

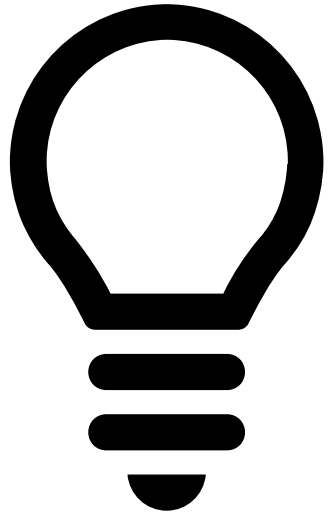


Life is a Struggle
Use Python

MAGGIE
CYNTHIA
LILY
EMMA

Content

- Introduction
- Process
- Result
- Future



Introduction

Introduction

- Use python to scrape 2000+ housing information records from housing rental housing websites
- Clean data and perform regression analysis
- Write a user interaction programme to recommend housings for students and professors near Carnegie Mellon University campus

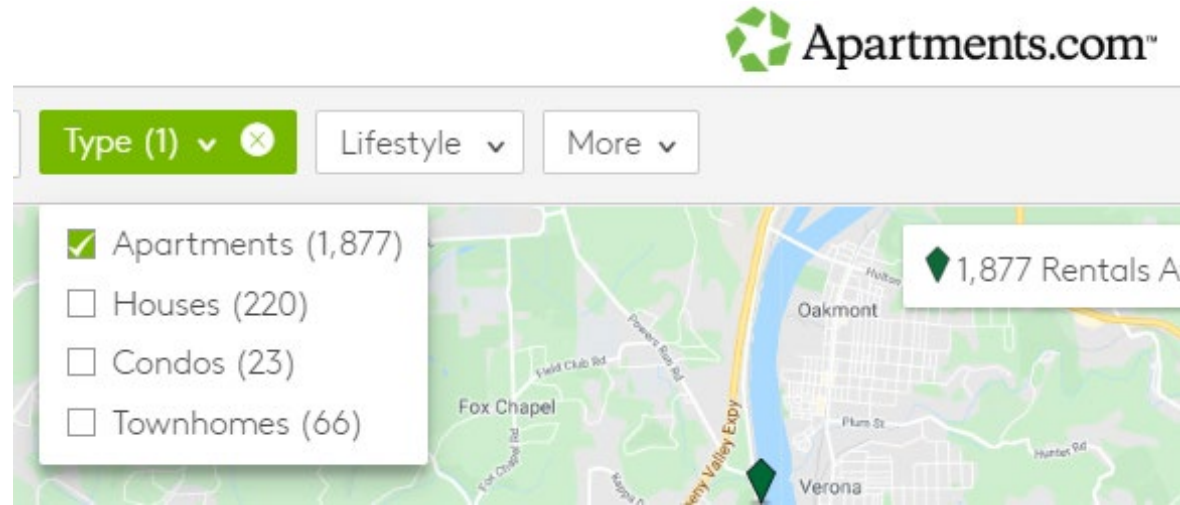


Process

Process - data scraping from Apartments

```
housing=pd.DataFrame(columns=['Address', 'District','Beds', 'Baths', \
                              'Longitudes', 'Latitudes', \
                              'Price', 'Size', 'Website','Distance', 'Type'])
```

Process - data scraping from Apartments



apartments.com/apartments/shadyside-pa/

español

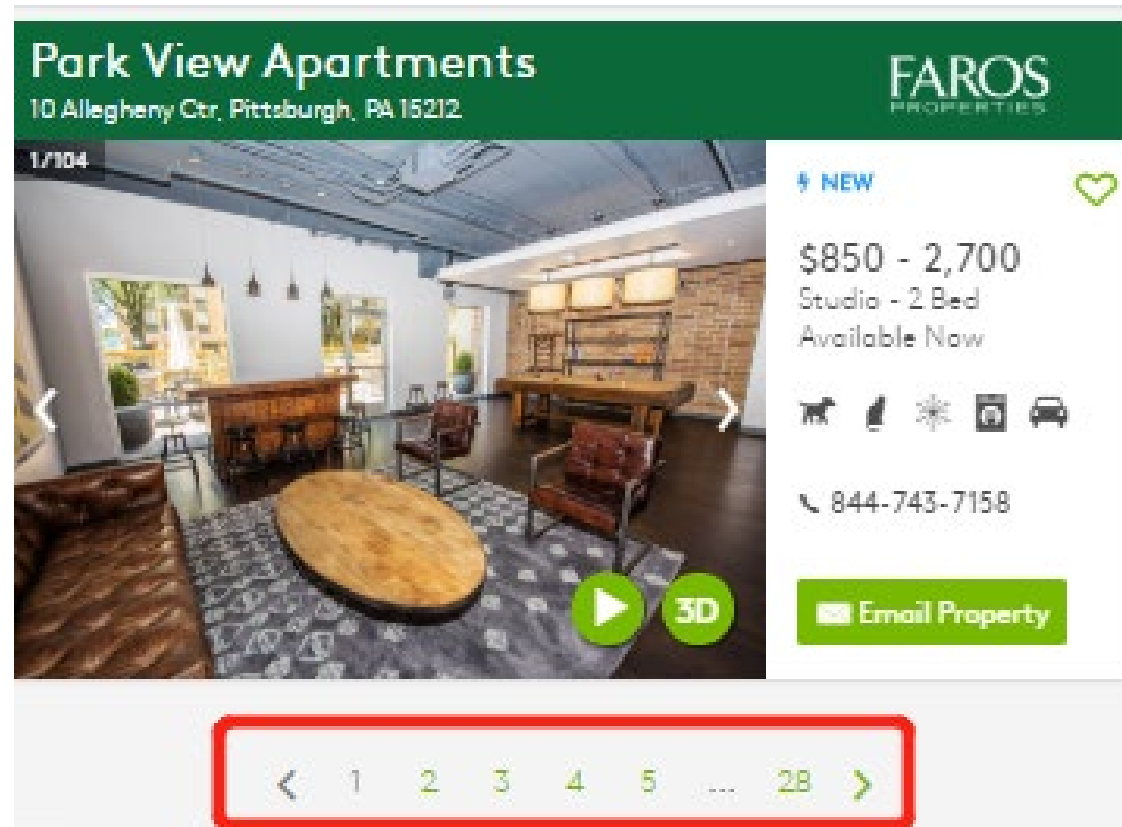
type district



Process - data scraping from Apartments

```
type_url='https://www.apartments.com/'+types+'/'+district+'-pa/'  
r = requests.get(type_url,headers=headers)  
c = r.content  
soup = BeautifulSoup(c,'html.parser')
```


Process - data scraping from Apartments



Process - data scraping from Apartments

```
if(soup.find('div',class_='placardContainer').find('nav',class_='paging')!=None):
    paging = soup.find('div',class_='placardContainer').find('nav',class_='paging').find_all('li')
    page=[]
    #get all possible pages in the bar
    for item in paging:
        if item.find('a') is None: continue
        page.append(item.find('a').get('data-page'))


    start_page = page[1]
    last_page = page[-2]
else:
    start_page = last_page = 0
```

Process - data scraping from Apartments

Heinz Lofts

300 Heinz St, Pittsburgh, PA 15212

1/58








Alterra
PROPERTY GROUP

7 HRS AGO

\$1,400 - 4,750

1-3 Bed

Available Now

412-847-5081

Email Property

Process - data scraping from Apartments

```
childUrl = Title[i]['href']
childResponse = requests.get(childUrl, headers=headers)
if childResponse.status_code == 200:
    childSoup = BeautifulSoup(childResponse.content, 'html.parser')
else:
    print('parsing error: code temporarily unavailable')
```

Process - data scraping from Apartments

Heinz LoftsCoStar Verified®

300 Heinz St, Pittsburgh, PA 15212 - Northshore

★★★★★ (3 Reviews) ▾


1 Bedroom \$1,400 - 2,705


2 Bedrooms \$1,700 - 4,499


3 Bedrooms \$2,730 - 4,750


Home / Pennsylvania / Pittsburgh / Heinz Lofts

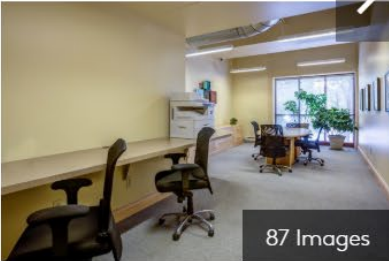
Today











87 Images

All1 Bedroom2 Bedrooms3 Bedrooms

1 Bedroom	1 Bathroom	\$1,400 - 2,350		573 Sq Ft	1 Bedroom - D1 Layout	Available Now	View Model
1 Bedroom	1 Bathroom	\$1,440 - 2,285	C102R	573 Sq Ft	1 Bedroom - D1 Layout	Available Now	View Unit
1 Bedroom	1 Bathroom	\$1,450 - 2,295	C204R	573 Sq Ft	1 Bedroom - D1 Layout	Available Now	View Unit


Process - data scraping from Apartments

```
<link rel="alternate" hreflang="x-default" href="https://  
www.apartments.com/heinz-lofts-pittsburgh-pa/m0k293k/">  
<link rel="alternate" hreflang="es-US" href="https://  
www.apartments.com/es/heinz-lofts-pittsburgh-pa/m0k293k/">  
<meta property="place:location:latitude" content="40.4544">  
<meta property="place:location:longitude" content="-79.9909">  
<link rel="preload" href="/a/01f5ab/content/fonts/  
apartmenticons.woff" as="font" crossorigin>  
<link rel="preload" href="/a/48178f/content/fonts/lineto-costar-  
brown-regular.woff" as="font" crossorigin>  
<link rel="preload" href="/a/0b286e/content/fonts/lineto-costar-  
brown-light.woff" as="font" crossorigin>  
<link rel="preload" href="/a/d21ed8/content/fonts/lineto-costar-  
brown-bold.woff" as="font" crossorigin>  
<link href="/content/style.bundle.css?  
v=8yJ29JXe08PTqGUI3PTL75pHP0bTwe4oR2VRy4wwQXQ" type="text/css"  
rel="stylesheet">
```

Process - data scraping from Apartments


```
latitude = childSoup.find('meta',property="place:location:latitude")["content"]
longitude = childSoup.find('meta',property="place:location:longitude")["content"]
location=(longitude,latitude)
distance=(geodesic(cmu,location).km)
childPanel = childSoup.find('table', class_=re.compile("availabilityTable"))
rows = childPanel.tbody.findAll('tr', class_=re.compile("rentalGridRow"))
```


Process - data scraping from Craigslist


 pittsburgh.craigslist.org/search/apartments=120&availabilityMode=0&housing_type=1&query=squirrel%20hill

▼ > housing ▼ > apts/housing for rent ▼

ig for

« squirrel hill 

s only

 gallery ▼

<< < prev 121 - 240 / 636 next >

Process - data scraping from Craigslist

```
response = requests.get(get_num + "&housing_type=" + str(housetype))
```

```
html_soup = BeautifulSoup(response.text, "html.parser")
```

```
results_num = html_soup.find("div", class_ = "search-legend")
```

```
web = get_web + str(page) + "&housing_type=" + str(housetype)
```

```
response = requests.get(web)
```

```
page_html = BeautifulSoup(response.text, "html.parser")
```

```
posts = page_html.find_all("li", class_ = "result-row")
```

Process - data scraping from Craigslist



Process - data scraping from Craigslist

```
post_title = post.find("a", class_ = "result-title hdrlnk")
```

```
post_title_text = post_title.text
```

```
post_price = float(post.find("span", class_ = "result-price").text.strip().replace("$", ""))
```

```
post_link = post_title["href"]
```

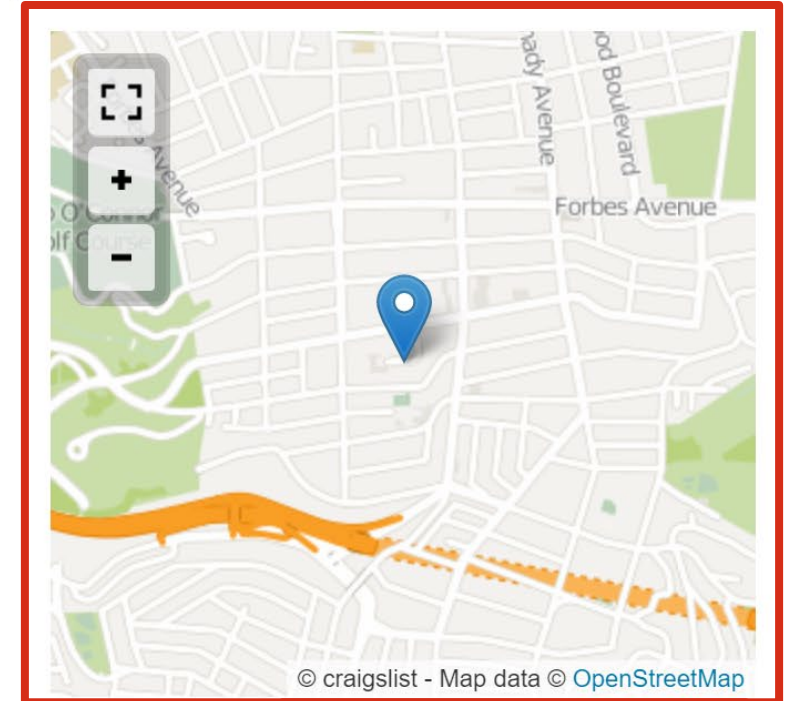
```
lists_html = requests.get(post_link)
```

```
lists = BeautifulSoup(lists_html.text, "html.parser")
```

Process - data scraping from Craigslist

\$965 / 1br - 700ft² - Perfect Squirrel Hill Location and New Kitchen (5646 Munhall Road)

image 1 of 8



Munhall Road near Beacon

([google map](#))

1BR / 1Ba

700ft²

available now

Process - data scraping from Craigslist

```
post_latitude = lists.find("div", id = "map")["data-latitude"]
```

```
post_longitude = lists.find("div", id = "map")["data-longitude"]
```

```
distance = (geodesic(cmu, (post_longitude, post_latitude)).km)
```

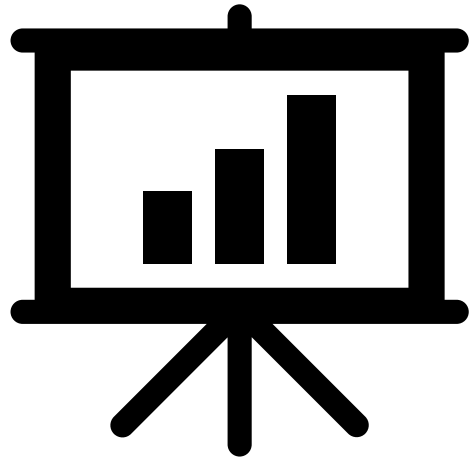
```
post_bedbathsize = lists.find_all(lambda tag: tag.name == "span" and tag.get("class") ==  
["shared-line-bubble"])
```

```
post_size = post_bedbathsize[1].find("b")
```

```
bedroom_count = float(post_bedbathsize[0].find_all("b")[0].text[:-2])
```

```
bathroom_count = float(post_bedbathsize[0].find_all("b")[1].text[:-2])
```

```
post_address = lists.find("div", class_ = "mapaddress").text
```



Result

Result – Welcome page

- Users could select several features of their ideal house to find the most suitable and optimal houses or apartments near Carnegie Mellon campus.

```
=====Welcome to LIFE IS A STRUGGLE USE PYTHON=====  
====Find ideal place to live in Shadyside, North Oakland and Squirrel Hills=====
```

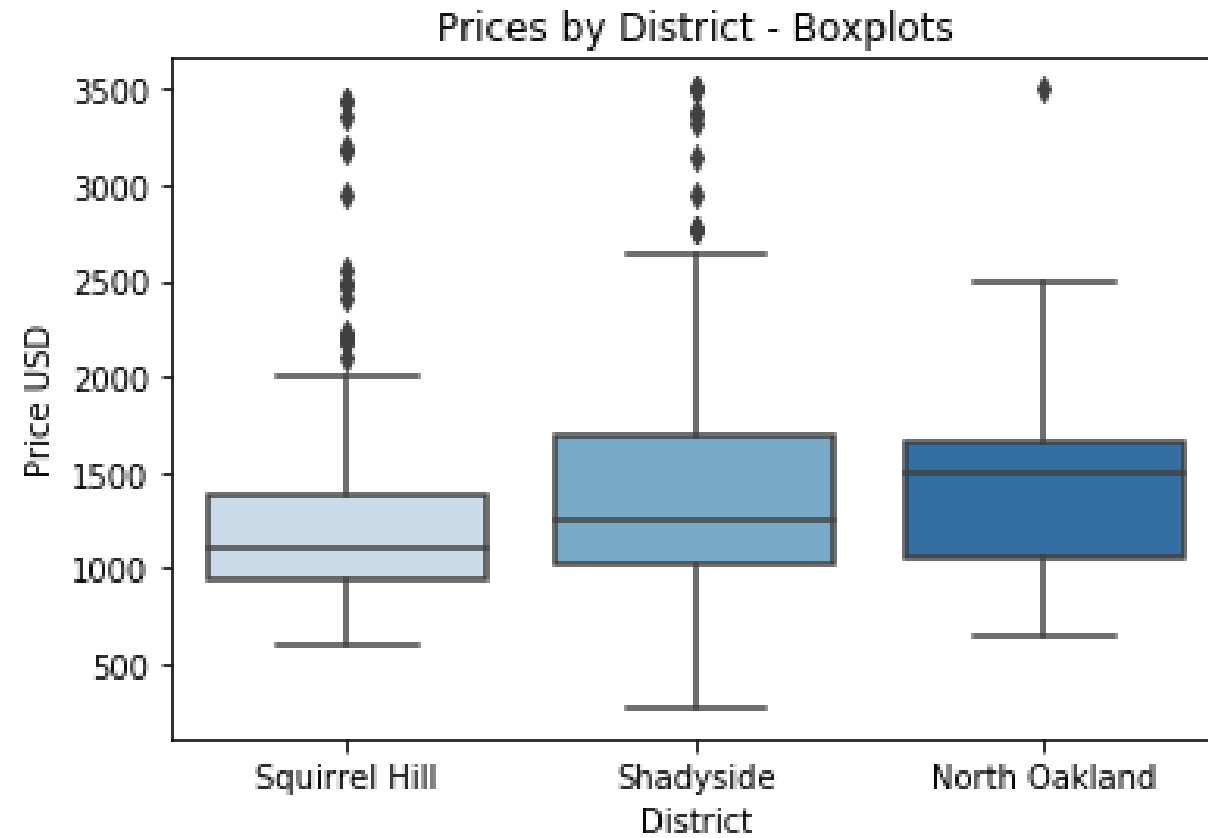
Result – Statistical information

- First, our menu would ask if the users want to show the statistical information on the housing in Pittsburgh.

Do you want some statistical information on the housing in Pittsburgh?(y/n): y

	Counts	Average	Highest	Lowest
=====				
Squirrel Hill	497	878.64	1779.50	300.00
North Oakland	99	702.67	1300.00	339.00
Shadyside	499	1036.95	2444.00	137.50
Overall	1095	934.87	2444.00	137.50

Result – Statistical information



Result – Statistical information

Do you want more details on the effects of the factors?(y/n): y

OLS Regression Results

```
=====
Dep. Variable:          Price      R-squared:          0.268
Model:                  OLS        Adj. R-squared:       0.262
Method:                 Least Squares  F-statistic:       40.07
Date:                  Mon, 09 Dec 2019  Prob (F-statistic): 1.50e-67
Time:                  18:25:09      Log-Likelihood:    -12605.
No. Observations:      1103         AIC:              2.523e+04
Df Residuals:          1092         BIC:              2.529e+04
Df Model:              10
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-4405.6927	2057.554	-2.141	0.032	-8442.900	-368.486
Beds	-6427.2640	918.842	-6.995	0.000	-8230.160	-4624.368
Baths	4844.2266	2103.532	2.303	0.021	716.805	8971.648
Size	15.9951	1.060	15.096	0.000	13.916	18.074
Distance	-455.8332	61.401	-7.424	0.000	-576.310	-335.356
NorthOakland	-435.0589	2085.103	-0.209	0.835	-4526.321	3656.203
Shadyside	-775.0989	1133.619	-0.684	0.494	-2999.416	1449.218
SquirrelHill	-3195.5349	1098.148	-2.910	0.004	-5350.255	-1040.815
Apartment	-454.5564	1546.214	-0.294	0.769	-3488.443	2579.331
Condo	-1.05e+04	1.6e+04	-0.657	0.511	-4.19e+04	2.09e+04
House	3.691e+04	4158.221	8.876	0.000	2.88e+04	4.51e+04
Townhomes	-4190.5436	8090.432	-0.518	0.605	-2.01e+04	1.17e+04

```
=====
Omnibus:                2043.879      Durbin-Watson:          1.661
Prob(Omnibus):          0.000         Jarque-Bera (JB):       4367436.408
Skew:                   12.732         Prob(JB):               0.00
Kurtosis:               310.216        Cond. No.                1.16e+19
=====
```

Result – Search page

- Second, our menu would ask if the users want more detail, if the users select y, we would display some features that users could choose to find their ideal house/apartment

Start a new search! Enter ALLDONE to quit.

Select the neighborhood:

1. Shadyside
2. North Oakland
3. Squirrel Hill
4. All the three areas

Please enter your choice: 1

Select the housing type:

1. House
2. Apartments
3. Condos
4. Townhomes
5. All the above types

Please enter your choice: 1

Enter the your acceptable minimum housing price: 0

Enter the your affordable maximum housing price: 9999

How do you want to sort the results

1. Price (high to low)
2. Price (low to high)
3. Distance to CMU (in miles)

Please enter your choice: 1

Result –Result page

Find 45 places in total.
Here are the top 5 place for you

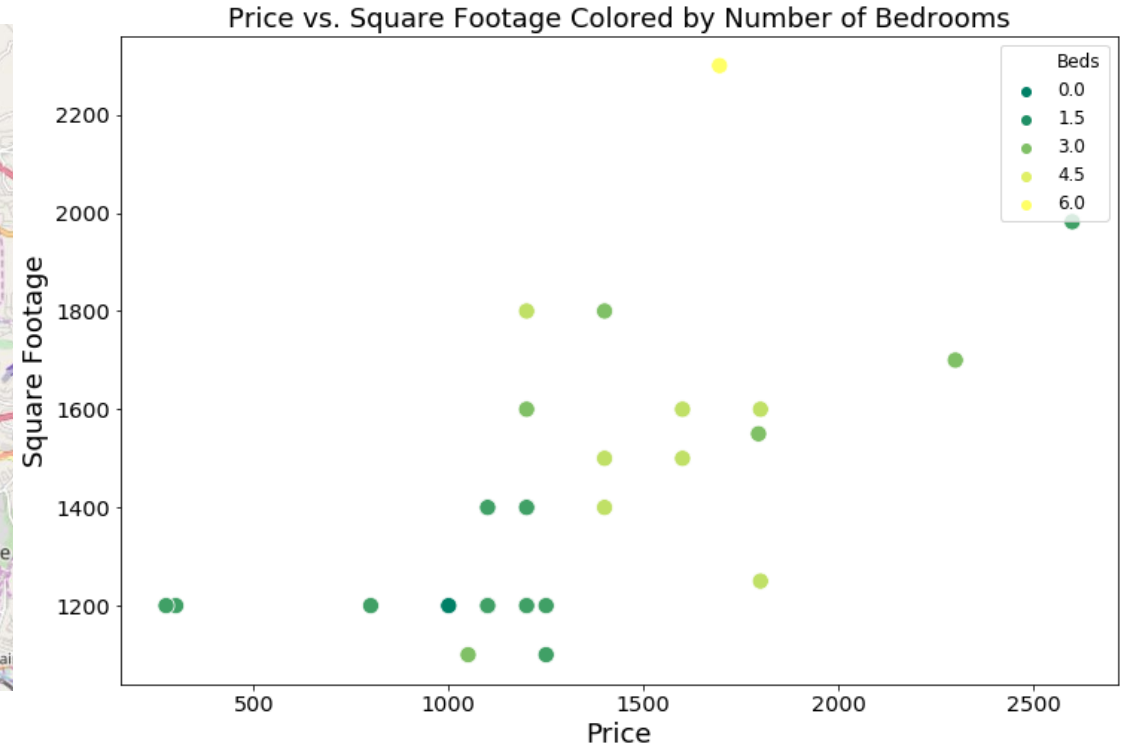
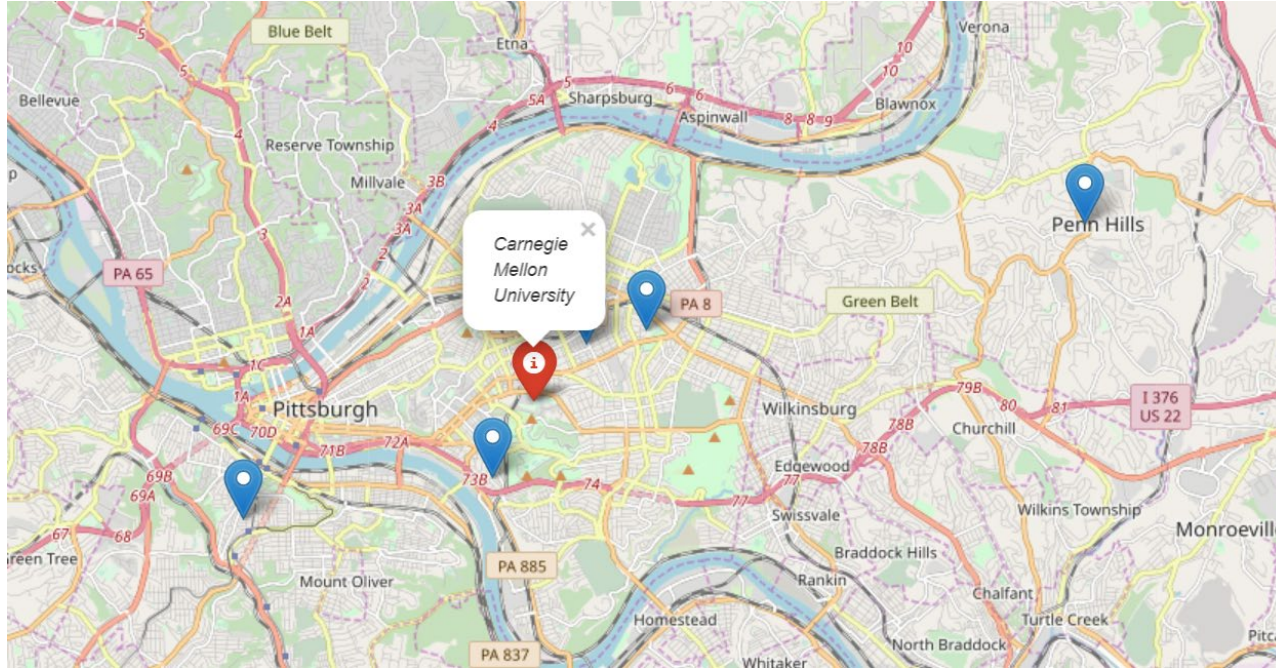
	Address	District	Beds	Baths	Price	Size(sqft)	Website	Distance(miles)	Type
0	Aiken	Shadyside	3.0	1.5	2300.0	1700.0	https://pittsburgh.craigslist.org/apa/d/pittsburgh-3br-house-for-rent-close-to/7029194632.html	1.3	House
1	234 Albert St near Boggs Ave.	Shadyside	2.0	1.0	1100.0	1200.0	https://pittsburgh.craigslist.org/apa/d/pittsburgh-spacious-2br-home-in-mt/7015640569.html	7.3	House
2	3232 Parkview Ave	Shadyside	4.0	2.0	1800.0	1250.0	https://pittsburgh.craigslist.org/apa/d/pittsburgh-beautiful-4-br-2-bath-house/7028961521.html	1.1	House
3	6426 Howe Street near Festival Street	Shadyside	2.0	2.5	2600.0	1982.0	https://pittsburgh.craigslist.org/apa/d/pittsburgh-quiet-cottage-from-january/7027490530.html	2.8	House
4	crescent hills road	Shadyside	3.0	1.5	1795.0	1550.0	https://pittsburgh.craigslist.org/apa/d/pittsburgh-good-credit-income-rent-to/7016494828.html	13.7	House

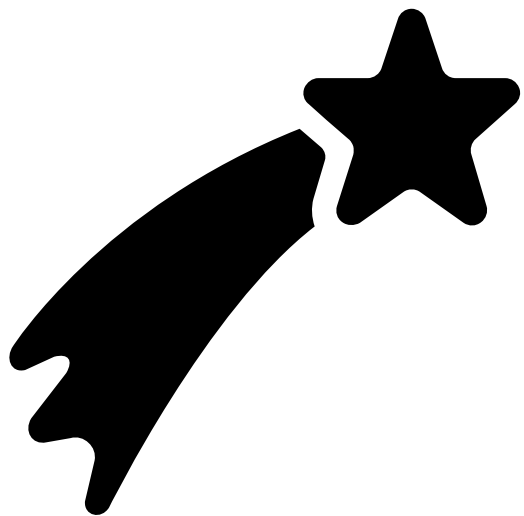
Result –Result page

Do you want the whole list(y/n): y

	Address	District	Beds	Baths	Price	Size(sqft)	Website	Distance(miles)	Type
0	Aiken	Shadyside	3.0	1.5	2300.0	1700.0	https://pittsburgh.craigslist.org/apad/pittsburgh-3br-house-for-rent-close-to/7029194632.html	1.3	House
1	234 Albert St near Boggs Ave.	Shadyside	2.0	1.0	1100.0	1200.0	https://pittsburgh.craigslist.org/apad/pittsburgh-spacious-2br-home-in-mt/7015640569.html	7.3	House
2	3232 Parkview Ave	Shadyside	4.0	2.0	1800.0	1250.0	https://pittsburgh.craigslist.org/apad/pittsburgh-beautiful-4-br-2-bath-house/7028961521.html	1.1	House
3	6426 Howe Street near Festival Street	Shadyside	2.0	2.5	2600.0	1982.0	https://pittsburgh.craigslist.org/apad/pittsburgh-quiet-cottage-from-january/7027490530.html	2.8	House
4	crescent hills road	Shadyside	3.0	1.5	1795.0	1550.0	https://pittsburgh.craigslist.org/apad/pittsburgh-good-credit-income-rent-to/7016494828.html	13.7	House
5	4030 Hoosac St. near Hoosac and Greenfield	Shadyside	2.0	1.0	1250.0	1200.0	https://pittsburgh.craigslist.org/apad/pittsburgh-large-story-house-with-2-br/7019692033.html	0.5	House
6	1502 Kenberma near Cape May	Shadyside	5.0	2.5	1695.0	2300.0	https://pittsburgh.craigslist.org/apad/pittsburgh-this-home-should-be-on-hgtv/7024145833.html	8.8	House
7	Near I 279 near RT 19 North	Shadyside	3.0	1.0	1400.0	1800.0	https://pittsburgh.craigslist.org/apad/pittsburgh-renovated-3-or-4-bedroom/6995049196.html	7.9	House
8	1316 Center St.	Shadyside	2.0	1.5	1100.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-2-bedroom-sleeps-3-fully/7030593132.html	89.2	House
9	1316 Center St.	Shadyside	3.0	1.5	1200.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7030392579.html	89.2	House
10	405 Clinton Ave.	Shadyside	4.0	2.0	1600.0	1600.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-3-4/7029439405.html	88.8	House
11	1807 Center St.	Shadyside	2.0	1.5	1100.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-2-and-3-br-all-inclusive/7028919694.html	88.7	House
12	1807 Center St.	Shadyside	2.0	1.5	1100.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-2-3/7028015880.html	88.7	House
13	405 Clinton Ave.	Shadyside	2.0	1.5	1100.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-1-bedroom-fully-furn-with/7027398942.html	88.8	House
14	405 Clinton Ave.	Shadyside	4.0	2.0	1200.0	1800.0	https://wheeling.craigslist.org/apad/shadyside-4-br-fully-furn-house-for/7027387917.html	88.8	House
15	1807 Center St.	Shadyside	2.0	1.5	1100.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7025921409.html	88.7	House
16	1807 Center St.	Shadyside	2.0	1.5	1100.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-2-3/7024903606.html	88.7	House
17	1807 Center St.	Shadyside	1.0	1.5	1100.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-1-bedroom-fully-furn-with/7024325456.html	88.7	House
18	1807 Center St.	Shadyside	2.0	1.5	1200.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7023691990.html	88.7	House
19	1807 Center St.	Shadyside	2.0	1.5	1200.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7023010556.html	88.7	House
20	1807 Center St.	Shadyside	2.0	1.5	1100.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers2/7022318626.html	88.7	House
21	1807 Center St.	Shadyside	2.0	1.5	1100.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7021827582.html	88.7	House
22	1807 Center St.	Shadyside	2.0	1.5	1200.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-2-3-and-4-br-all-inclusive/7021136034.html	88.7	House
23	1807 Center St.	Shadyside	2.0	1.5	1200.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7020473489.html	88.7	House
24	1807 Center St.	Shadyside	2.0	1.5	1100.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-2-3-and-4-br-all-inclusive/7019810772.html	88.7	House
25	1807 Center St.	Shadyside	2.0	1.5	1200.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-og-workers-br-available-ill/7019164331.html	88.7	House
26	405 Clinton Ave.	Shadyside	2.0	1.5	1100.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7019140362.html	88.8	House
27	1807 Center St.	Shadyside	1.0	1.0	1000.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-1-bedroom-fully-furn-with/7019135919.html	88.7	House
28	1807 Center St.	Shadyside	2.0	1.5	300.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7018504059.html	88.7	House
29	301 W 41st St near Adams	Shadyside	3.0	1.5	1050.0	1100.0	https://wheeling.craigslist.org/apad/shadyside-home-for-rent/7017478117.html	98.2	House
30	1807 Center St.	Shadyside	2.0	1.5	275.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-br-all-inclusive-in/7016680878.html	88.7	House
31	1807 Center St.	Shadyside	2.0	1.5	1200.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-2-bedroom-sleeps-3-fully/7015351605.html	88.7	House
32	405 Clinton Ave.	Shadyside	4.0	1.0	1600.0	1500.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-4-br/7015339587.html	88.8	House
33	405 Clinton Ave.	Shadyside	3.0	1.5	1200.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7014618209.html	88.8	House
34	1807 Center St.	Shadyside	2.0	1.5	1200.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-2-3/7014078457.html	88.7	House
35	405 Clinton Ave.	Shadyside	4.0	1.0	1400.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-4-br/7013975067.html	88.8	House
36	405 Clinton Ave.	Shadyside	3.0	1.5	1200.0	1600.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-br/7013202667.html	88.8	House
37	405 Clinton Ave.	Shadyside	4.0	1.0	1400.0	1500.0	https://wheeling.craigslist.org/apad/shadyside-att-construction-workers-4-br/7012805539.html	88.8	House
38	1807 Center St.	Shadyside	2.0	1.5	1200.0	1400.0	https://wheeling.craigslist.org/apad/shadyside-2-bedroom-sleeps-3-fully/7011985635.html	88.7	House
39	203 7th Street	Shadyside	4.0	2.0	1600.0	1600.0	https://wheeling.craigslist.org/apad/shadyside-4-bedroom-2-bath-fully/7011292012.html	90.7	House
40	1807 Center St.	Shadyside	2.0	1.5	1200.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-2-and-3-br-all-inclusive/7010046460.html	88.7	House
41	203 7th Street	Shadyside	4.0	4.0	1800.0	1600.0	https://wheeling.craigslist.org/apad/shadyside-4-bedroom-2-bath-fully/7007221721.html	90.7	House
42	1316 Center St.	Shadyside	2.0	1.5	800.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-2-and-3-br-all-inclusive/7006684079.html	89.2	House
43	1316 Center St.	Shadyside	2.0	1.5	1200.0	1200.0	https://wheeling.craigslist.org/apad/shadyside-2-3-and-4-br-all-inclusive/7004202136.html	89.2	House

Result –Result page





Future

Future

- Scarping more information from other housing rental housing websites
- Perform further analysis
- Building a more interactive user interface



Thanks
