Applied Data Science Capstone

A brief idea of a renting system construction for workers (interns)

Changfa FU

17.02.2019

Introduction

As it becomes a trend to have an **internship** for college student in the United States. So let us assume that there are a group of students who wants to work in Washington DC and look for **places** to live. And they have **requirements** as following:

- surrounded by restaurants
- not far away from workplace
- a balanced life with regards to work and relax

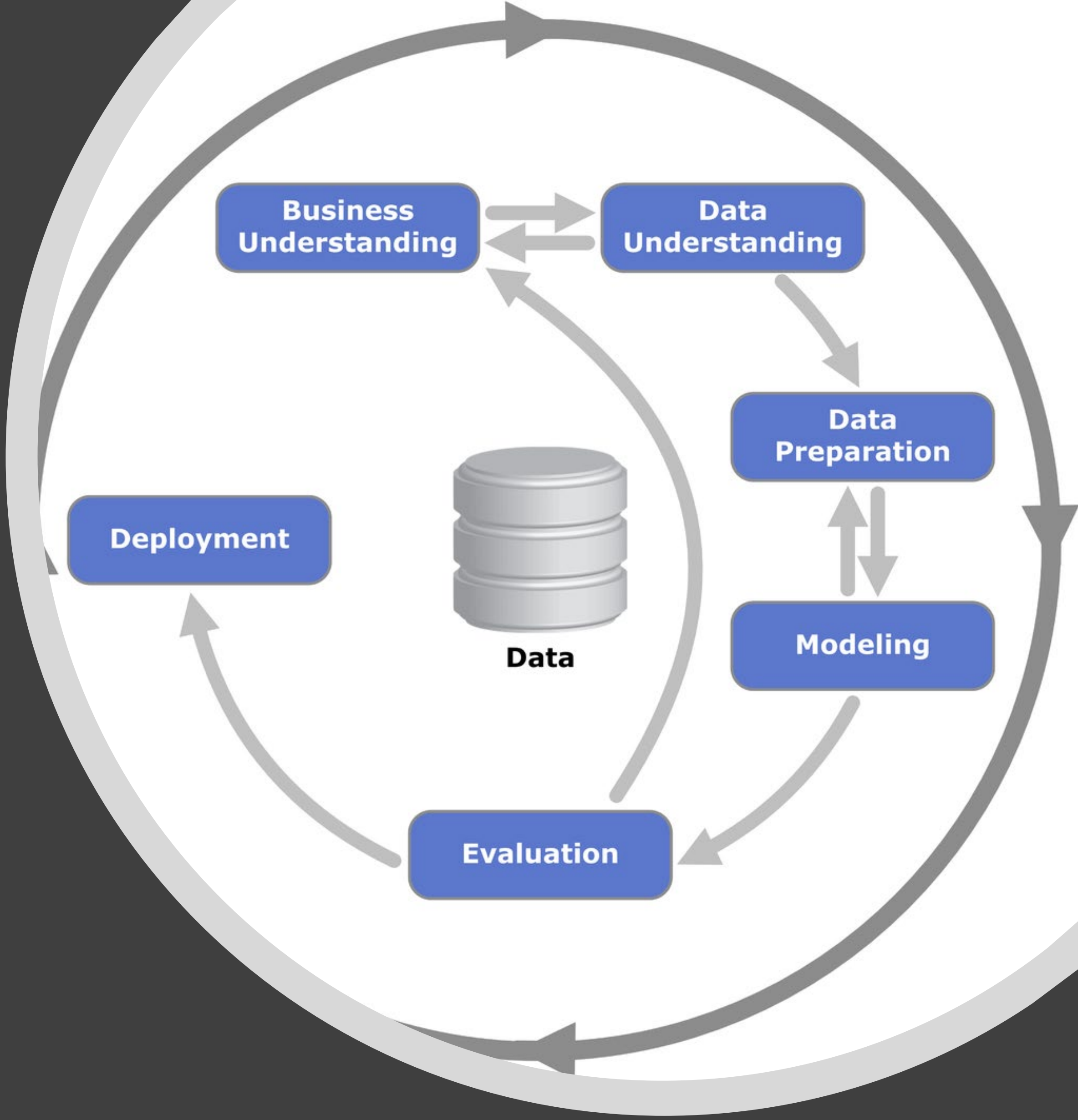
- Therefore, where will be the best or preferred location for them to settle down?
- It is where the challenge comes from!!!

Data

- In order to find best place to locate for interns, we will need their **company location** (Foursquare API);
- Meanwhile, Foursquare location data will be used to explore the **surrounding** area of the company with set radius;
- After **cluster** the neighborhood and determine the region of **settlement** compared to DC address points seeking for **availability**
- Foursquare API accessing
- https://opendata.arcgis.com/datasets/aa514416aaf74fdc94748f1e56e7cc8a_0.csv?session=361463279.1550275769

Methodology

- **CRISP-DM – Cr**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining
- DISCRIPTIVE => EXPLORING
    - **Locating**
    - **Exploring**
    - **Determining**
- Highly depends on the data from Foursquare API and D.C. Address Points



Locating

```
[*]: company=input('Which company are you going to intern for:')
```

Which company are you going to intern for: KPMG

```
[11]: url_c = 'https://api.foursquare.com/v2/venues/search?&client_id={}&client_secret={}&v={}&ll={},{}&query={}'.format(
          CLIENT_ID,
          CLIENT_SECRET,
          VERSION,
          38.8976797,
          -77.0365191,
          company)
```

```
[12]: results = requests.get(url_c).json()
      results
```

Results - Locating


```
{'meta': {'code': 200, 'requestId': '5c698ca7dd57977bd434296a'},
 'response': {'venues': [{'id': '4f03018d6da1cd035bcc7484',
    'name': 'KPMG',
    'location': {'address': '1801 K Street NW',
     'crossStreet': 'btwn 18th St & 19th St NW',
     'lat': 38.90317883226485,
     'lng': -77.04223495398159,
     'labeledLatLngs': [{'label': 'display',
       'lat': 38.90317883226485,
       'lng': -77.04223495398159}],
     'distance': 787,
     'postalCode': '20006',
     'cc': 'US',
     'city': 'Washington',
     'state': 'D.C.',
     'country': 'United States',
     'formattedAddress': ['1801 K Street NW (btwn 18th St & 19th St NW)',
      'Washington, D.C. 20006',
      'United States']},
    'categories': [{'id': '4bf58dd8d48988d124941735',
      'name': 'Office',
      'pluralName': 'Offices',
      'shortName': 'Office',
      'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/building/default_',
       'suffix': '.png'},
      'primary': True}],
    'referralId': 'v-1550421159',
    'hasPerk': False},
```

Exploring

```
[13]: radius=float(input('How far away from your company can you stand for commuting (in meters):'))
```

```
How far away from your company can you stand for commuting (in meters): 2000
```

```
[14]: number=float(input('How many venues in the neighbourhood do you want to check at a glance:'))
```

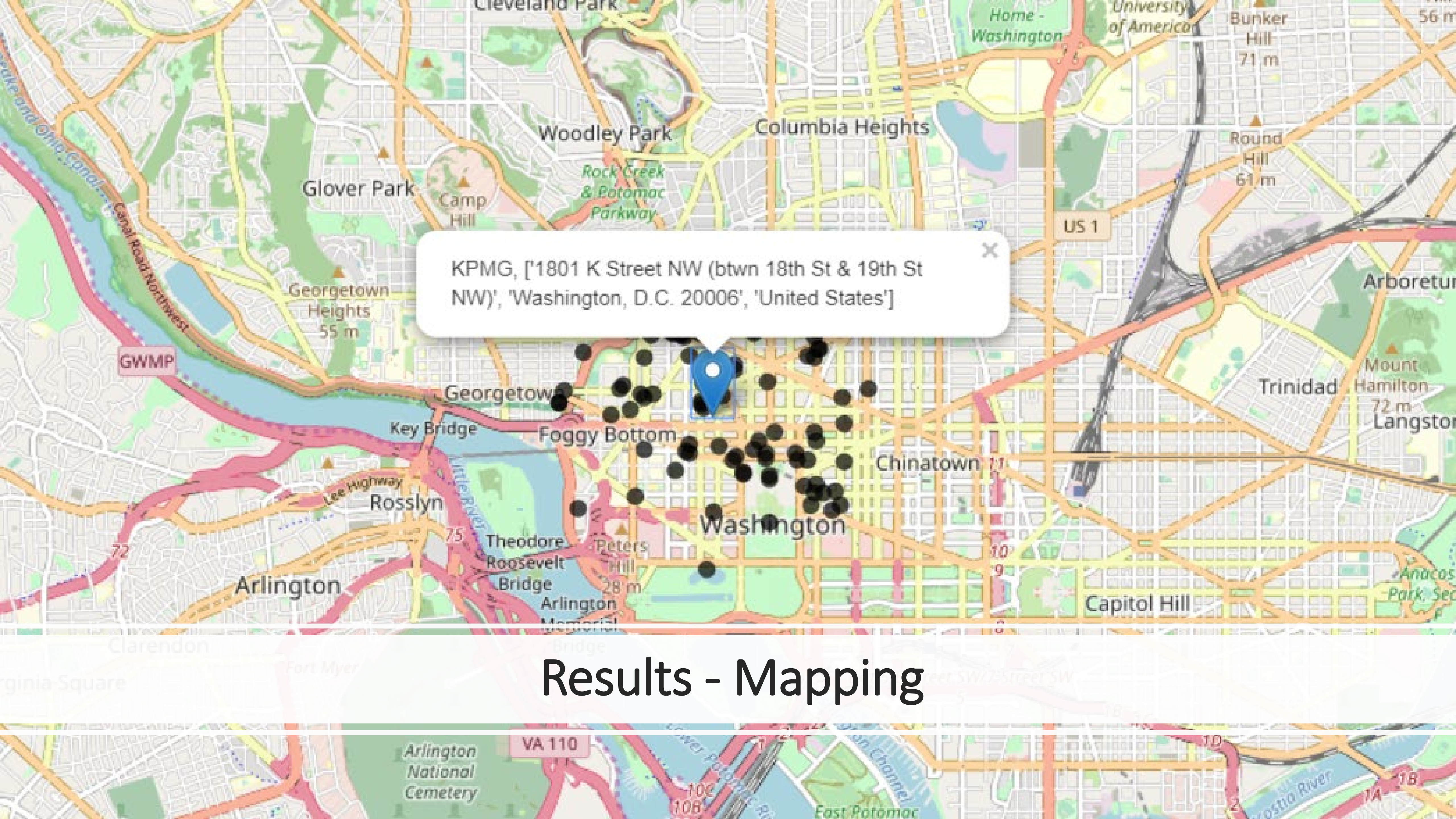
```
How many venues in the neighbourhood do you want to check at a glance: 100
```

```
[17]: url_s = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
          CLIENT_ID,
          CLIENT_SECRET,
          VERSION,
          lat_,
          lng_,
          radius,
          number)
```

```
[18]: results = requests.get(url_s).json()
      results
```

Results - Exploring

	postalCode	name	categories	lat	lng
0	20036	Wawa	Convenience Store	38.904303	-77.043356
1	20036	DC Pizza	Pizza Place	38.904038	-77.043186
2	20006	Compass Coffee	Coffee Shop	38.901041	-77.041680
3	20036	Greek Deli & Catering	Greek Restaurant	38.904517	-77.043428
4	20036	DC Improv Comedy Club	Comedy Club	38.904894	-77.041208
5	20036	Bub and Pop's	Sandwich Place	38.905712	-77.042335
6	20036	Boqueria	Spanish Restaurant	38.905921	-77.043140
7	20006	Founding Farmers	American Restaurant	38.900573	-77.044519
8	20006	Filter Coffeehouse & Espresso Bar	Coffee Shop	38.901133	-77.044608
9	20006	AKA White House	Hotel	38.899970	-77.040290
10	20036	Shake Shack	Burger Joint	38.906512	-77.041893
11	20006	The Bombay Club	Indian Restaurant	38.900822	-77.038246
12	20005	BLT Steak	Steakhouse	38.901483	-77.037692
13	20036	CAVA	Mediterranean Restaurant	38.906639	-77.042132
14	20006	Bindaas	Indian Restaurant	38.900598	-77.044875
15	20006	Compass Coffee	Coffee Shop	38.900199	-77.039970
16	20500	Renwick Gallery	Art Museum	38.898962	-77.039189



Results - Mapping

Clustering (into 6)

[33]:
```
company_merged.loc[company_merged['Cluster Labels'] == 0, company_merged.columns[[1] + list(range(5, company_merged.shape[1]))]]
```
...

[34]:
```
company_merged.loc[company_merged['Cluster Labels'] == 1, company_merged.columns[[1] + list(range(5, company_merged.shape[1]))]]
```
...

[35]:
```
company_merged.loc[company_merged['Cluster Labels'] == 2, company_merged.columns[[1] + list(range(5, company_merged.shape[1]))]]
```
...

[36]:
```
company_merged.loc[company_merged['Cluster Labels'] == 3, company_merged.columns[[1] + list(range(5, company_merged.shape[1]))]]
```
...

[37]:
```
company_merged.loc[company_merged['Cluster Labels'] == 4, company_merged.columns[[1] + list(range(5, company_merged.shape[1]))]]
```
...

[38]:
```
company_merged.loc[company_merged['Cluster Labels'] == 5, company_merged.columns[[1] + list(range(5, company_merged.shape[1]))]]
```
...

Results – Clustering 0

	name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Wawa	0.0	Hotel	Salad Place	Coffee Shop	Museum	Pizza Place	Sandwich Place	Mediterranean Restaurant	Steakhouse	Café	Gym / Fitness Center
1	DC Pizza	0.0	Hotel	Salad Place	Coffee Shop	Museum	Pizza Place	Sandwich Place	Mediterranean Restaurant	Steakhouse	Café	Gym / Fitness Center
2	Greek Deli & Catering	0.0	Hotel	Salad Place	Coffee Shop	Museum	Pizza Place	Sandwich Place	Mediterranean Restaurant	Steakhouse	Café	Gym / Fitness Center
3	DC Improv Comedy Club	0.0	Hotel	Salad Place	Coffee Shop	Museum	Pizza Place	Sandwich Place	Mediterranean Restaurant	Steakhouse	Café	Gym / Fitness Center
4	Compass Coffee	0.0	Hotel	Coffee Shop	Sandwich Place	Hotel Bar	Indian Restaurant	American Restaurant	Historic Site	Mexican Restaurant	Breakfast Spot	Café
5	Bub and Pop's	0.0	Hotel	Salad Place	Coffee Shop	Museum	Pizza Place	Sandwich Place	Mediterranean Restaurant	Steakhouse	Café	Gym / Fitness Center
6	Boqueria	0.0	Hotel	Salad Place	Coffee Shop	Museum	Pizza Place	Sandwich Place	Mediterranean Restaurant	Steakhouse	Café	Gym / Fitness Center
7	Kaz Sushi Bistro	0.0	Hotel	Coffee Shop	Sandwich Place	Hotel Bar	Indian Restaurant	American Restaurant	Historic Site	Mexican Restaurant	Breakfast Spot	Café

92 rows × 12 columns

Results – Clustering 1

	name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Con
98	Lafayette Square Park	1.0	Park	Yoga Studio	Hotel	Historic Site	Health & Beauty Service	Gym / Fitness Center	Gym	Grocery

Results – Clustering 2

	name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
50	Crepes Parfait	2.0	Food Truck	Yoga Studio	Thai Restaurant	History Museum	Historic Site	Health & Beauty Service	Gym / Fitness Center	Gym

Results – Clustering 3

	name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Cor
22	Nando's PERi-PERi	3.0	Portuguese Restaurant	Hotel	Historic Site	Health & Beauty Service	Gym / Fitness Center	Gym	Grocery Store	Rest

Results – Clustering 4

	name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Mo: Commo Venu
60	GWU University Yard	4.0	College Quad	Yoga Studio	Deli / Bodega	History Museum	Historic Site	Health & Beauty Service	Gym / Fitne: Cente

Results – Clustering 5

	name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Co
52	Renwick Gallery	5.0	Art Museum	Government Building	Yoga Studio	Deli / Bodega	History Museum	Historic Site	Health & Beauty Service	Gym /
95	The White House	5.0	Art Museum	Government Building	Yoga Studio	Deli / Bodega	History Museum	Historic Site	Health & Beauty Service	Gym /

Discussion

- Therefore, where should the student locate to meeting requirements above?
  - Recap of requirements:
    - surrounded by restaurants
      - **The majority are concentrated in cluster 0,**
    - not far away from workplace
      - **In the radius of 1000 m centered at the company**
    - a balanced life with regards to work and relax
      - **Depending on venues nearby**
      - **Cluster 0 gets the most**
- **THUS, it is quite evident that cluster 0 will be the best area for the student to settle in**
- Refer back to common postal code in cluster 0
- With help of dataset of D.C. Address Points, we can have recommendations for the student to find a preferred house

Conclusion

- Location Icon stands for the company
- Blue points are active housing options
- Black points are TOP 100 venues surrounded