Neighborhood_Battle

June 25, 2019

Battle of the Neighborhoods

0.1 Intorduction

- Client/Stake Holders: ACME Co. is an international precious stones and metals investment firm which has jewelry stores in various countries selling jewelries from low grade to the highest of unique gems.
- Overall Business Goal: Due to improved economic conditions, ACME Co. has decided to establish one or two jewelry stores providing not only highest quality and grade of jewelries but also provide services such as creating custom jewelry dependent on high net-worth clients' tastes and requests.
- Detailed Business Goal: Los Angeles county is culturally diverse, close to international shipping routs (Long Beach shipyards), center for some of the most famous fashion industries and large number of high net-worth individuals. Due to the facts above, ACME has decided to open high-end jewelry stores in Los Angeles county. Because of this they need to find the best LA city/cities to open one or two exclusive jewelry stores. This would also provide an opportunity for ACME Co. to introduce and build a brand for itself among the high net-worth communities.

0.2 Ideas / Solutions

- Ideas: So the questions that comes to mind are where are the best Los Angeles county areas to provide jewelry services to high net-worth individuals? What is the best way to find these individuals and have them visit our store(s)? How can we find high net-worth individuals?
- Solution: In my opinion the best answer is not to focus on such individuals because they could be spread out through out Los Angeles. In some areas more and in some areas with less density. Additionally, some could also leave in middle class areas. So instead of focusing on distinguished clientele, we will focus on properties which such individuals own and reside in, specifically their homes. In conclusion, we will be focusing on Los Angeles zip codes that have the highest >average< house values.
- Solution in Detail: So should we just find the first zip code with the highest average house price, and put up our jewelry store there? The answer is NO, because some expensive zones could be mostly residential, or the zones might not have the right type business environment or the zones/centers might not be the right type for jewelry store foot traffic. So we will grab the top 10 zip codes with the highest average house prices and cluster them into three groups (k-clusters) and then evaluate each group by the type of popular venues/businesses nearby.

0.3 Data Resources / Research

- Resources / Data: Our analysis will be based on several types of resources and data:
- 1 a) Description: A list of all Los Angeles county zip codes, names and geo positions
- 1 b) Resource: https://simplemaps.com/data/us-zips >
- 2 a) Description: A ist of all Los Angeles county zip codes and their average "single family" home values
- 2 b) Reesource: https://www.zillow.com/los-angeles-county-ca/home-values/ >
- 3 a) Description: Top five venues for each of the "top ten zipcodes" above
- 3 b) Resource: https://www.Foursquare.com

0.3.1 Import Libraries

```
In [2]: import pandas as pd

        import numpy as np

        import requests as requests

        # Matplotlib
        import matplotlib as mpl
        import matplotlib.pyplot as plt
        #associated plotting modules
        import matplotlib.cm as cm
        import matplotlib.colors as colors

        from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

        # import k-means from clustering stage
        from sklearn.cluster import KMeans

        #!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't completed the Fours
        import folium # map rendering library

        print('Libraries imported.')
```

```
Libraries imported.
```

```
In [3]: df_zip_info = pd.read_csv('uszips.csv')
```

Resource: https://simplemaps.com/data/us-zips The file containing whole U.S. Zipcode information such as Latitude, Longitude, city name is "uszips.csv" Since this file had to be manually downloaded, I cleaned it up ahead of time and copied it hear localy. Much easier to configure data in excel than Pandas!

File containing Zipcode information in Los Angesles county ...

```
In [4]: df_zip_info.head()
```

```
Out[4]:   Zipcode  Latitude  Longitude    City_Name         Zip_City
       0    90001   33.9740  -118.2495  Los Angeles  90001-Los Angeles
       1    90002   33.9491  -118.2467  Los Angeles  90002-Los Angeles
       2    90003   33.9641  -118.2737  Los Angeles  90003-Los Angeles
       3    90004   34.0762  -118.3108  Los Angeles  90004-Los Angeles
       4    90005   34.0591  -118.3064  Los Angeles  90005-Los Angeles
```

Resource: https://www.zillow.com/los-angeles-county-ca/home-values/. The file "Home_Prices_By_Zip.csv" contains all average home prices in Los Angeles county by zipcodes. Since I had to download it manually, I cleaned it up in excel before bringing it in here.

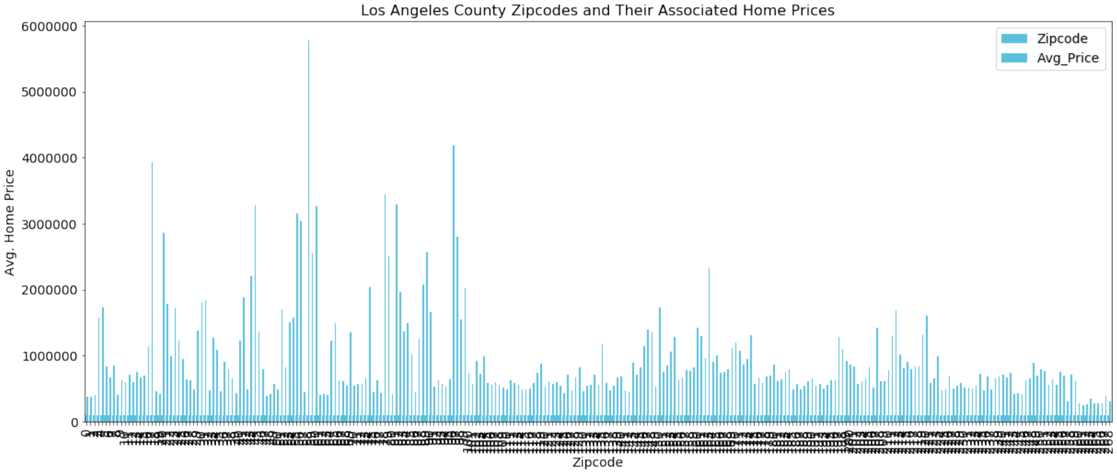
```
In [5]: df_zip_prices = pd.read_csv('Home_Prices_By_Zip.csv')
        df_zip_prices.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 269 entries, 0 to 268
Data columns (total 2 columns):
Zipcode      269 non-null int64
Avg_Price    269 non-null int64
dtypes: int64(2)
memory usage: 4.3 KB
```

```
In [6]: #df_zip_prices["Avg_Price"] = df_zip_prices["Avg_Price"].astype(dtype=np.int64)
        df_zip_prices.head()
```

```
Out[6]:   Zipcode  Avg_Price
       0    90001     378200
       1    90002     376500
       2    90003     397600
       3    90004    1576100
       4    90005    1732000
```

```
In [7]: bar_color = ('#5bc0de')
        bar_chart = df_zip_prices.plot(kind='bar', width=0.8, figsize=(20, 8), fontsize=14 , color=bar_color )
        bar_chart.set_title("Los Angeles County Zipcodes and Their Associated Home Prices", fontsize=16)
        bar_chart.set_xlabel("Zipcode", fontsize=14)
        bar_chart.set_ylabel("Avg. Home Price", fontsize=14)
        bar_chart.legend(fontsize=14)
        plt.show()
```



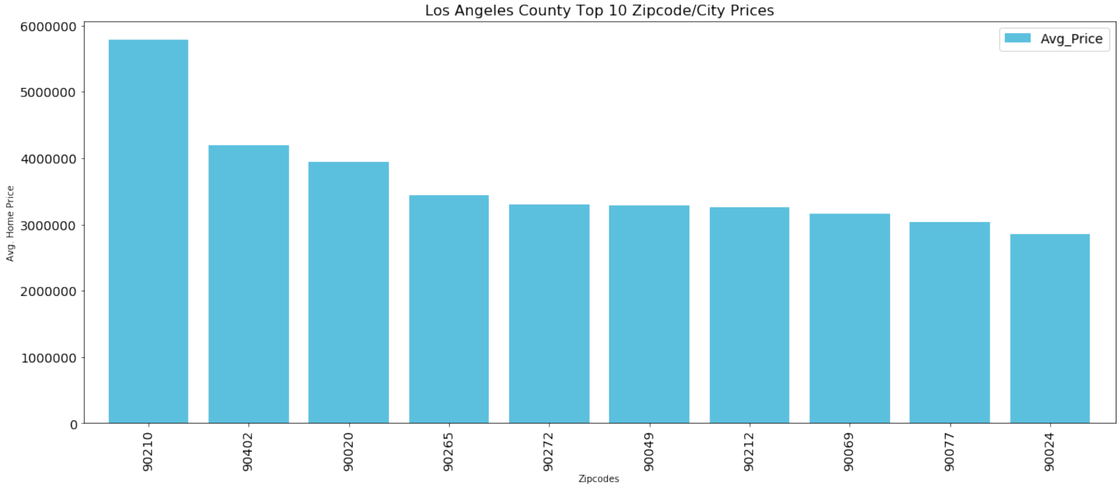
Lets take a look at all Los Angeles county zipcodes and their home price ranges ...

0.3.2 Next, lets grab the top ten zipcodes by home prices

In [8]:
```
df_zip_top10 = df_zip_prices.nlargest(10, 'Avg_Price')
df_zip_top10.sort_values(['Avg_Price'], ascending=False, axis=0, inplace=True)
df_zip_top10
```

Out[8]:
```
    Zipcode  Avg_Price
58    90210    5782800
96    90402    4190600
17    90020    3937900
78    90265    3442600
81    90272    3295600
44    90049    3286300
60    90212    3265100
55    90069    3156200
56    90077    3039300
20    90024    2859000
```

In [9]:
```
df_bar_top10 = df_zip_top10
#Lets remove the index column so it wouldn't show in the bar chart
df_bar_top10.set_index('Zipcode', inplace=True)
bar_color = ('#5bc0de')
bar_chart = df_bar_top10.plot(kind='bar', width=0.8, figsize=(20, 8), fontsize=14 , color=bar_color )
bar_chart.set_title("Los Angeles County Top 10 Zipcode/City Prices", fontsize=16)
bar_chart.set_xlabel("Zipcodes")
bar_chart.set_ylabel("Avg. Home Price")
bar_chart.legend(fontsize=14)
plt.show()
```



0.3.3 Now lets combine dataframe which contains zipcode information with the dataframe which contains top10 zipcode homeprices

```
In [10]: df_final_zip = pd.merge(df_zip_info, df_zip_top10, on='Zipcode')
         df_final_zip
```

```
Out[10]:   Zipcode  Latitude  Longitude          City_Name                Zip_City  \
        0    90020   34.0664  -118.3099        Los Angeles        90020-Los Angeles
        1    90024   34.0657  -118.4350        Los Angeles        90024-Los Angeles
        2    90049   34.0872  -118.4893        Los Angeles        90049-Los Angeles
        3    90069   34.0938  -118.3817     West Hollywood     90069-West Hollywood
        4    90077   34.1055  -118.4561        Los Angeles        90077-Los Angeles
        5    90210   34.1010  -118.4148      Beverly Hills      90210-Beverly Hills
        6    90212   34.0622  -118.4019      Beverly Hills      90212-Beverly Hills
        7    90265   34.0719  -118.8499             Malibu             90265-Malibu
        8    90272   34.0799  -118.5422  Pacific Palisades  90272-Pacific Palisades
        9    90402   34.0356  -118.5036       Santa Monica       90402-Santa Monica

          Avg_Price
        0   3937900
        1   2859000
        2   3286300
        3   3156200
        4   3039300
        5   5782800
        6   3265100
        7   3442600
        8   3295600
        9   4190600
```

0.3.4 Next lets put up a map to show how these top 10 zipcodes are situated against each other ...

0.3.5 But before we show the map, lets initialize it (center it) with a random city in our top 10 list ...

```
In [11]: #Lets get Latitude and Longitude of "90069, West Hollywood"
         Init_latitude = df_final_zip.loc[df_final_zip['Zipcode'] == 90069, 'Latitude'].iat[0]
         Init_longitude = df_final_zip.loc[df_final_zip['Zipcode'] == 90069, 'Longitude'].iat[0]
         print(Init_latitude)
         print(Init_longitude)
```

```
34.0938
-118.3817
```

```
In [12]: # create map of Los Angeles county using latitude and longitude values
         map_LosAngeles = folium.Map(location=[Init_latitude, Init_longitude], zoom_start=10)

         # add neighborhood/markers to map
         for lat, lng, Zip_City in zip(df_final_zip['Latitude'], df_final_zip['Longitude'], df_final_zip['Zip_City']):
             label = '{}'.format(Zip_City)
             label = folium.Popup(label, parse_html=True)
             folium.CircleMarker(
                 [lat, lng],
                 radius=5,
                 popup=label,
                 color='blue',
                 fill=True,
                 fill_color='#3186cc',
                 fill_opacity=0.7,
                 parse_html=False).add_to(map_LosAngeles)

         map_LosAngeles
```

```
Out[12]: <folium.folium.Map at 0x7f16c8d3d828>
```

If map not visible, click here to see map!

0.3.6 Next lets explore each zipcode/city top 5 venues using Foursquare API ...

0.3.7 Note: Judging by how zipcodes are situated against each other on the map, its best to focus on top 5 venues that are no further than 2 kilometers away!

```
In [13]: CLIENT_ID = 'CITQUKHYRNH24TDRB3E5FWK03UFKVSIEWLLX5R2DV1Q5G3JT' # your Foursq
         CLIENT_SECRET = 'TOEXJBCWJHPXLK5EFKMEBKX5FBOMOLZPGEHJ3CULSPYM03WE' # yo
         VERSION = '20180605' # Foursquare API version

         print('Your credentails:')
         print('CLIENT_ID: ' + CLIENT_ID)
         print('CLIENT_SECRET:' + CLIENT_SECRET)
```

Your credentails:
CLIENT_ID: CITQUKHYRNH24TDRB3E5FWK03UFKVSIEWLLX5R2DV1Q5G3JT
CLIENT_SECRET:TOEXJBCWJHPXLK5EFKMEBKX5FBOMOLZPGEHJ3CULSPYM03WE

In [15]:
```
#Lets repeat the same process for all neighborhoods in North York
def getNearbyVenues(names, latitudes, longitudes, radius=2000):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Zip_City',
                  'Zip_City Latitude',
                  'Zip_City Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

In [16]:
```
#Lets get the top 5 venues within the 2 kilometer radius
LIMIT = 5 # limit of number of venues returned by Foursquare API
radius = 2000 # define radius
```

```
In [17]: #Call the function for each of the top 10 Zipcodes in Los Angeles county in to a new dataframe
         LosAngeles_venues = getNearbyVenues(names=df_final_zip['Zip_City'],
                                             latitudes=df_final_zip['Latitude'],
                                             longitudes=df_final_zip['Longitude']
                                            )
```

```
90020-Los Angeles
90024-Los Angeles
90049-Los Angeles
90069-West Hollywood
90077-Los Angeles
90210-Beverly Hills
90212-Beverly Hills
90265-Malibu
90272-Pacific Palisades
90402-Santa Monica
```

```
In [18]: print(LosAngeles_venues.shape)
         LosAngeles_venues
```

```
(45, 7)
```

```
Out[18]:                Zip_City  Zip_City Latitude  Zip_City Longitude  \
         0     90020-Los Angeles            34.0664           -118.3099
         1     90020-Los Angeles            34.0664           -118.3099
         2     90020-Los Angeles            34.0664           -118.3099
         3     90020-Los Angeles            34.0664           -118.3099
         4     90020-Los Angeles            34.0664           -118.3099
         5     90024-Los Angeles            34.0657           -118.4350
         6     90024-Los Angeles            34.0657           -118.4350
         7     90024-Los Angeles            34.0657           -118.4350
         8     90024-Los Angeles            34.0657           -118.4350
         9     90024-Los Angeles            34.0657           -118.4350
         10    90049-Los Angeles            34.0872           -118.4893
         11    90049-Los Angeles            34.0872           -118.4893
         12    90049-Los Angeles            34.0872           -118.4893
         13    90049-Los Angeles            34.0872           -118.4893
         14    90049-Los Angeles            34.0872           -118.4893
         15 90069-West Hollywood            34.0938           -118.3817
         16 90069-West Hollywood            34.0938           -118.3817
         17 90069-West Hollywood            34.0938           -118.3817
         18 90069-West Hollywood            34.0938           -118.3817
         19 90069-West Hollywood            34.0938           -118.3817
         20    90077-Los Angeles            34.1055           -118.4561
         21    90077-Los Angeles            34.1055           -118.4561
         22    90077-Los Angeles            34.1055           -118.4561
         23    90077-Los Angeles            34.1055           -118.4561
```

```
24      90077-Los Angeles        34.1055       -118.4561
25     90210-Beverly Hills       34.1010       -118.4148
26     90210-Beverly Hills       34.1010       -118.4148
27     90210-Beverly Hills       34.1010       -118.4148
28     90210-Beverly Hills       34.1010       -118.4148
29     90210-Beverly Hills       34.1010       -118.4148
30     90212-Beverly Hills       34.0622       -118.4019
31     90212-Beverly Hills       34.0622       -118.4019
32     90212-Beverly Hills       34.0622       -118.4019
33     90212-Beverly Hills       34.0622       -118.4019
34     90212-Beverly Hills       34.0622       -118.4019
35  90272-Pacific Palisades      34.0799       -118.5422
36  90272-Pacific Palisades      34.0799       -118.5422
37  90272-Pacific Palisades      34.0799       -118.5422
38  90272-Pacific Palisades      34.0799       -118.5422
39  90272-Pacific Palisades      34.0799       -118.5422
40      90402-Santa Monica        34.0356       -118.5036
41      90402-Santa Monica        34.0356       -118.5036
42      90402-Santa Monica        34.0356       -118.5036
43      90402-Santa Monica        34.0356       -118.5036
44      90402-Santa Monica        34.0356       -118.5036

                                          Venue  Venue Latitude  \
0                       Han Bat Sul Lung Tang       34.065474
1                       Here's Looking at You       34.063635
2                                     Jun Won       34.066871
3                          Montys Good Burger       34.064797
4                                  Beer Belly       34.064302
5         UCLA Mildred E. Mathias Botanical Garden       34.064753
6              Wilshire Margot Luxury Apartments       34.061865
7                                iPic Theatres       34.059093
8                 Beverly Hills Plaza Hotel & Spa       34.067635
9                             Geffen Playhouse       34.063626
10                      Getty Sculpture Garden       34.087560
11                        J. Paul Getty Museum       34.077605
12                         Getty Center Garden       34.078157
13                 Getty Center North Building       34.078287
14                              Central Garden       34.076484
15                                  H Lorenzo       34.091727
16                          Sushiya on Sunset       34.092004
17                Tocaya Organica - Sunset Plaza       34.091639
18     The Butcher, The Baker, The Cappuccino Maker       34.092099
19                                   Eveleigh       34.091197
20                              Bel Air Foods       34.116383
21                               Beverly Glen       34.101815
22                            Bev Glen Market       34.100067
23                                Glen Market       34.099795
24                               The Clay Pit       34.116263
```

```
25                          Franklin Canyon Park        34.096533
26                         Coldwater Canyon Park        34.091264
27                       Greystone Mansion & Park        34.093137
28                     Goldstein House by Lautner        34.093408
29                              Dumpling Station        34.115127
30                                    Urth Caffé        34.062620
31                               XIV Karats Ltd.        34.061520
32                                 Cafe Istanbul        34.061220
33                       Ruth's Chris Steak House        34.064029
34  Beverly Wilshire Hotel (A Four Seasons Hotel) ...        34.066402
35                      Summit Club At Palisades        34.076921
36                              Top Of The World        34.064698
37                                    Skull Rock        34.070771
38                                    High Point        34.062871
39                                   Summit Club        34.076871
40                      North Santa Monica beach        34.034963
41                            Revolution Fitness        34.031074
42                                  Aero Theatre        34.031833
43                        YogaWorks Montana Ave        34.032706
44                      Rori's Artisanal Creamery        34.028677

    Venue Longitude                          Venue Category
0       -118.309471                       Korean Restaurant
1       -118.307999                 New American Restaurant
2       -118.308867                       Korean Restaurant
3       -118.309034                            Burger Joint
4       -118.308763                                     Bar
5       -118.440427                                  Garden
6       -118.433834  Residential Building (Apartment / Condo)
7       -118.441475                           Movie Theater
8       -118.426062                                   Hotel
9       -118.444523                         College Theater
10      -118.475748                              Art Museum
11      -118.474594                              Art Museum
12      -118.475079                                  Garden
13      -118.474994                                Building
14      -118.474741                                  Garden
15      -118.380404                                Boutique
16      -118.380219                         Sushi Restaurant
17      -118.381135                      Mexican Restaurant
18      -118.380536                                    Café
19      -118.381904                     American Restaurant
20      -118.464182                           Grocery Store
21      -118.445677                                    Road
22      -118.443871                           Grocery Store
23      -118.443645                           Grocery Store
24      -118.464130                       Indian Restaurant
25      -118.412046                                    Park
```

```
26   -118.411829                     Park
27   -118.401254      Other Great Outdoors
28   -118.434407             Historic Site
29   -118.422940                Food Truck
30   -118.399206                      Café
31   -118.398887             Jewelry Store
32   -118.399037  Mediterranean Restaurant
33   -118.398768                Steakhouse
34   -118.400691                     Hotel
35   -118.553413            Nightlife Spot
36   -118.549638            Scenic Lookout
37   -118.542775                     Trail
38   -118.537758                     Trail
39   -118.553121                       Gym
40   -118.505775                     Beach
41   -118.496779                       Gym
42   -118.495437       Indie Movie Theater
43   -118.494150               Yoga Studio
44   -118.499492            Ice Cream Shop
```

In [19]:
```
LosAngeles_venues.groupby('Zip_City').count()
```

Out[19]:
```
                         Zip_City Latitude  Zip_City Longitude  Venue  \
Zip_City
90020-Los Angeles                        5                   5      5
90024-Los Angeles                        5                   5      5
90049-Los Angeles                        5                   5      5
90069-West Hollywood                     5                   5      5
90077-Los Angeles                        5                   5      5
90210-Beverly Hills                      5                   5      5
90212-Beverly Hills                      5                   5      5
90272-Pacific Palisades                  5                   5      5
90402-Santa Monica                       5                   5      5

                       Venue Latitude  Venue Longitude  Venue Category
Zip_City
90020-Los Angeles                   5                5               5
90024-Los Angeles                   5                5               5
90049-Los Angeles                   5                5               5
90069-West Hollywood                5                5               5
90077-Los Angeles                   5                5               5
90210-Beverly Hills                 5                5               5
90212-Beverly Hills                 5                5               5
90272-Pacific Palisades             5                5               5
90402-Santa Monica                  5                5               5
```

In [20]:
```
print('There are {} uniques categories.'.format(len(LosAngeles_venues['Venue Category'].unique())))
```

```
There are 34 uniques categories.
```

0.4 Very Interesting !!!

0.4.1 If we pay special attention, we see here that one of the top zipcodes is missing. I.E. Foursquare did NOT have any venues for one of our zipcodes.

0.4.2 It seems that "90265-Malibu" does not have any venues. Most probably its either all residential area or there are no commercial zones near by.

0.4.3 We will keep this in mind because when time comes for clustering cities, we will need to remove "90265-Malibu" from our analysis.

```
In [21]: # one hot encoding
         LosAngeles_onehot = pd.get_dummies(LosAngeles_venues[['Venue Category']], prefix="", prefix_sep=""

         # add Zip_City column back to dataframe
         LosAngeles_onehot['Zip_City'] = LosAngeles_venues['Zip_City']

         # move Zip_City column to the first column
         fixed_columns = [LosAngeles_onehot.columns[-1]] + list(LosAngeles_onehot.columns[:-1])
         LosAngeles_onehot = LosAngeles_onehot[fixed_columns]

         LosAngeles_onehot.head()
```

```
Out[21]:          Zip_City  American Restaurant  Art Museum  Bar  Beach  Boutique  \
         0  90020-Los Angeles                    0           0    0      0         0
         1  90020-Los Angeles                    0           0    0      0         0
         2  90020-Los Angeles                    0           0    0      0         0
         3  90020-Los Angeles                    0           0    0      0         0
         4  90020-Los Angeles                    0           0    1      0         0

            Building  Burger Joint  Café  College Theater  ...  Nightlife Spot  \
         0         0             0     0                0  ...               0
         1         0             0     0                0  ...               0
         2         0             0     0                0  ...               0
         3         0             1     0                0  ...               0
         4         0             0     0                0  ...               0

            Other Great Outdoors  Park  Residential Building (Apartment / Condo)  Road  \
         0                     0     0                                         0     0
         1                     0     0                                         0     0
         2                     0     0                                         0     0
         3                     0     0                                         0     0
         4                     0     0                                         0     0

            Scenic Lookout  Steakhouse  Sushi Restaurant  Trail  Yoga Studio
         0               0           0                 0      0            0
         1               0           0                 0      0            0
         2               0           0                 0      0            0
         3               0           0                 0      0            0
         4               0           0                 0      0            0
```

```
[5 rows x 35 columns]
```

0.4.4 Convert Categorical values into dummy/indicator values

```
In [22]: LosAngeles_onehot.shape
```

```
Out[22]: (45, 35)
```

```
In [23]: #Next, let's group rows by Zip_City and by taking the mean of the frequency of occurrence of each categ
         LosAngeles_grouped = LosAngeles_onehot.groupby('Zip_City').mean().reset_index()
         LosAngeles_grouped
```

```
Out[23]:                  Zip_City  American Restaurant  Art Museum  Bar  Beach  \
         0        90020-Los Angeles                  0.0         0.0  0.2    0.0
         1        90024-Los Angeles                  0.0         0.0  0.0    0.0
         2        90049-Los Angeles                  0.0         0.4  0.0    0.0
         3     90069-West Hollywood                  0.2         0.0  0.0    0.0
         4        90077-Los Angeles                  0.0         0.0  0.0    0.0
         5      90210-Beverly Hills                  0.0         0.0  0.0    0.0
         6      90212-Beverly Hills                  0.0         0.0  0.0    0.0
         7  90272-Pacific Palisades                  0.0         0.0  0.0    0.0
         8       90402-Santa Monica                  0.0         0.0  0.0    0.2

            Boutique  Building  Burger Joint  Café  College Theater  ...  \
         0       0.0       0.0           0.2   0.0              0.0  ...
         1       0.0       0.0           0.0   0.0              0.2  ...
         2       0.0       0.2           0.0   0.0              0.0  ...
         3       0.2       0.0           0.0   0.2              0.0  ...
         4       0.0       0.0           0.0   0.0              0.0  ...
         5       0.0       0.0           0.0   0.0              0.0  ...
         6       0.0       0.0           0.0   0.2              0.0  ...
         7       0.0       0.0           0.0   0.0              0.0  ...
         8       0.0       0.0           0.0   0.0              0.0  ...

            Nightlife Spot  Other Great Outdoors  Park  \
         0             0.0                   0.0   0.0
         1             0.0                   0.0   0.0
         2             0.0                   0.0   0.0
         3             0.0                   0.0   0.0
         4             0.0                   0.0   0.0
         5             0.0                   0.2   0.4
         6             0.0                   0.0   0.0
         7             0.2                   0.0   0.0
         8             0.0                   0.0   0.0

            Residential Building (Apartment / Condo)  Road  Scenic Lookout  Steakhouse  \
         0                                       0.0   0.0             0.0         0.0
         1                                       0.2   0.0             0.0         0.0
```

```
2                                  0.0   0.0           0.0         0.0
3                                  0.0   0.0           0.0         0.0
4                                  0.0   0.2           0.0         0.0
5                                  0.0   0.0           0.0         0.0
6                                  0.0   0.0           0.0         0.2
7                                  0.0   0.0           0.2         0.0
8                                  0.0   0.0           0.0         0.0

   Sushi Restaurant  Trail  Yoga Studio
0               0.0    0.0          0.0
1               0.0    0.0          0.0
2               0.0    0.0          0.0
3               0.2    0.0          0.0
4               0.0    0.0          0.0
5               0.0    0.0          0.0
6               0.0    0.0          0.0
7               0.0    0.4          0.0
8               0.0    0.0          0.2

[9 rows x 35 columns]
```

In [26]:
```python
#Let's print each Zip_City along with the top 5 most common venues
num_top_venues = 5

for hood in LosAngeles_grouped['Zip_City']:
    print("----"+hood+"----")
    temp = LosAngeles_grouped[LosAngeles_grouped['Zip_City'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
```

```
----90020-Los Angeles----
                    venue  freq
0       Korean Restaurant   0.4
1                     Bar   0.2
2            Burger Joint   0.2
3  New American Restaurant   0.2
4                    Park   0.0
----90024-Los Angeles----
                                  venue  freq
0                                Garden   0.2
1                         Movie Theater   0.2
2                                 Hotel   0.2
3  Residential Building (Apartment / Condo)   0.2
4                       College Theater   0.2
----90049-Los Angeles----
```

```
                      venue  freq
0                Art Museum   0.4
1                    Garden   0.4
2                  Building   0.2
3  Mediterranean Restaurant   0.0
4        Mexican Restaurant   0.0
----90069-West Hollywood----
                 venue  freq
0  American Restaurant   0.2
1     Sushi Restaurant   0.2
2             Boutique   0.2
3   Mexican Restaurant   0.2
4                 Café   0.2
----90077-Los Angeles----
                 venue  freq
0        Grocery Store   0.6
1    Indian Restaurant   0.2
2                 Road   0.2
3  American Restaurant   0.0
4       Nightlife Spot   0.0
----90210-Beverly Hills----
                      venue  freq
0                      Park   0.4
1                Food Truck   0.2
2      Other Great Outdoors   0.2
3             Historic Site   0.2
4  Mediterranean Restaurant   0.0
----90212-Beverly Hills----
                      venue  freq
0  Mediterranean Restaurant   0.2
1                Steakhouse   0.2
2                     Hotel   0.2
3                      Café   0.2
4             Jewelry Store   0.2
----90272-Pacific Palisades----
                 venue  freq
0                Trail   0.4
1       Nightlife Spot   0.2
2       Scenic Lookout   0.2
3                  Gym   0.2
4  American Restaurant   0.0
----90402-Santa Monica----
                 venue  freq
0  Indie Movie Theater   0.2
1                  Gym   0.2
2       Ice Cream Shop   0.2
3          Yoga Studio   0.2
4                Beach   0.2
```

```
In [28]: #Now let's create the new dataframe and display the top 5 venues for each Zip_City.
         num_top_venues = 5

         indicators = ['st', 'nd', 'rd']

         # create columns according to number of top venues
         columns = ['Zip_City']
         for ind in np.arange(num_top_venues):
             try:
                 columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
             except:
                 columns.append('{}th Most Common Venue'.format(ind+1))

         # create a new dataframe
         Zip_City_venues_sorted = pd.DataFrame(columns=columns)
         Zip_City_venues_sorted['Zip_City'] = LosAngeles_grouped['Zip_City']

         for ind in np.arange(LosAngeles_grouped.shape[0]):
             Zip_City_venues_sorted.iloc[ind, 1:] = return_most_common_venues(LosAngeles_grouped.iloc[ind, :

         Zip_City_venues_sorted
```

```
Out[28]:                 Zip_City 1st Most Common Venue    2nd Most Common Venue  \
         0        90020-Los Angeles     Korean Restaurant  New American Restaurant
         1        90024-Los Angeles                 Hotel          College Theater
         2        90049-Los Angeles            Art Museum                   Garden
         3     90069-West Hollywood   American Restaurant                 Boutique
         4        90077-Los Angeles         Grocery Store        Indian Restaurant
         5      90210-Beverly Hills                  Park            Historic Site
         6      90212-Beverly Hills                  Café               Steakhouse
         7  90272-Pacific Palisades                 Trail           Scenic Lookout
         8       90402-Santa Monica           Yoga Studio                    Beach

           3rd Most Common Venue                       4th Most Common Venue  \
         0                   Bar                                 Burger Joint
         1                Garden  Residential Building (Apartment / Condo)
         2              Building                                  Yoga Studio
         3    Mexican Restaurant                                         Café
         4                  Road                                         Café
         5            Food Truck                         Other Great Outdoors
         6         Jewelry Store                     Mediterranean Restaurant
         7                   Gym                                Nightlife Spot
         8                   Gym                                Ice Cream Shop

           5th Most Common Venue
         0           Yoga Studio
```

```
1          Movie Theater
2             Food Truck
3      Sushi Restaurant
4          Historic Site
5            Yoga Studio
6                  Hotel
7                    Bar
8    Indie Movie Theater
```

```
In [27]: #Lets sort the venues in descending order
         def return_most_common_venues(row, num_top_venues):
             row_categories = row.iloc[1:]
             row_categories_sorted = row_categories.sort_values(ascending=False)

             return row_categories_sorted.index.values[0:num_top_venues]
```

0.4.5 Now let's create the new dataframe and display the top 5 venues for each Zip_City.

```
In [29]: #Run k-means to cluster the neighborhood into 3 clusters.
         # set number of clusters
         kclusters = 3

         LosAngeles_grouped_clustering = LosAngeles_grouped.drop('Zip_City', 1)

         # run k-means clustering
         kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(LosAngeles_grouped_clustering)

         # check cluster labels generated for each row in the dataframe
         kmeans.labels_[0:10]
```

```
Out[29]: array([0, 0, 0, 0, 2, 0, 0, 1, 1], dtype=int32)
```

```
In [30]: #Let's create a new dataframe that includes the cluster as well as the top 5 venues for each Zip_City
         # add clustering labels
         Zip_City_venues_sorted.insert(0, 'Cluster_Labels', kmeans.labels_)
         #Zip_City_venues_sorted.head()
         LosAngeles_merged = df_final_zip

         # merge LosAngeles_grouped with LosAngeles_data to add latitude/longitude for each Zip_City
         LosAngeles_merged = LosAngeles_merged.join(Zip_City_venues_sorted.set_index('Zip_City'), on='Zip

         #LosAngeles_merged['Cluster Labels'] = LosAngeles_merged['Cluster Labels'].astype(int)

         LosAngeles_merged.head(20) # check the last columns!
```

```
Out[30]:    Zipcode  Latitude  Longitude       City_Name            Zip_City  \
         0    90020   34.0664  -118.3099     Los Angeles   90020-Los Angeles
         1    90024   34.0657  -118.4350     Los Angeles   90024-Los Angeles
         2    90049   34.0872  -118.4893     Los Angeles   90049-Los Angeles
```

```
3   90069  34.0938  -118.3817     West Hollywood    90069-West Hollywood
4   90077  34.1055  -118.4561        Los Angeles       90077-Los Angeles
5   90210  34.1010  -118.4148      Beverly Hills     90210-Beverly Hills
6   90212  34.0622  -118.4019      Beverly Hills     90212-Beverly Hills
7   90265  34.0719  -118.8499             Malibu            90265-Malibu
8   90272  34.0799  -118.5422  Pacific Palisades  90272-Pacific Palisades
9   90402  34.0356  -118.5036       Santa Monica      90402-Santa Monica

   Avg_Price  Cluster_Labels 1st Most Common Venue    2nd Most Common Venue  \
0    3937900             0.0     Korean Restaurant  New American Restaurant
1    2859000             0.0                 Hotel          College Theater
2    3286300             0.0            Art Museum                   Garden
3    3156200             0.0   American Restaurant                 Boutique
4    3039300             2.0         Grocery Store        Indian Restaurant
5    5782800             0.0                  Park            Historic Site
6    3265100             0.0                  Café               Steakhouse
7    3442600             NaN                   NaN                      NaN
8    3295600             1.0                 Trail          Scenic Lookout
9    4190600             1.0           Yoga Studio                    Beach

  3rd Most Common Venue                      4th Most Common Venue  \
0                   Bar                               Burger Joint
1                Garden  Residential Building (Apartment / Condo)
2              Building                                Yoga Studio
3    Mexican Restaurant                                       Café
4                  Road                                       Café
5            Food Truck                       Other Great Outdoors
6         Jewelry Store                    Mediterranean Restaurant
7                   NaN                                        NaN
8                   Gym                              Nightlife Spot
9                   Gym                             Ice Cream Shop

  5th Most Common Venue
0           Yoga Studio
1         Movie Theater
2            Food Truck
3       Sushi Restaurant
4         Historic Site
5           Yoga Studio
6                 Hotel
7                   NaN
8                   Bar
9   Indie Movie Theater
```

0.4.6 Lets show the cluster assignments and top 5 venues for each zip/city ...

0.4.7 As we remember earlier on, "90265-Malibu" did not have any venues, so lets delete the row so it will be excluded from mapping

```
In [31]: LosAngeles_merged = LosAngeles_merged[LosAngeles_merged.Zipcode != 90265]
         LosAngeles_merged.sort_values(['Cluster_Labels'], ascending=False, axis=0, inplace=True)
         LosAngeles_merged = LosAngeles_merged.reset_index()
         LosAngeles_merged.head(20)
```

```
Out[31]:   index  Zipcode  Latitude  Longitude          City_Name  \
         0      4    90077   34.1055  -118.4561        Los Angeles
         1      8    90272   34.0799  -118.5422  Pacific Palisades
         2      9    90402   34.0356  -118.5036       Santa Monica
         3      0    90020   34.0664  -118.3099        Los Angeles
         4      1    90024   34.0657  -118.4350        Los Angeles
         5      2    90049   34.0872  -118.4893        Los Angeles
         6      3    90069   34.0938  -118.3817     West Hollywood
         7      5    90210   34.1010  -118.4148      Beverly Hills
         8      6    90212   34.0622  -118.4019      Beverly Hills

                            Zip_City  Avg_Price  Cluster_Labels 1st Most Common Venue  \
         0         90077-Los Angeles    3039300             2.0         Grocery Store
         1   90272-Pacific Palisades    3295600             1.0                 Trail
         2        90402-Santa Monica    4190600             1.0           Yoga Studio
         3         90020-Los Angeles    3937900             0.0     Korean Restaurant
         4         90024-Los Angeles    2859000             0.0                 Hotel
         5         90049-Los Angeles    3286300             0.0            Art Museum
         6      90069-West Hollywood    3156200             0.0   American Restaurant
         7       90210-Beverly Hills    5782800             0.0                  Park
         8       90212-Beverly Hills    3265100             0.0                  Café

           2nd Most Common Venue 3rd Most Common Venue  \
         0     Indian Restaurant                  Road
         1        Scenic Lookout                   Gym
         2                 Beach                   Gym
         3 New American Restaurant                 Bar
         4       College Theater                Garden
         5                Garden              Building
         6              Boutique    Mexican Restaurant
         7         Historic Site            Food Truck
         8            Steakhouse         Jewelry Store

                         4th Most Common Venue 5th Most Common Venue
         0                                Café         Historic Site
         1                       Nightlife Spot                   Bar
         2                       Ice Cream Shop    Indie Movie Theater
         3                         Burger Joint           Yoga Studio
         4 Residential Building (Apartment / Condo)      Movie Theater
```

```
5                   Yoga Studio           Food Truck
6                          Café    Sushi Restaurant
7          Other Great Outdoors          Yoga Studio
8      Mediterranean Restaurant                Hotel
```

0.4.8 The map function does not like cluster values as floats, so lets convert the Cluster_Label values to integer

```
In [32]: LosAngeles_merged['Cluster_Labels'] = LosAngeles_merged['Cluster_Labels'].astype(int)
         LosAngeles_merged.head(20)
```

```
Out[32]:    index  Zipcode  Latitude  Longitude          City_Name  \
         0      4    90077   34.1055  -118.4561        Los Angeles
         1      8    90272   34.0799  -118.5422  Pacific Palisades
         2      9    90402   34.0356  -118.5036       Santa Monica
         3      0    90020   34.0664  -118.3099        Los Angeles
         4      1    90024   34.0657  -118.4350        Los Angeles
         5      2    90049   34.0872  -118.4893        Los Angeles
         6      3    90069   34.0938  -118.3817     West Hollywood
         7      5    90210   34.1010  -118.4148      Beverly Hills
         8      6    90212   34.0622  -118.4019      Beverly Hills

                      Zip_City  Avg_Price  Cluster_Labels 1st Most Common Venue  \
         0       90077-Los Angeles    3039300               2        Grocery Store
         1 90272-Pacific Palisades    3295600               1                Trail
         2      90402-Santa Monica    4190600               1          Yoga Studio
         3       90020-Los Angeles    3937900               0    Korean Restaurant
         4       90024-Los Angeles    2859000               0                Hotel
         5       90049-Los Angeles    3286300               0           Art Museum
         6    90069-West Hollywood    3156200               0  American Restaurant
         7     90210-Beverly Hills    5782800               0                 Park
         8     90212-Beverly Hills    3265100               0                 Café

              2nd Most Common Venue 3rd Most Common Venue  \
         0        Indian Restaurant                  Road
         1           Scenic Lookout                   Gym
         2                    Beach                   Gym
         3  New American Restaurant                   Bar
         4          College Theater                Garden
         5                   Garden              Building
         6                 Boutique    Mexican Restaurant
         7            Historic Site            Food Truck
         8               Steakhouse         Jewelry Store

                      4th Most Common Venue 5th Most Common Venue
         0                            Café         Historic Site
         1                   Nightlife Spot                   Bar
         2                   Ice Cream Shop    Indie Movie Theater
```

```
3                          Burger Joint          Yoga Studio
4  Residential Building (Apartment / Condo)         Movie Theater
5                           Yoga Studio            Food Truck
6                                 Café       Sushi Restaurant
7                    Other Great Outdoors          Yoga Studio
8                 Mediterranean Restaurant                Hotel
```

In [33]:
```
#Lets visualize the resulting clusters
# create map
map_clusters = folium.Map(location=[Init_latitude, Init_longitude], zoom_start=11)

# set color scheme for the clusters
#x = np.arange(kclusters)
#ys = [i + x + (i*x)**2 for i in range(kclusters)]
#colors_array = cm.rainbow(np.linspace(0.0, 1.0, len(ys)))
#rainbow = [colors.rgb2hex(i) for i in colors_array]

# In this case I prefer to use my onw custom colors instead of using randon color "rainbow" generator

rainbow = ['#8000ff', '#ffaa00', '#ff0000']

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(LosAngeles_merged['Latitude'], LosAngeles_merged['Longitude'], LosAngele
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster],
        fill=True,
        fill_color=rainbow[cluster],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[33]:
```
<folium.folium.Map at 0x7f16c9530400>
```

If map not visible, click here to see map!

0.5 Observations:

0.6 Cluster/Group Dynamics:

First Cluster(#0): We can see that Cluster(#0) has various venues which are very different in different zipcodes. It seems that the best two candidates for this group are "90212-Beverly Hills" and "90024-Los Angeles". Other cities' venues in this group have more of a casual type of commerce which would not be compatible with our high-end jewelry store.

Second Cluster(#1): Here in this group we can see that most popular venues are either casual type commerce or free outdoor activities. Therefore, none of these cities seem to offer us proper commerce centers for our business.

Third Cluster(#2): In this last group we have only one zipcode and the venues seem to be not only casual commerce but also center for tourists which are not the type of clientele we're going after.

City Map Placement: By looking at the map above, we can see that the cities in each group are somewhat closely situated to other cities in the same group.

0.7 Final Decision

So to review our original goal, in summary, we would like to open one or two high-end jewelry stores in proper zipcode(s). The second cluster(#1) and the third cluster(#2) do not seem to have compatible top venues for foot traffic of high-end clientele we are looking for. However, first cluster(#0) seems to provide couple of opportunities for our store.

0.7.1 Decision in more Detail

> It seems that "90212-Beverly Hills" seems to offer the best type of venues because not only commerce centers are the compatible with our type of high-end clientele, additionally there are other jewelry stores nearby. Other jewelry stores here are not only not competition, but if we put our business there, as a group we could attract more foot traffic for shoppers interested in high-end jewelry. Moreover, since the zipcodes in this group are physically very close to each other, it would not make any sense to open a second store in other cities in this group. So finally, establishing one jewelry store in "90212-Beverly Hills" should satisfy all of our requirements.