



Battle of Neighbourhoods

Identifying interesting places to live, using data and analytics...



Motivation

When you want to find a house/place to live, one turns to Redfin, Zillow, Foursquare and

- Looks at housing prices
- Looks at interesting places/activities around the area - nightlife, bars, etc.
- Looks at schools and proximity to work, etc.

But there are more:

- How do we look for part-time job opportunities in grocery shops, supermarkets, etc.
- How do we look for specific interests like special needs schools



Drawbacks of current applications

- Provides the usual features for filtering (home prices, sq. ft. area, etc.)
- Provides info on interesting activities, places nearby (restaurants, bars, gyms, etc.)
- But ...
 - Hardly provides other specific categories
 - Filtering data based on specific interests forces one to go to multiple websites
 - One has to merge the disparate sources of information oneself



Objectives of the work

An attempt to:

1. Use Foursquare API to collect interesting places based on specific interests
2. Use Redfin dataset to try to estimate housing prices at specific geographic locations

Merge 1. and 2. to arrive at a consolidated application that one can use to collect interesting places from 1. and estimate home prices at interesting places using 2.

Data Sources

Redfin data

The screenshot shows the Redfin website interface. At the top, there's a navigation bar with the Redfin logo and a search bar containing "City, Address, School, A...". Below the search bar, there's a map of the San Francisco Bay Area with a red pin in Livermore. To the right of the map, there's a section titled "Livermore Homes for Sale" with a "Listings" tab selected. Below the tab, there's a table showing 21 of 1776 homes. The table has columns for Price, No min, to, No max, Filters, Photos, and Table. The first three rows of the table are:

Price	No min	to	No max	Filters	Photos	Table
35145 Adriano St	\$1100,000				4	3
5236 Ironwood Ln	\$490,000				3	2
193 Misty Cir	\$869,900				3	3

At the bottom of the page, there's a footer with "Map data ©2021 Google" and "Terms of Use".

Foursquare API

FOURSQUARE /developers

Blog

Docs

Log In

Contact Us

Documentation

Places API Reference

Venues / Get Venue Categories

Search Documentation

Places API Reference

Venues

- Search for Venues
- Get Venue Recommendations
- Get Details of a Venue
- Report Venue Selection
- Get a Venue's Photos
- Get a Venue's Tips
- Get a Venue's Hours
- Get a Venue's Menu
- Get a Venue's Links
- Get Trending Venues
- Suggest Completion
- Get Users Who Liked a Venue

Get Venue Categories

Description

- Request
- Authentication
- Response Fields
- Response
- Get Similar Venues
- Get Events at a Venue
- Get Next Venues
- Get Lists a Venue is On

Venue Categories

Description

Returns a hierarchical list of categories applied to venues. This list is also available on our [categories](#) page.

Request

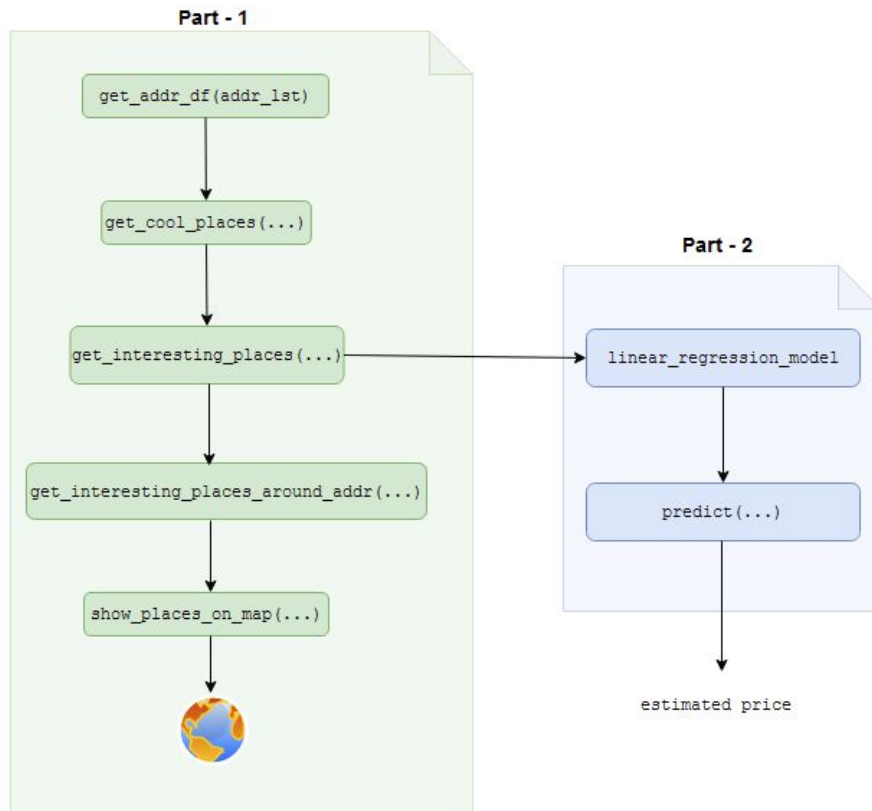
GET <https://api.foursquare.com/v2/venues/categories>

Authentication

This endpoint supports **userless** or **user** authentication.

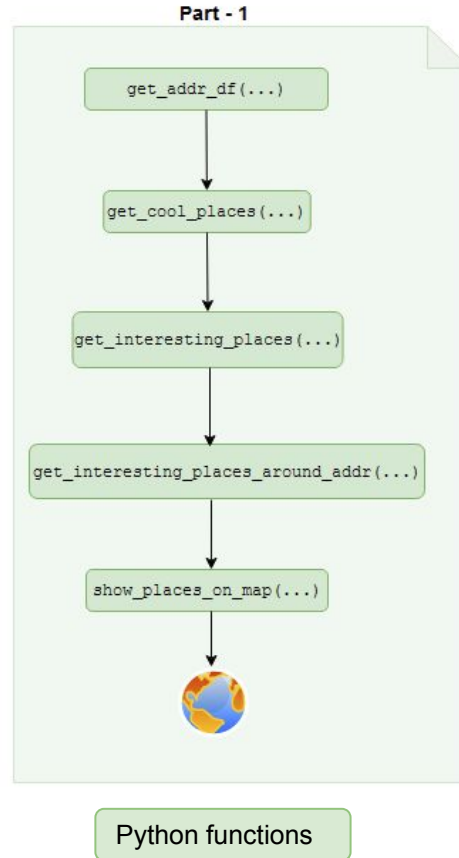
Supported	Description
Userless	required unless OAuth token is provided. A valid Client ID and Secret in the query string of each request. Eg. <code>{client_id=XXXX&client_secret=XXXX}</code>
User	required unless Client ID and Secret are provided. A valid OAuth access token in the query string of each request. Eg. <code>{oauth_token=XXXX}</code>

Architecture



Part - 1

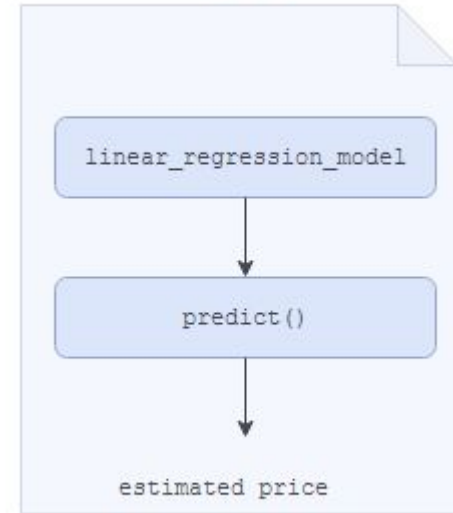
- Foursquare API
- Folium (for map)
- Custom functions



Part - 2

- Redfin dataset
- Scikit-learn
 - Imputing, encoding the data
 - Linear regression
- Custom python functions

Part - 2



Python functions



Snapshots

```
my_addr_lst = ['927 Jackson St, Mountain View, CA 94043'  
, '2200 Monroe St, Santa Clara, CA'  
, '668 Madrone Ave, Sunnyvale, CA'  
, '10755 Tressler Ct, Cupertino, CA'  
, '328 Hill St, San Francisco, CA 94114'  
, '2424 Palm Ave, Redwood City, CA 94061'  
, '3159 Kenland Dr, San Jose, CA 95111'  
, '916 Cameron Cir, Milpitas, CA 95035'  
, '4041 Crestwood St, Fremont, CA 94538'  
, '4400 Canterbury Way, Union City, CA 94587'  
, '2219 Curtis St, Oakland, CA 94607'  
, '1801 University Ave, Berkeley, CA 94703'  
, '2421 Lincoln Ave, Richmond, CA 94804'  
, '1415 Speers Ave, San Mateo, CA 94403'  
, '855 Standish Rd, Pacifica, CA 94044'  
, '1170 Paula Dr, Campbell, CA 95008'  
, '222 Johnson Ave Los Gatos, CA 95030']
```

get_addr_df(addr_lst)

	Address	Latitude	Longitude
0	927 Jackson St, Mountain View, CA 94043	37.398051	-122.079750
1	2200 Monroe St, Santa Clara, CA	37.335526	-121.943685
2	668 Madrone Ave, Sunnyvale, CA	37.393987	-122.025583
3	10755 Tressler Ct, Cupertino, CA	37.312231	-122.060524
4	328 Hill St, San Francisco, CA 94114	37.755851	-122.428753

Snapshots

	Address	Latitude	Longitude
0	927 Jackson St, Mountain View, CA 94043	37.398051	-122.079750
1	2200 Monroe St, Santa Clara, CA	37.335526	-121.943685
2	668 Madrone Ave, Sunnyvale, CA	37.393987	-122.025583
3	10755 Tressler Ct, Cupertino, CA	37.312231	-122.060524
4	328 Hill St, San Francisco, CA 94114	37.755851	-122.428753

`get_cool_places(...)`

	Anchor_Address	Anchor_Latitude	Anchor_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Address	Venue_Category
5	927 Jackson St, Mountain View, CA 94043	37.398051	-122.07975	Zareen's	37.426834	-122.144179	365 California Ave	Indian Restaurant
14	927 Jackson St, Mountain View, CA 94043	37.398051	-122.07975	Luna Mexican Restaurant	37.333935	-121.915180	1495 The Alameda	Mexican Restaurant
16	927 Jackson St, Mountain View, CA 94043	37.398051	-122.07975	Ping's Bistro 留湘	37.575651	-122.044134	34145 Fremont Blvd	Hunan Restaurant

Snapshots

	Anchor_Address	Anchor_Latitude	Anchor_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Address	Venue_Category
1451	3159 Kenland Dr, San Jose, CA 95111	37.28712	-121.842567	Souvlaki Greek Skewers	37.309358	-121.887078	577 W Alma Ave	Greek Restaurant
1452	3159 Kenland Dr, San Jose, CA 95111	37.28712	-121.842567	Gio Cha Duc Huong Sandwich	37.331458	-121.854906	1020 Story Rd	Vietnamese Restaurant
1454	3159 Kenland Dr, San Jose, CA 95111	37.28712	-121.842567	Pho Hà Nội	37.332473	-121.858145	San Jose, CA 95122	Vietnamese Restaurant
1455	3159 Kenland Dr, San Jose, CA 95111	37.28712	-121.842567	Luna Mexican	37.333935	-121.915180	1495 The	Mexican

```
my_new_addr = '3159 Kenland Dr, San Jose, CA 95111'  
my_restrictions = ['Vegan', 'Vegetarian']  
my_food_places = get_interesting_places_around_addr(my_new_addr, cool_places, my_restrictions)
```

	Anchor_Address	Anchor_Latitude	Anchor_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Address	Venue_Category
1632	3159 Kenland Dr, San Jose, CA 95111	37.28712	-121.842567	Blossom Vegan Restaurant	37.69981	-121.869719	4000 Pimlico Dr #112	Vegetarian / Vegan Restaurant

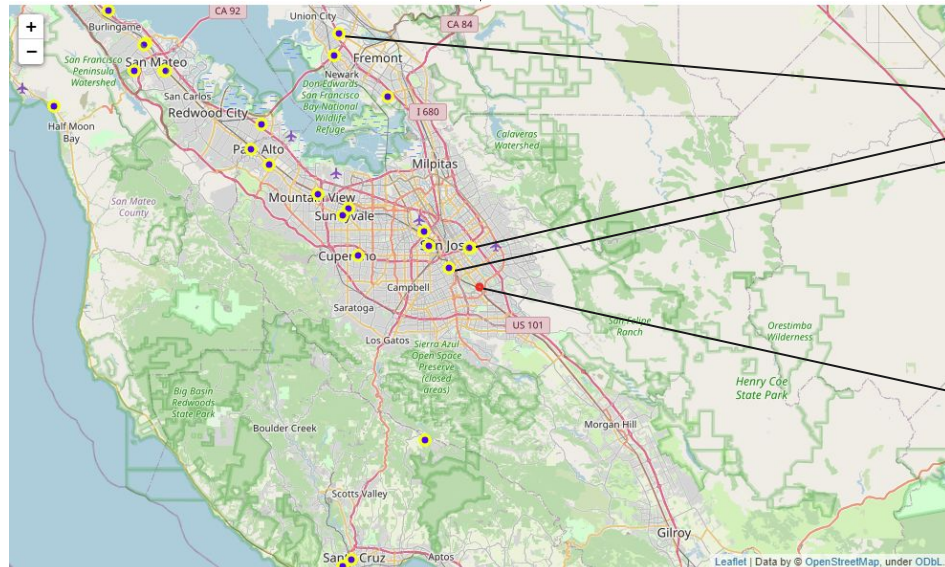
Snapshots

```
my_work_categories = ['Restaurant', 'market']
```

```
my_work_places = get_interesting_places(cool_places, my_work_categories)
```

```
my_new_addr = '3159 Kenland Dr, San Jose, CA 95111'
```

```
map = show_places_on_map(my_new_addr, my_work_places)  
map
```



My_work_places
(blue, yellow)

My_new_addr
(red)



Snapshots

Preprocessed, prepared Redfin data with listed prices as labels



```
lr = LinearRegression()  
lr.fit(training_prepared, training_labels)
```



Linear Regression model

Snapshots

```
my_addr = 'Monroe St, Santa Clara, CA'
beds = [2.0, 3.0, 4.0]
baths = [2.0, 2.5, 2.0]
property_types = ['Single Family Residential', 'Condo/Co-op', 'Multi-Family (2-4 Unit)',
                  'Townhouse', 'Mobile/Manufactured Home', 'Multi-Family (5+ Unit)']
```

```
my_df = create_fake_df(1000000, beds, baths, my_addr, property_types)
```

(1000000 is just a placeholder value for budget)

index	property_type	zipcode	beds	baths	sq_ft	year_built	lot_size	latitude	longitude	price
0	Single Family Residential	95051	2.0	2.0	2490	1963	4347	37.364014	-121.968601	1000000
1	Single Family Residential	95051	2.0	2.5	3698	1969	3817	37.364014	-121.968601	1000000
...										
...										
52	Multi-Family (5+ Unit)	95051	4.0	2.5	3501	1992	4472	37.364014	-121.968601	1000000
53	Multi-Family (5+ Unit)	95051	4.0	2.0	2492	1978	4220	37.364014	-121.968601	1000000

Preprocess, prepare

```
full_pipeline.fit_transform(...)
```

```
# A specific house with a specific characteristic
rand_predictions = lr.predict(rand_test_prepared[2:3])
print(rand_predictions)
```

[1460456.41893491]

Estimated price



Possible future extensions

- Get more details from housing data (school district boundaries, number/types of jobs, etc.)
- Get demographic data in neighbourhoods (diversity, etc.)
- Use GPU accelerated libraries like Tensorflow to do Deep Learning on the data
- Obtain details of Venue sub-categories from Foursquare API