

Where to find the best living location for interns in DC

Changfa FU

Introduction

Description:

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 Washinton 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

Problem:

Therefore, where will be the best or preferred location for them to settle down?

Interest:

Interns, companies or even rental agents will have interest in this.

Data

2.1 Data Source

1. Foursquare API

2. D.C. Address Points

(https://opendata.arcgis.com/datasets/aa514416aaf74fdc94748f1e56e7cc8a_0.csv?session=361463279.1550275769)

2.2 Data Clearing

For information from Foursquare API, only keep information of name, postal code, categories, latitude and longitude of venues

For D.C. Address Points, keep ADDRESS_ID, LATITUDE, LONGITUDE, STATUS, RES_TYPE, STNAME, ASSESSMENT_NBHD and ZIPCODE of each house. Meanwhile remove all data of 'STATUS'!=ACTIVE and 'RES_TYPE'=='NON RESIDENTIAL'.

Methodology

CRISP-DM – Cross Industry Standard Process for Data Mining

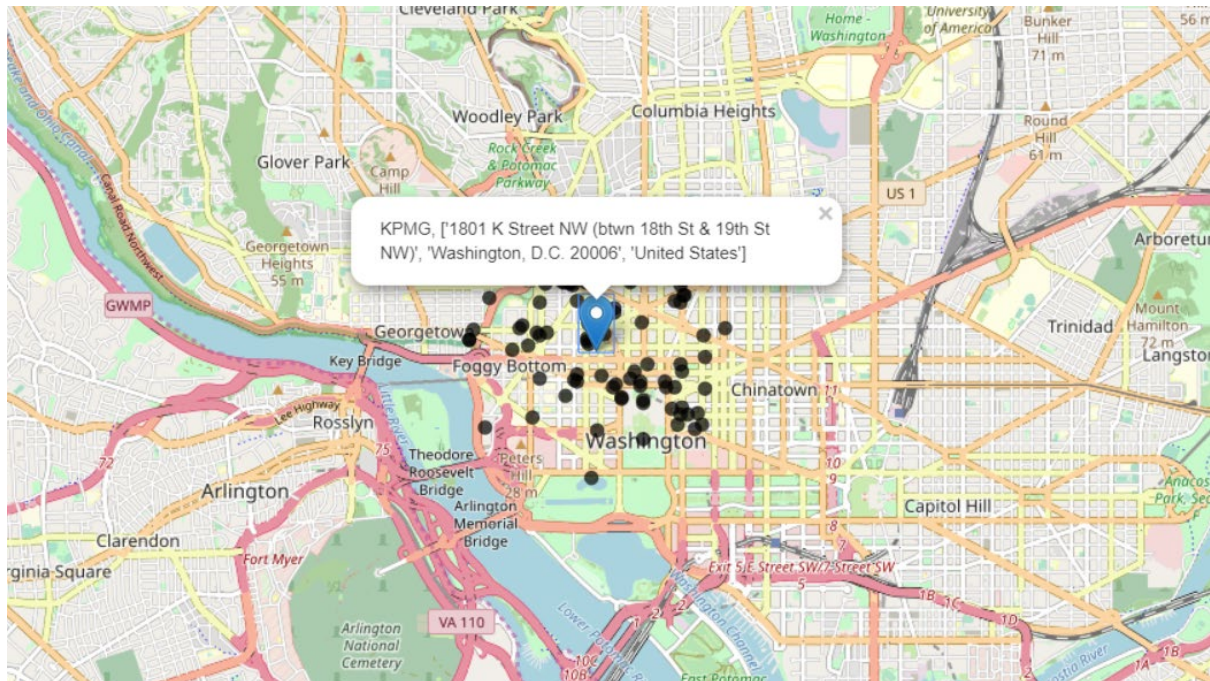
DISCRIPTIVE => EXPLORING

- Locating
- Exploring
- Determining (with clustering)

Highly depends on the data from Foursquare API and D.C. Address Points

In order to find best place to locate for interns, we will need their company location. Meanwhile, Foursquare location data will be used to get list of locations of restaurants & bars. After, the location of settlement has been chosen, Foursquare can help to give a best choice of restaurants amongst all in the neighbourhood

Results



Discussion

Find company location and all venues surrounded.

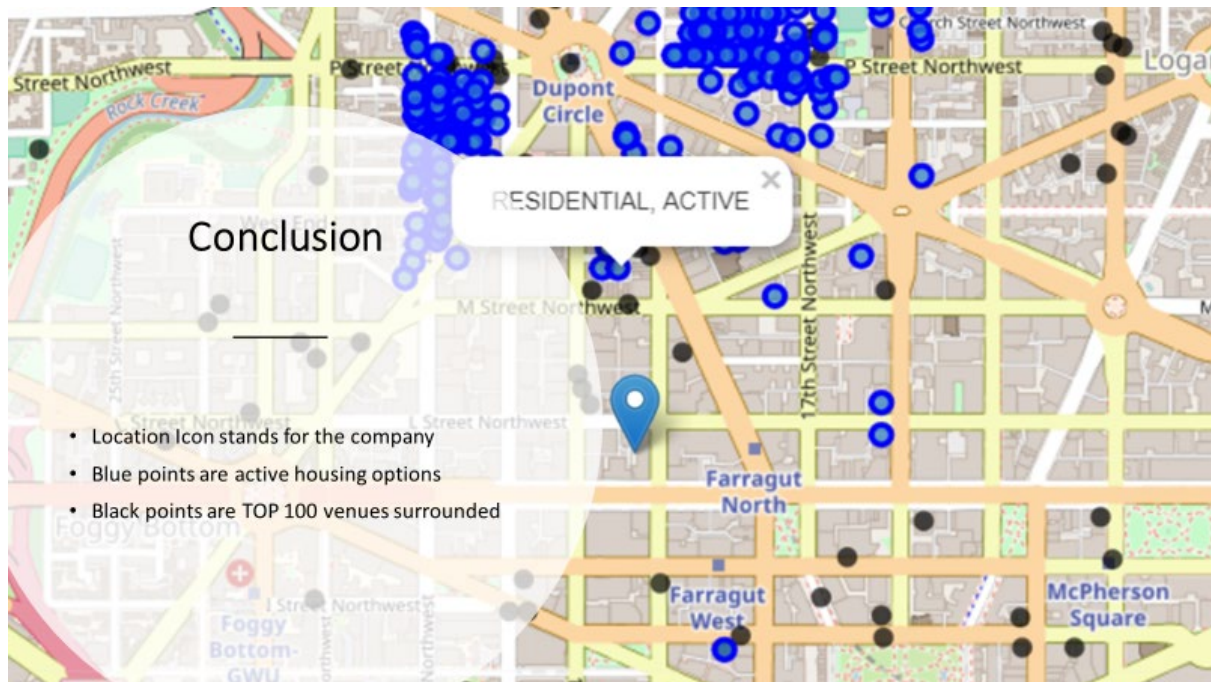
Clustering (into 6) all venues to find the preferred area for housing.

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



Implication and Improvement

This can be applied to rental/housing companies for their customers

This could be a tool for companies for well settle their employees and provide information

Python is not efficient enough to deal with large data sets