A description of the problem and a discussion of the background.

The **Washington D.C. Metropolitan Area** is the metropolitan area centered on Washington, D.C., the capital of the United States. The area includes all of the District of Columbia, and parts of the US States of Virginia, Maryland and West Virginia. Northern Virginia (locally referred to as NoVA or NOVA) comprises cities and counties in Virginia surrounding Washington D.C.

It is the most populous region of Virginia and Washington Metropolitan area. There are about 3.5million people working in various jobs in this area.

DC is home to all types of amenities and attractions. World famous Smithsonian Museums are based in D.C. D.C. Metro Area boasts of hundreds of Performing Arts/theaters, public transportation, metro rail systems, Convention Halls, and Resorts etc... List is endless.

When people want to retire, they move out of DC area to have a relaxed retirement life. Many though, want to live close to DC with similar amenities and if desired, make a trip to D. C. area.

According to Kiplinger Personal Finance magazine, in its list of '12 Smart Places to Retire' lists Richmond, Virginia as one of the top in the list.

When people move, they want to make sure place has needed amenities and can move out quickly, if it does not suit them. House Price Index (HPI) measures the price changes of residential housing. HPI can be used to evaluate whether place is worth moving in without incurring a huge financial loss.

The project tries to cluster Richmond area neighborhoods with Northern Virginia neighborhoods, based on Venues. Then, show areas with HPI and Venue along with cluster information, for making a informed decision about where to move within Richmond.

A description of the data and how it will be used to solve the problem.

Following data will be used:

For ease of identifying neighborhood, each Zip Code is considered as a Neighborhood.

Thus, a file with Zip code and its longitude and latitude is needed for use with Foursquare API.

US state of Virginia specific zipcode data is downloaded from website:
https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/

The file is edited to have just Zip, City, State, Latitude and Longitude, values for places in Northern Virginia (NOVA) and Richmond, VA , in a CSV format and saved as NOVA_city_richmond_selected.csv.

House Price Index is available by Zip Code at web site:
https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx#qpo
Downloaded Five-Digit ZIP Codes (Developmental Index; Not Seasonally Adjusted) XLSX file.

The file has following columns:
Five-Digit ZIP Code, Year, Annual Change (%), HPI, HPI with 1990 base, HPI with 2000 base.

The downloaded file was filtered for data of Zip Code from 20001 thru 24000, year=2018 and created a CSV file HPI_AT_BDL_ZIP5_2018_VA.csv with following column : Five-Digit ZIP Code, Year, Annual Change (%),HPI.

NOVA_city_richmond_selected.csv fil merged with HPI_AT_BDL_ZIP5_2018_VA.csv file by Zip Code. This will make sure any combined data will have both geo-coordinates and HPI values.

The combined data will be stored in pandas dataframe as **nova_data**.

Using nova_data, call FourSquare API to get Venue details for each Zip Code (will be renamed as Neighborhood for consistency with Battle of Neighborhood theme. Once Venue data is obtained, do following steps:

1. Identify top 10 Venues for each Neighborhood.
2. Normalize the data,
3. Use K-Mean clustering to cluster the data
4. Show cluster on a folium map.
5. Create range for HPI
6. Try to show HPI data, cluster data on a Folium map
7. Try to show a Chloropleth map with Virginia as focus.

If Neighborhood in Richmond area is in same cluster as that of Northern Virginia, that means, it meets the amenities criteria, with the assumption that amenities are tied to Venues.
Once cluster association is done, person can check on HPI for given location from the map to narrow down the choice.

Will try to analyze the data at each stage of processing and will make observation on results.
At the end, draw some conclusion; identify any shortcomings of the approach, data inadequacy, and suggestion for further refinement.