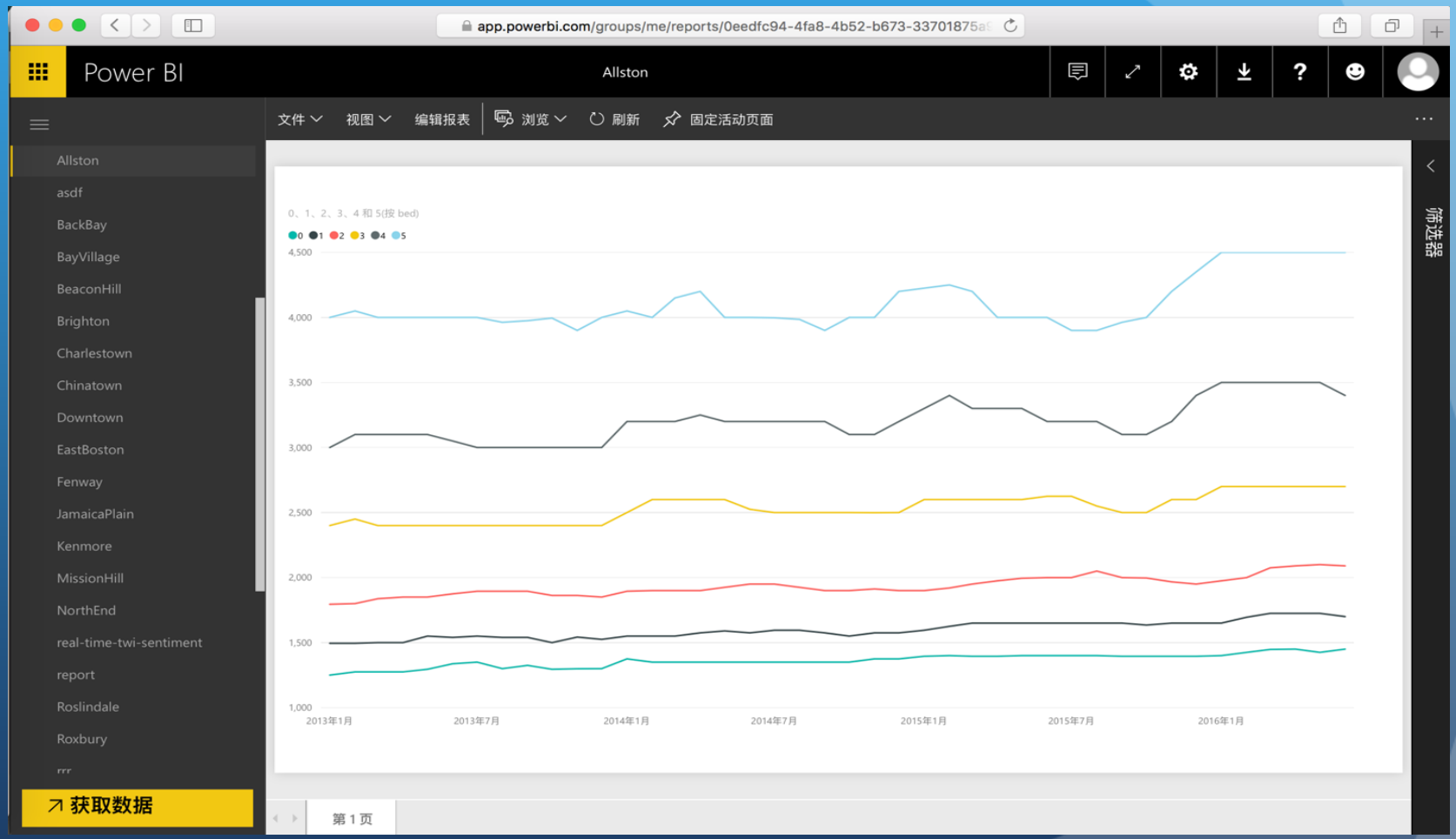
3 Tools We Used:

- (2) Azure Portal



3 Tools We Used:

- (3) Power BI



Web application

http://adsfinalproject.xigtazbqwr.us-east-1.elasticbeanstalk.com/