



Northeastern University



# Web Crawler for Boston Realty Sites

Empowering Smart Living  
Choices

- Achyut Ramakrishna Kowshik
- Karan Mudaliar
- Sashank Reddy Vasepalli

# Objectives of this Project

---

- Develop an efficient web crawler to collect details of rental listings.
- Organize the collected data into a structured database through SQL and Python 3, enabling easy querying and analysis of the rental housing market in Boston.
- Create an interactive Tableau dashboard to dynamically update Rental Data provided by the web crawler pipeline in real-time.
- This enables more informed choices through comprehensive housing information and user-friendly tools.

# Why this topic?

- Only two cities had rent grow faster than the Boston area between July 2021 and July 2022.
- Students are struggling to find affordable housing within the vicinity of their universities.
- We want to provide students with a tool that aids them in finding affordable housing within Boston.



# Methodology

- Web Crawling & Data Scraping
- Data Transformation & Storage
- Data Visualization
- Automation and Deployment

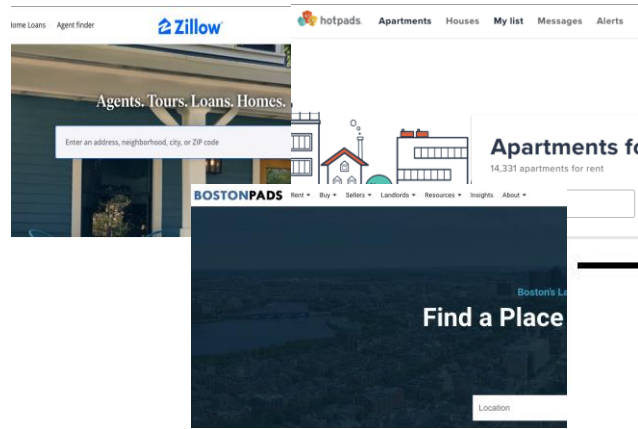
# Web Crawling & Data Scrapping

BOSTONPADS

Zillow

hotpads.com™  
the place to find your place™

YouGotListings



Website



Web Scrapping



Structured Data

## Breadth-First Search (BFS) Algorithm

- Implemented BFS for systematic web crawling
- Ensured exploration of webpages in a layer-by-layer manner
- Visited all relevant pages before moving to the next level

# Data Transformation and Storage

BeautifulSoup

```
<!DOCTYPE html>
<html lang="en">
  <head> ... </head>
  <body>
    <header> ... </header>
    <main>
      <div class="search_page">
        <div class="search_area"> ... </div>
        <div class="main_wrapper"> ... </div>
        <div class="main_wrapper">
          <div class="search_results_area"> (flex)
            <div class="property_item">
              <a href="https://ygl.is/bryn-allen-1/rental/215274545" c
                _icon" style="background-image: url('https://www.uds.com/
                dc9266-23ae-8a2f-2019-35a06c7a1b04.jpeg');"> ... </a>
              <div class="details"> == $0
                <div class="item_props"> ... </div> (flex)
                  <a href="https://ygl.is/bryn-allen-1/rental/215274545"
                    Harrison Ave., #Ph12, Boston (South End)</a>
                </div>
```

```
{*total":1274,*per_page":40,*current_page":1,*last_page":32,*next
_page_url":"http://m.bostonpads.com/api/listings-short?
page=2","prev_page_url":null,"from":1,"to":40,"data":
[{"advertiser_id":null,"agent_ad_num":113,"ad_id":4399633,"sub_area
_name":"BOSTON - WEST END","zip":"02114","building_name":"Emerson
Place Apartments","building_address":{"street_address":"1 Emerson
Place","city":"Boston","zip":"02114","state":"Massachusetts"},"styl
e":"Apartment","luxury_building":true,"pets":1,"bed_room":3,"bath
s":2,"heat":true,"hot_water":1,"electricity":false,"extra_stora
ge":false,"parking":"Garage","dish_washer":true,"laundry_location":
"Common
Area","air_conditioning":"Central","furnished":"","create_date":"20
12-02-17 00:00:00","date_modified":"2023-08-06
21:25:49","date_available":"2023-08-03
00:00:00","price":$5,355,"status":"Active","first_month":4490,"
last_month":4490,"security":4490,"fee":1,"feed_type":"RENTAL","
masked_lng":-71.0685066,"masked_lat":42.3647924,"virtual_tour_url":
null,"video_vistia_id":null,"price_drop":null,"agent_email":"cwalsh
@rentboardwalk.com","agent_phone":"(617) 433-
0392","agent_full_name":"Connor
Walsh","is_short_term":false,"advertiser_type":null,"short_term_ame
nities":null,"logo_url":null,"picCount":30,"vidCount":0,"availabili
ty":"Now","feeText":"","pets_description":"Pet
OK","primaryPhotoAspectRatio":0.56,"firstPhoto":"/media/v2/listi
ngs/boston-west-end/photos/12-17/1514567884-14.jpg"},
{"advertiser_id":null,"agent_ad_num":1,"ad_id":4422782,"sub_area_na
me":"BOSTON - DORCHESTER - UPHAMS
```


	last_updated	price	n_beds	n_baths	avbl_date	location
0	2023-08-05	3073	1.0	1.0	2023-08-02 00:00:00	3611-15 Washington St.
1	2023-08-05	3794	2.0	1.0	2023-08-02 00:00:00	3611-15 Washington St.
2	2023-08-05	2667	1.0	1.0	2023-08-02 00:00:00	3611-15 Washington St.
3	2023-08-05	5355	3.0	2.0	2023-08-03 00:00:00	1 Emerson Place
4	2023-08-05	8214	3.0	2.0	2023-07-31 00:00:00	131 Seaport Boulevard
5	2023-08-05	6291	2.0	2.0	2023-08-01 00:00:00	100 Sudbury St.
6	2023-08-05	6750	5.0	2.0	2023-08-15 00:00:00	BOSTON - ALLSTON
7	2023-08-05	4500	2.0	2.0	2023-07-15 00:00:00	BOSTON - NORTH END
8	2023-08-05	2800	2.0	1.0	2023-09-01 00:00:00	BOSTON - BRIGHTON - OAK SQUARE
9	2023-08-05	3300	3.0	1.0	2023-09-01 00:00:00	BOSTON - EAST BOSTON - MAVERICK
10	2023-08-05	3000	4.0	1.0	2023-09-01 00:00:00	BOSTON - DORCHESTER/SOUTH BOSTON BORDER
11	2023-08-05	2700	3.0	1.0	2023-09-01 00:00:00	BOSTON - DORCHESTER/SOUTH BOSTON BORDER
12	2023-08-05	3000	4.0	1.0	2023-09-01 00:00:00	BOSTON - DORCHESTER/SOUTH BOSTON BORDER
13	2023-08-05	2300	1.0	1.0	2023-09-01 00:00:00	BOSTON - MISSION HILL
14	2023-08-05	5900	4.0	3.0	2023-09-01 00:00:00	BOSTON - DORCHESTER - SAVIN HILL
15	2023-08-05	3200	1.0	1.0	2023-09-01 00:00:00	BOSTON - SOUTH END



# Data Transformation and Storage

SQLite as our database of choice. Why?



- Lightweight, serverless and embedded
  - Good performance for small-moderate sized data
  - No dependencies
- 



# Database connection to Tableau

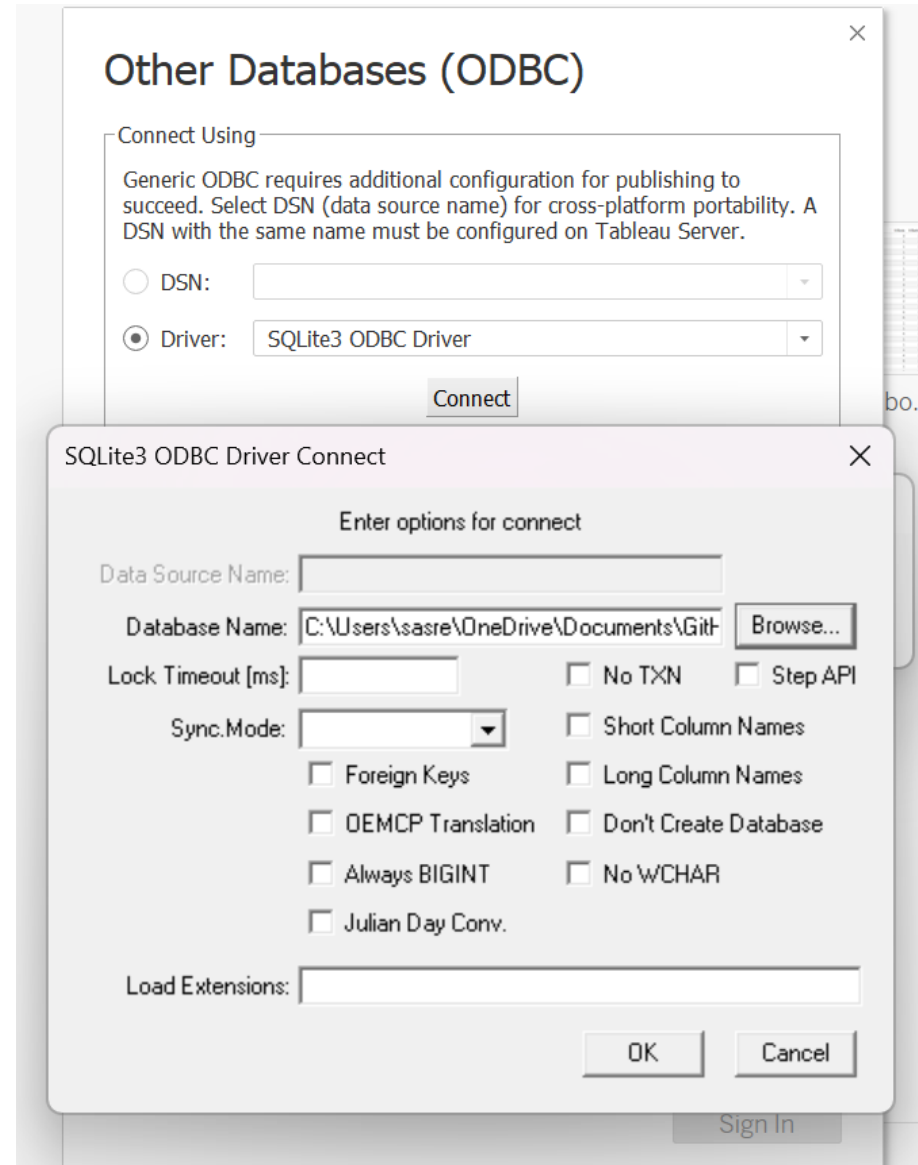
## Connection with SQLite

- Connect Tableau to the SQLite database generated by the Python.
- However, SQLite is not natively supported by Tableau.
- Custom WIN64 driver for SQLite and connect it to tableau using the Microsoft Open Database Connectivity (ODBC) interface.
- Ensure live connection through server settings.





# Custom SQLite ODBC Drivers



# Data Visualization Dashboard



Tableau is used to create easy to use, interactive Dashboards.



We can search and filter listings from several websites in a single location.



We can use analytics to identify trends and make more informed decisions on our rental houses.

# Rental Search Dashboard

## Rental Data

Id	area	Location	Month of A..	Day of A..	N Beds	N Baths	Price
1	West End	1 Emerson Place	August	3	3	2	5,355
2	Boston	BOSTON - NORTHEASTERN/SYMPHONY	September	1	1	1	1,900
3	Boston	BOSTON - DOWNTOWN	September	1	1	1	3,000
4	Charlestown	BOSTON - CHARLESTOWN	September	1	1	1	1,500
5	Back Bay	BOSTON - BACK BAY	September	1	2	1	3,900
6	Boston	BOSTON - FENWAY/KENMORE	August	15	1	1	2,395
7	Boston	BOSTON - DOWNTOWN	September	1	1	1	2,500
8	Allston	BOSTON - ALLSTON	August	31	1	1	2,300
9	Jamaica Plain	BOSTON - JAMAICA PLAIN - FOREST HILLS	September	2	3	1	3,200
10	Jamaica Plain	BOSTON - JAMAICA PLAIN - FOREST HILLS	September	1	3	1	3,200
11	Mission Hill	BOSTON - MISSION HILL	September	1	2	1	3,000
12	Allston	BOSTON - ALLSTON	August	15	5	2	6,500
13	North End	BOSTON - NORTH END	August	15	2	2	4,950
14	Brighton	BOSTON - BRIGHTON - CLEVELAND CIRC..	September	1	1	1	2,775
15	Dorchester	BOSTON - DORCHESTER - UPHAMS COR..	September	1	6	3	7,000
16	Fort Hill	BOSTON - FORT HILL	September	1	6	3	6,000
17	Boston	BOSTON - FENWAY/KENMORE	September	1	1	1	2,995
18	Boston	BOSTON - FENWAY/KENMORE	September	1	1	1	3,080
19	Brighton	BOSTON - BRIGHTON - BOSTON COLLEGE	September	1	6	2	11,130
20	Boston	BOSTON - FENWAY/KENMORE	September	1	1	1	2,845
21	Boston	BOSTON - FENWAY/KENMORE	September	1	1	1	2,895
22	Boston	BOSTON - FENWAY/KENMORE	September	1	1	1	2,795

### Areas

- ☒ (All)
- ☒ Allston
- ☒ Back Bay
- ☒ Bay Village
- ☒ Beacon Hill
- ☒ Boston
- ☒ Brighton
- ☒ Brookline
- ☒ Cambridge
- ☒ Charlestown
- ☒ Chinatown
- ☒ Dorchester
- ☒ East Boston
- ☒ Fenway

### Beds

1.00 11.00



### Baths

0.00 11.50



### Price

0 50000

## Rental Details

Price: 2,895 \$

Beds: 1.000

Baths: 1.000

Location: BOSTON - FENWAY/KENMORE

Last Updated: August 6, 2023

Utilities paid for: Heat - No, Hot Water - No, Electricity - No.

Amenities provided: Parking - No, Dish Washer - No, Laundry - No, Air Conditioning - No.

## Agent Details

Nicholas Rizzolo

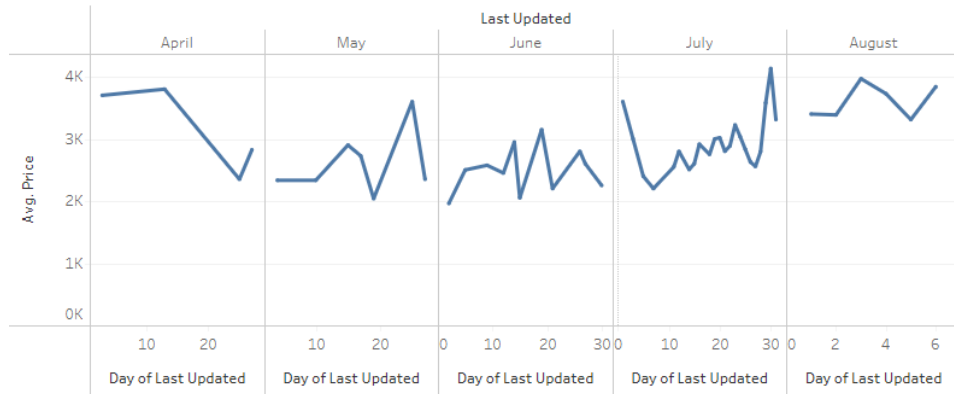
(857) 693-0029

nrizzolo@rentboardwalk.com

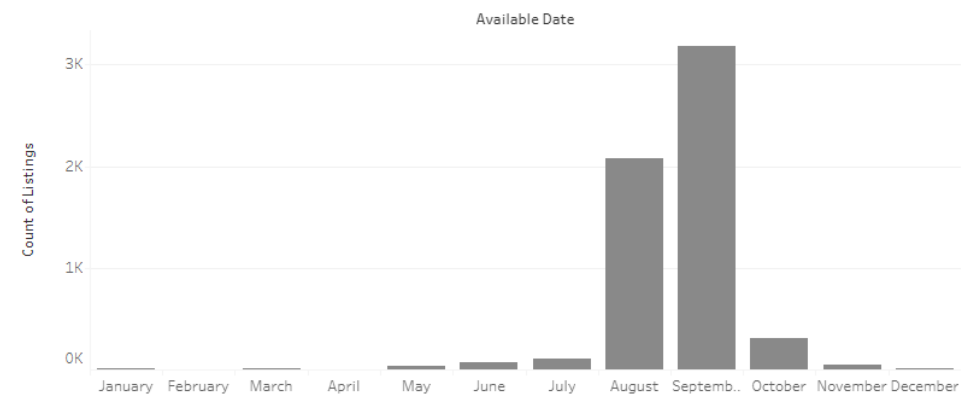
Agent Fees: Yes

# Analytics Dashboards

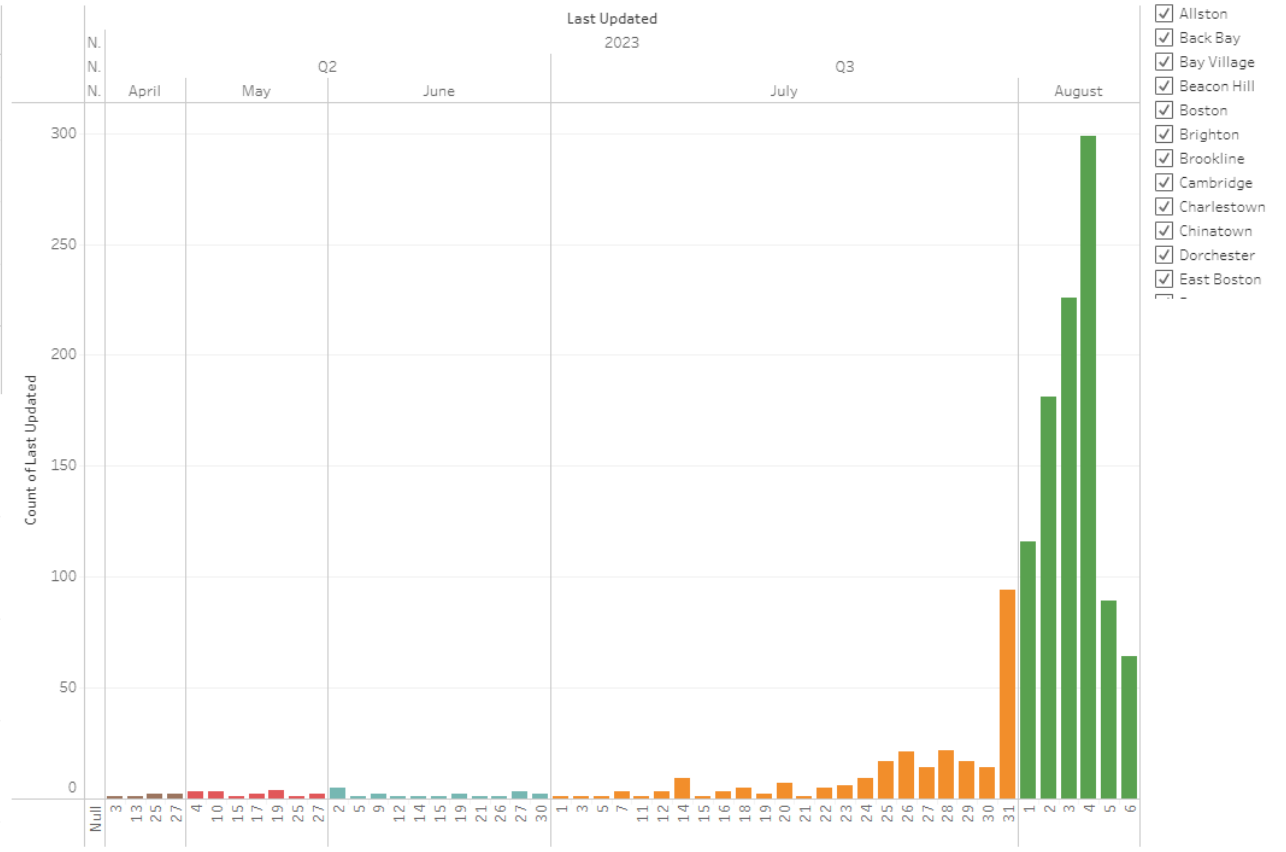
Price change MoM



Availability Histogram



Listings over time



# Deployment

- For live deployment, databricks acts as an alternative way to extract, store, transform and load the scraped data
- Since databricks is cloud hosted and allows parallel processing, it can load and process the data much faster.
- Lastly, it has built in connectors to tableau and also has a built in scheduler to schedule tasks at given intervals.



**databricks**

Thank you