BEST NEIGHBOURHOOD IN KARACHI FOR PURCHASING NEW HOMES

JUNE 14, 2021

MUHAMMAD AMIN GHIAS

Table of Contents

ABSTRACT:

Buying new house is in Karachi is difficult task, this project will help in finding a neighborhood which will be better for buying house and moving in considering the desired requirements of people. We will utilize and manipulate the data to find the best locations based on venues, area size and number of bedrooms.

INTRODUCTION

People in this fast paced and competitive world have to move and change homes. Changing homes within a big city takes a lot of effort and time and becomes a huge task which has severe consequences on their life. This project will compare different neighborhoods and will help people in their decision of which neighborhood is better.

BUSINESS PROBLEM

In a big and expanding city with a population of approximately 2 crore new residencies are situated far from the center of the city. Shifting home/neighborhood in a large city is quite a challenging task and requires a wise decision. Many factors are to be taken in consideration when one shifts home.

The problem people shifting home are the availability of following in the new area:

- Restaurants
- Shopping centers
- Grocery Stores
- Parks
- Cinemas
- Transport buses

Everyone has their own favorite restaurant, coffee shop, the biggest problem comes in shopping centers as big shopping complexes are few in city, all of this results in people traveling larger distances when facilities are not available in their new area.

This assignment will help the people in making this decision and provide a comparison which locality is better in regards to facilities. We will cluster different localities and identify the facilities within their radius and formulate the results. With the help of this project people can chose houses based on which neighborhood has better facilities.

DATA ACQUISITION AND WRANGLING

DATA SOURCE

For this project we will use online data the source is from kaggle originally from zameen.com

CSV file with url:" https://opendata.com.pk/dataset/9e959916-1cfc-4e28-85c8-f10ff63e5df2/resource/8c882a14-ee3f-4534-a15f-5fec2319d5e5/download/zameen-updated.csv "

The data set lists all the advertisements for houses, flats, farm houses etc in the city of Karachi along with their coordinates longitude, latitude, price size etc from different localities of the city

DATA CLEANING

The CSV dataset contains many homes from a locality, with different area size and number of rooms and prices, as shown below:

	property_id	location_id	page_url	property_type	price	location	city	province_name	latitude	longitude
0	237062	3325	https://www.zameen.com/Property/g_10_g_10_2_gr...	Flat	10000000	G-10	Islamabad	Islamabad Capital	33.679890	73.012640
1	346905	3236	https://www.zameen.com/Property/e_11_2_service...	Flat	6900000	E-11	Islamabad	Islamabad Capital	33.700993	72.971492
2	386513	764	https://www.zameen.com/Property/islamabad_g_15...	House	16500000	G-15	Islamabad	Islamabad Capital	33.631486	72.926559
3	656161	340	https://www.zameen.com/Property/islamabad_bani...	House	43500000	Bani Gala	Islamabad	Islamabad Capital	33.707573	73.151199
4	841645	3226	https://www.zameen.com/Property/dha_valley_dha...	House	7000000	DHA Defence	Islamabad	Islamabad Capital	33.492591	73.301339

The dataset is modified so it contains data for Karachi:

	property_id	location_id	page_url	property_type	price	location	city	province_name	latitude	longitu
156	86575	6649	https://www.zameen.com/Property/faisal_cantonm...	House	450000000	Cantt	Karachi	Sindh	24.889395	67.0986
157	342005	232	https://www.zameen.com/Property/karachi_gulist...	House	35000000	Gulistan-e-Jauhar	Karachi	Sindh	24.914988	67.1387
158	466607	1484	https://www.zameen.com/Property/d_h_a_dha_phas...	Flat	21000000	DHA Defence	Karachi	Sindh	24.814367	67.0720
159	678919	9594	https://www.zameen.com/Property/malir_malir_ka...	House	6500000	Malir	Karachi	Sindh	24.882302	67.1846
160	813506	6732	https://www.zameen.com/Property/surjani_town_s...	House	13000000	Gadap Town	Karachi	Sindh	25.018156	67.0668

A new dataframe is made by copying the important parameters required for our analysis which is:

	property_type	price	location	city	latitude	longitude	Area Size	bedrooms
156	House	450000000	Cantt	Karachi	24.889395	67.098627	4.0	6
157	House	35000000	Gulistan-e-Jauhar	Karachi	24.914988	67.138702	16.0	6
158	Flat	21000000	DHA Defence	Karachi	24.814367	67.072083	8.9	3
159	House	6500000	Malir	Karachi	24.882302	67.184677	3.2	2
160	House	13000000	Gadap Town	Karachi	25.018156	67.066864	9.6	4

In order to make data better and find the mean coordinates of a location, the data was grouped on basis of 'location' and aggregation of mean latitudes and longitude area, mean min max, prices and rooms , house value count was added it looked like below:

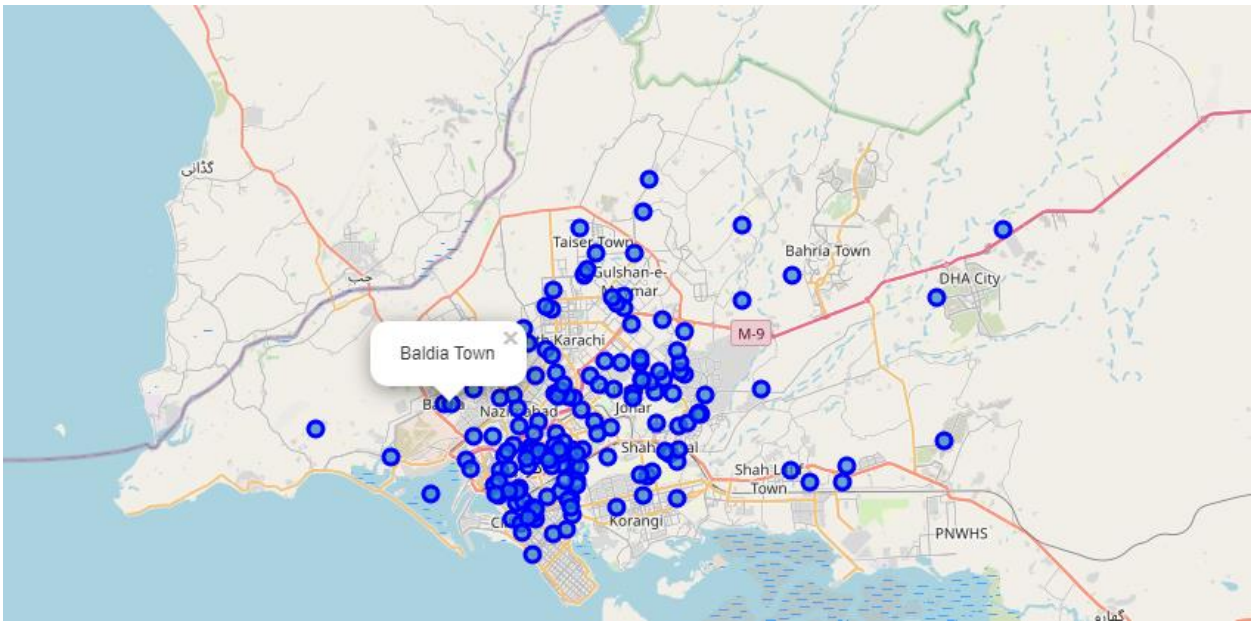
	property_type	location	city	mean_latitude	mean_longitude	price_mean	price_min	price_max	area_mean	are_min	area_max	bedroom_mean	b
0	Farm House	Bahria Town Karachi	Karachi	24.991752	67.255541	6.273810e+07	25000000.0	140000000.0	9.409524	6.0	19.4	0.523810	
1	Farm House	Bin Qasim Town	Karachi	24.865622	67.284917	3.003625e+07	20000.0	120000000.0	17.000000	12.0	20.0	3.500000	
2	Farm House	DHA City Karachi	Karachi	25.015873	67.418095	3.457357e+07	15000.0	70000000.0	4.142857	2.0	7.0	2.714286	
3	Farm House	DHA Defence	Karachi	24.819653	67.077749	2.500000e+04	25000.0	25000.0	7.000000	7.0	7.0	2.000000	
4	Farm House	Damba	Karachi	24.972520	67.259372	3.650000e+04	28000.0	45000.0	8.000000	8.0	8.0	0.000000	

Further dataframe is modified so it contains data only for houses:

	index	property_type	location	city	mean_latitude	mean_longitude	price_mean	price_min	price_max	area_mean	are_min	area_max	bedroom_
0	185	House	APP Employees Co-operative Housing Society	Karachi	24.928572	67.157725	3.900000e+07	39000000.0	39000000.0	16.00	16.0	16.0	8.0
1	186	House	ASF Housing Scheme	Karachi	24.937917	67.153713	7.000000e+06	7000000.0	7000000.0	4.80	4.8	4.8	4.0
2	187	House	Abdullah Ahmed Road	Karachi	24.891224	66.988458	6.645000e+07	2900000.0	130000000.0	3.00	1.0	5.0	2.5
3	188	House	Abdullah Haroon Road	Karachi	24.847829	67.031185	1.088000e+07	400000.0	30000000.0	8.36	1.0	16.0	4.2
4	189	House	Abid Town	Karachi	24.924261	67.082496	1.493257e+07	28000.0	28000000.0	4.80	1.8	9.6	4.7

DATA UTILIZATION:

Using the longitude and latitude we plot a map of locations in Karachi as shown:



Using the data of mean latitude and longitude from the above data frame we find the venues in the radius using **Foursquare API** the results are stored as JSON file then the data is formulated into a dataframe, the dataset contains venue names, categories, latitude and longitude the table is shown below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	APP Employees Co-operative Housing Society	24.928572	67.157725	Saien's Chai	24.927310	67.160069	Coffee Shop
1	ASF Housing Scheme	24.937917	67.153713	Bin Hashim	24.938233	67.151470	Department Store
2	ASF Housing Scheme	24.937917	67.153713	Muslim Pharmacy & Super Market	24.938128	67.151982	Shopping Mall
3	ASF Housing Scheme	24.937917	67.153713	Food Time	24.940111	67.151352	Fast Food Restaurant
4	ASF Housing Scheme	24.937917	67.153713	Sindh Green Restaurant	24.939285	67.157438	BBQ Joint
5	Abdullah Ahmed Road	24.891224	66.988458	Shell	24.893470	66.989346	Gas Station
6	Abdullah Haroon Road	24.847829	67.031185	Frere Hall	24.846787	67.032243	Historic Site
7	Abdullah Haroon Road	24.847829	67.031185	Karachi Club	24.844083	67.029199	Social Club
8	Abdullah Haroon Road	24.847829	67.031185	Sind Club Bistro	24.849628	67.031978	Bistro

The dataframe is grouped on basis of location and aggregated on venue categories value count, this forms table below:

		Venue Category value_counts
Neighborhood	Venue Category	
APP Employees Co-operative Housing Society	Coffee Shop	1
ASF Housing Scheme	BBQ Joint	1
	Department Store	1
	Fast Food Restaurant	1
	Shopping Mall	1
Abdullah Ahmed Road	Gas Station	1
Abdullah Haroon Road	Café	2
	Social Club	2
	Asian Restaurant	1
	Bistro	1
	Fast Food Restaurant	1
	Historic Site	1
	Hotel	1
	Restaurant	1

The above data set shows venues for each location

The data set is unstacked to find the venues in area, the column of total number of venues is added in dataset

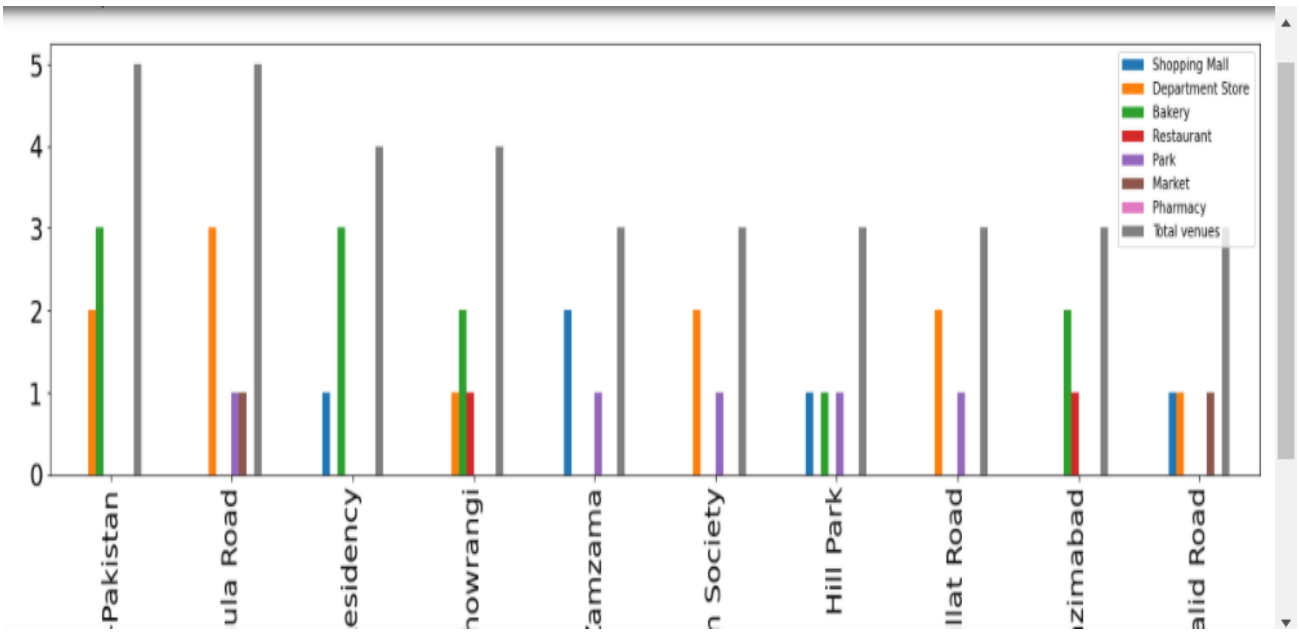
	Venue Category value_counts													
Venue Category	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Asian Restaurant	Athletics & Sports	Auto Workshop	BBQ Joint	Bakery	Bar	Beach	Bistro	Bookstore	Boutique	B A
Neighborhood															
Shahrah-e-Pakistan	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN	NaN	
Siraj-ud-Daula Road	NaN	NaN	NaN	1.0	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	
Suparco Road	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	
Super Highway	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Surti Muslim Co-Operative Housing Society	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Tariq Road	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	
Teacher Society	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Tipu Sultan Road	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN	

The venues categories which are useful for analysis is formed into a new dataframe. Further we sort locations which have highest number of venues as show below:

	location	Shopping Mall	Department Store	Bakery	Restaurant	Park	Market	Pharmacy	Total venues
0	Shahrah-e-Pakistan	NaN	2.0	3.0	NaN	NaN	NaN	NaN	5.0
1	Siraj-ud-Daula Road	NaN	3.0	NaN	NaN	1.0	1.0	NaN	5.0
2	Blue Bell Residency	1.0	NaN	3.0	NaN	NaN	NaN	NaN	4.0
3	Maskan Chowrangi	NaN	1.0	2.0	1.0	NaN	NaN	NaN	4.0
4	Zamzama	2.0	NaN	NaN	NaN	1.0	NaN	NaN	3.0
5	Others	1.0	NaN	NaN	1.0	NaN	1.0	NaN	3.0
6	Darul Aman Society	NaN	2.0	NaN	NaN	1.0	NaN	NaN	3.0
7	Hill Park	1.0	NaN	1.0	NaN	1.0	NaN	NaN	3.0
8	Shaheed Millat Road	NaN	2.0	NaN	NaN	1.0	NaN	NaN	3.0
9	North Nazimabad	NaN	NaN	2.0	1.0	NaN	NaN	NaN	3.0

EXPLORATORY DATA ANALYSIS:

A bar plot of top 10 locations with most venues is plotted:



Lastly the two datasets are joined to compare locality venue counts with different parameters of the older dataframe.

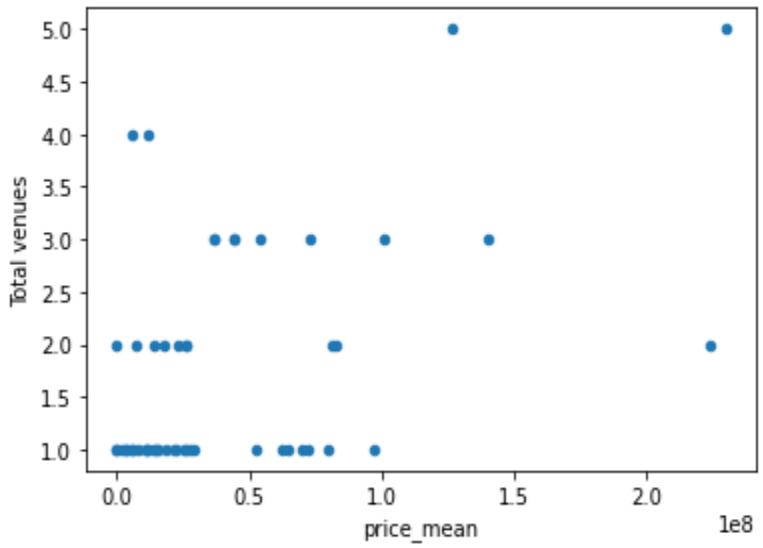
	location	Shopping Mall	Department Store	Bakery	Restaurant	Park	Market	Pharmacy	Total venues	property_type	city	mean_latitude	mean_longitude	price_r
0	Shahrah-e-Pakistan	NaN	2.0	3.0	NaN	NaN	NaN	NaN	5.0	House	Karachi	24.934436	67.072266	1.265000
1	Siraj-ud-Daula Road	NaN	3.0	NaN	NaN	1.0	1.0	NaN	5.0	House	Karachi	24.879682	67.068985	2.300000
2	Blue Bell Residency	1.0	NaN	3.0	NaN	NaN	NaN	NaN	4.0	House	Karachi	24.837117	67.035573	6.000000
3	Maskan Chowrangi	NaN	1.0	2.0	1.0	NaN	NaN	NaN	4.0	House	Karachi	24.934847	67.105248	1.160000
4	Zamzama	2.0	NaN	NaN	NaN	1.0	NaN	NaN	3.0	House	Karachi	24.822799	67.040011	1.005878
6	Darul Aman Society	NaN	2.0	NaN	NaN	1.0	NaN	NaN	3.0	House	Karachi	24.875484	67.068230	1.400000
7	Hill Park	1.0	NaN	1.0	NaN	1.0	NaN	NaN	3.0	House	Karachi	24.868021	67.071147	3.638095
	Shaheed													

CORRELATION OF PARAMETERS

Hypothesis 1

Null Hypothesis $H0$: price_mean is not correlated with Total_venues

Alternative Hypothesis $H1$: price_mean is correlated with Total_venues



Hypothesis 1

Null Hypothesis $H0$: price_mean is not correlated with Total_venues

Alternative Hypothesis $H1$: price_mean is correlated with Total_venues

```
In [46]: import scipy.stats
```

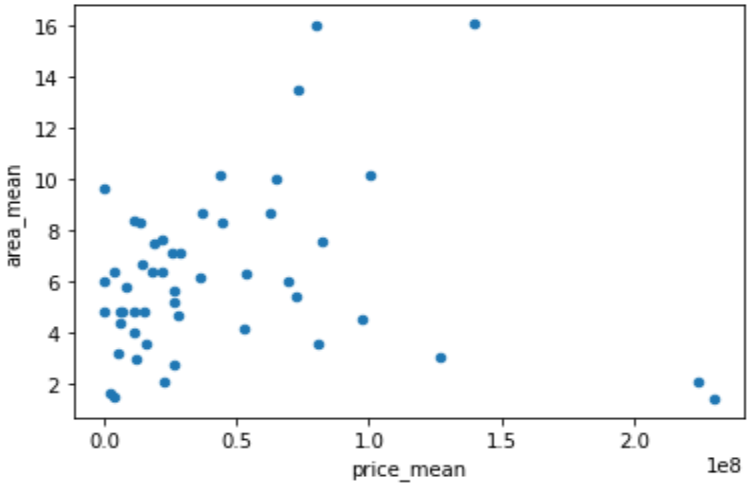
```
In [47]: scipy.stats.pearsonr(df_combined['price_mean'], df_combined['Total venues'])
Out[47]: (0.3845605883805333, 0.004880352224622747)
```

Conclusion: Since the p-value < 0.05, we reject the Null hypothesis and conclude that there exists a relationship between price_mean and Total_venues

Hypothesis 2

Null Hypothesis $H0$: price_mean is not correlated with area_mean

Alternative Hypothesis $H1$: price_mean is correlated with area_mean



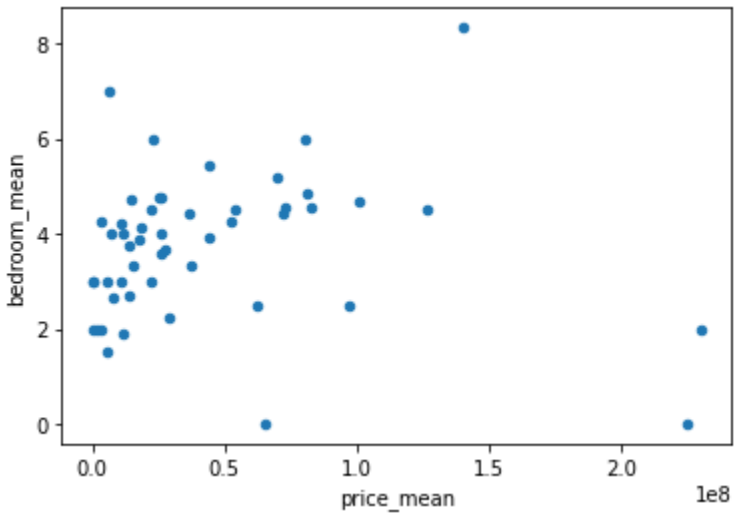

```
In [49]: scipy.stats.pearsonr(df_combined['price_mean'], df_combined['area_mean'])
Out[49]: (0.05193646097101594, 0.7146173762718203)
```

Conclusion: Since the p-value > 0.05, we accept the Null hypothesis and conclude that there exists no relationship between price_mean and area_mean

Hypothesis 3

Null Hypothesis $H0$: price_mean is not correlated bedroom_mean

Alternative Hypothesis $H1$: price_mean is correlated with bedroom_mean




```
In [51]: scipy.stats.pearsonr(df_combined['price_mean'], df_combined['bedroom_mean'])
Out[51]: (-0.033782528434371864, 0.812070976289614)
```

Conclusion: Since the p-value > 0.05, we accept the Null hypothesis and conclude that there exists no relationship between price_mean and bedroom_mean

Next we separately find the top 20 loctions on the basis of:

1. Area
2. Number of venues
3. Number of bedrooms

1. The combined table is sorted in descending order on basis of area_mean and stored as a new dataframe

	location	Shopping Mall	Department Store	Bakery	Restaurant	Park	Market	Pharmacy	Total venues	property_type	city	mean_latitude	mean_longitude	price_m
6	Darul Aman Society	NaN	2.0	NaN	NaN	1.0	NaN	NaN	3.0	House	Karachi	24.875484	67.068230	1.400000e
30	Allama Iqbal Town	NaN	NaN	NaN	NaN	NaN	1.0	NaN	1.0	House	Karachi	24.926879	67.026472	8.000000e
12	Naval Housing Scheme	NaN	NaN	NaN	1.0	NaN	2.0	1.0	3.0	House	Karachi	24.840295	67.057387	7.293821e
4	Zamzama	2.0	NaN	NaN	NaN	1.0	NaN	NaN	3.0	House	Karachi	24.822799	67.040011	1.005878e
10	Khalid Bin Walid Road	1.0	1.0	NaN	NaN	NaN	1.0	NaN	3.0	House	Karachi	24.874922	67.057322	4.392111e

2. The combined table is sorted in descending order on basis of total venues and stored as a new dataframe

	location	Shopping Mall	Department Store	Bakery	Restaurant	Park	Market	Pharmacy	Total venues	property_type	city	mean_latitude	mean_longitude	price_me
0	Shahrah-e-Pakistan	NaN	2.0	3.0	NaN	NaN	NaN	NaN	5.0	House	Karachi	24.934436	67.072266	1.265000e+
1	Siraj-ud-Daula Road	NaN	3.0	NaN	NaN	1.0	1.0	NaN	5.0	House	Karachi	24.879682	67.068985	2.300000e+
2	Blue Bell Residency	1.0	NaN	3.0	NaN	NaN	NaN	NaN	4.0	House	Karachi	24.837117	67.035573	6.000000e+
3	Maskan Chowrangi	NaN	1.0	2.0	1.0	NaN	NaN	NaN	4.0	House	Karachi	24.934847	67.105248	1.160000e+
4	Zamzama	2.0	NaN	NaN	NaN	1.0	NaN	NaN	3.0	House	Karachi	24.822799	67.040011	1.005878e+

3. The combined table is sorted in descending order on basis of bedroom_mean available and stored as a new dataframe

	location	Shopping Mall	Department Store	Bakery	Restaurant	Park	Market	Pharmacy	Total venues	property_type	city	mean_latitude	mean_longitude	price_m
6	Darul Aman Society	NaN	2.0	NaN	NaN	1.0	NaN	NaN	3.0	House	Karachi	24.875484	67.068230	1.400000e
2	Blue Bell Residency	1.0	NaN	3.0	NaN	NaN	NaN	NaN	4.0	House	Karachi	24.837117	67.035573	6.000000e
30	Allama Iqbal Town	NaN	NaN	NaN	NaN	NaN	1.0	NaN	1.0	House	Karachi	24.926879	67.026472	8.000000e
15	Old Clifton	2.0	NaN	NaN	NaN	NaN	NaN	NaN	2.0	House	Karachi	24.817883	67.032915	2.294000e
10	Khalid Bin Walid Road	1.0	1.0	NaN	NaN	NaN	1.0	NaN	3.0	House	Karachi	24.874922	67.057322	4.392111e

Now We analyzed the newly formed 3 datasets sorted on basis of venue counts, area_mean, bedroom_mean respectively. To find the locations which are common in all 3

The three dataset shows different top 20 locations from each analysis, we compare the common localities from top 20 on all 3 sorted dataset to find the locations:

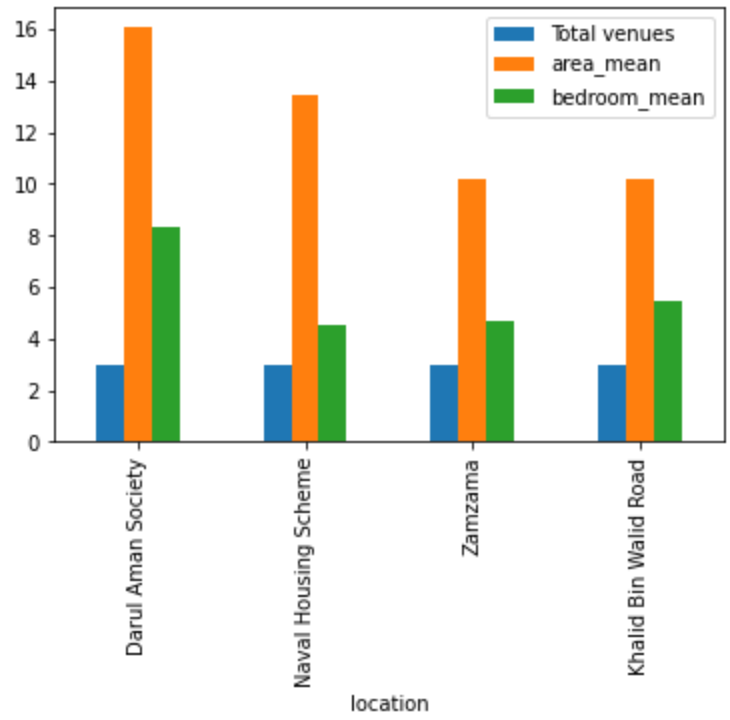
	location
0	Darul Aman Society
1	Naval Housing Scheme
2	Zamzama
3	Khalid Bin Walid Road

The above locations have mores venues and bedrooms and larger houses.

Now for these localities we join with older data set to know other parameters

	location	Shopping Mall	Department Store	Bakery	Restaurant	Park	Market	Pharmacy	Total venues	property_type	city	mean_latitude	mean_longitude	price_mea
0	Darul Aman Society	NaN	2.0	NaN	NaN	1.0	NaN	NaN	3.0	House	Karachi	24.875484	67.068230	1.400000e+
1	Naval Housing Scheme	NaN	NaN	NaN	1.0	NaN	2.0	1.0	3.0	House	Karachi	24.840295	67.057387	7.293821e+
2	Zamzama	2.0	NaN	NaN	NaN	1.0	NaN	NaN	3.0	House	Karachi	24.822799	67.040011	1.005878e+
3	Khalid Bin Walid Road	1.0	1.0	NaN	NaN	NaN	1.0	NaN	3.0	House	Karachi	24.874922	67.057322	4.392111e+

Making a bar graph to compare the 4 locations



RESULTS:

We formulated our result focusing on number of venues, area and number of bedrooms, in this way we found 4 locations based on our data and analysis which are best for buying new homes.

	location
0	Darul Aman Society
1	Naval Housing Scheme
2	Zamzama
3	Khalid Bin Walid Road

A common trend in scatter plots showed that:

1. The price of house is not affected much by the number of venues that are present near

2. The prices somewhat increases as the area of house increase

3. Also the increase number of bedrooms tends to increase the price

Hence we considered a person will be at more convenience and comfort having more venues, larger houses with more bedrooms.

For those people who focus mainly on the venues around a house, and in particular what kind of venues are present around, we clustered the location of Karachi and found the top most common venues of a location, this will also help in people making a decision to choose a neighborhood

DISCUSSION:

In our analysis we have not focused our study to find the cheapest house we have preferred facilities over money.

Analysis can further be done based on the personal priorities

Analysis is done on basis of data available and venues that are showing in Foursquare API, many venues and location may not be shown in analysis, further data would help enhancing our analysis.

CONCLUSION

Using our dataset and after data cleaning and manipulation we found that the locations which have larger houses and more bedrooms and greater venues around them are the best location for buying new home

After all the analysis we found the neighborhoods as shown above, which have best facilities like venues larger houses with more bedrooms are the locations where one should tend to buy a house provided money is not a restraint.

Using our analysis and list of top 10 most common venues of very location a person can buy bigger homes with best features and venues available to him.