In [6]:

```
import pandas as pd
import numpy as np
from bs4 import BeautifulSoup as bs
from requests import get
from urllib.request import urlopen
import json
```

In [69]:

```
urls = []
title=[]
typelist = []
pricelist = []
locationlist =[]
arealist =[]
geolistlat= []
geolistlng= []
zameenid=[]
phonenumber=[]
purpose=[]
hot=[]
title=[]
photoCount=[]
videoCount=[]
agencyID=[]
agencyproduct=[]
isVerified=[]
exactGeography=[]
productScore=[]
```

First lets try with the first page with 25 listings on it

In [50]:

```
one_url = 'https://www.zameen.com/Plots/Lahore-1-1.html'
headers = {'User-Agent':'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KH
response = get(one_url, headers=headers)
soup = bs(response.content, 'html.parser')
jscript = soup.find_all('script')[5].text.strip()[15:-1].split('}};')[0] + '}}'
script = json.loads(jscript)
short = script["algolia"]["content"]["hits"]

for index in range(0, len(short)):
    geolistlat.append(short[index]['geography']['lat'])
    geolistlng.append(short[index]['geography']['lng'])
    pricelist.append(short[index]['price'])
    arealist.append(short[index]['area'])
    zameenid.append(short[index]['externalID'])
    title.append(short[index]['title'])
    purpose.append(short[index]['purpose'])
    photoCount.append(short[index]['photoCount'])
    videoCount.append(short[index]['videoCount'])
    agencyID.append(short[index]["agency"]['externalID'])
    agencyproduct.append(short[index]["agency"]['product'])
    isVerified.append(short[index]['isVerified'])
    productScore.append(short[index]["agency"]['productScore'])
    locationlist.append(short[index]['location'][-1]['slug'])
    typelist.append(short[index]['category'][-1]['nameSingular'])
```

In [51]:

```
len(locationlist)
```

Out[51]:

25

In [52]:

```
locationlist
```

Out[52]:

```
['/Lahore_Central_Park_Housing_Scheme-1013',
 '/Lahore_Bahria_Nasheman-4510',
 '/Lahore_DHA_Defence_DHA_9_Town_DHA_9_Town___Block_C-8175',
 '/Lahore_Central_Park_Housing_Scheme-1013',
 '/Lahore_Raiwind_Road_Lake_City_Lake_City___Sector_M_3-8346',
 '/Lahore_Bahria_Orchard_Bahria_Orchard_Phase_2_Low_Cost_Sector_Low_Cost__
_Block_D_Extension-12389',
 '/Lahore_Bahria_Town_Bahria_Town___Sector_F_Bahria_Town___Tauheed_Block-1
1864',
 '/Lahore_Bahria_Orchard_Bahria_Orchard_Phase_2_Low_Cost_Sector_Low_Cost__
_Block_D_Extension-12389',
 '/Lahore_Raiwind_Road_Al_Kabir_Town_Al_Kabir_Town___Phase_2_Al_Kabir_Phas
e_2___Block_E-13022',
 '/Lahore_DHA_11_Rahbar_DHA_11_Rahbar_Phase_2-3717',
 '/Lahore_Raiwind_Road_Lake_City_Lake_City___Golf_Estate_2-15748',
 '/Lahore_Raiwind_Road_Lake_City_Lake_City___Sector_M_7_Lake_City___Sector
_M7___Block_C1-15743',
 '/Lahore_New_Lahore_City-8152',
 '/Lahore_Park_View_City-1466',
 '/Lahore_Defence__DHA__Phase_6_Block_C-1612',
 '/Lahore_DHA_Defence_DHA_Phase_5_DHA_Phase_5___Block_M_Extension-15734',
 '/Lahore_DHA_Defence_DHA_Phase_9_Prism_DHA_Phase_9_Prism___Block_A-9268',
 '/Lahore_Bedian_Road-378',
 '/Lahore_Bahria_Orchard_Bahria_Orchard_Phase_4_Bahria_Orchard_Phase_4___B
lock_G6-15718',
 '/Lahore_Bahria_Orchard_Bahria_Orchard_Phase_4_Bahria_Orchard_Phase_4___B
lock_G6-15718',
 '/Lahore_Raiwind_Road_Al_Kabir_Town_Al_Kabir_Town___Phase_2-11704',
 '/Lahore_Raiwind_Road_Al_Kabir_Town_Al_Kabir_Town___Phase_2-11704',
 '/Lahore_Raiwind_Road_Al_Kabir_Town_Al_Kabir_Town___Phase_2-11704',
 '/Lahore_Bahria_Orchard_Bahria_Orchard_Phase_1_Bahria_Orchard_Phase_1___E
astern-8239',
 '/Lahore_DHA_Defence_DHA_Phase_8_DHA_Phase_8___Block_S-1634']
```

In [14]:

```
for index in range(0, len(short)):
    geolistlat.append(short[index]['geography']['lat'])
    geolistlng.append(short[index]['geography']['lng'])
    pricelist.append(short[index]['price'])
    arealist.append(short[index]['area'])
    zameenid.append(short[index]['externalID'])
    title.append(short[index]['title'])
    purpose.append(short[index]['purpose'])
    photoCount.append(short[index]['photoCount'])
    videoCount.append(short[index]['videoCount'])
    agencyID.append(short[index]["agency"]['externalID'])
    agencyproduct.append(short[index]["agency"]['product'])
    isVerified.append(short[index]['isVerified'])
    productScore.append(short[index]["agency"]['productScore'])
    locationlist.append(short[index]['location'][len(script[index]['location'])-1]['slug'
    typelist.append(short[index]['category'][len(script[index]['category'])-1]['nameSingu
```

In [18]:

```
firstpg = pd.DataFrame(list(zip(geolistlat,geolistlng,pricelist,arealist,zameenid,title,p
                                agencyproduct,isVerified,productScore,locationlist,typeli
                        columns=['Latitude','Longitude','Price','Area SqM','Zameen ID','Tit
                                 'VideoCount','Agency ID','Agency Product','Is Verified','Pr

firstpg
```

Out[18]:

	Latitude	Longitude	Price	Area SqM	Zameen ID	Title	Purpose	PhotoCount	VideoCount	A
0	31.329231	74.226724	2500000	83.612736	44485679	4 Marla Commercial Plot File On Main Boulevard...	for-sale	32	1	1
1	31.502056	74.429791	242500000	167.225472	42162705	8 Marla Possession Plot On Main Commercial 200...	for-sale	54	1	1
2	31.321674	74.380860	3200000	104.515920	44825697	A 5 Marla Residential Plot For Sale At Prime L...	for-sale	7	1	1

Now Lets Do for All Pages

In [70]:

```
for i in range(1,1492):
    link= 'https://www.zameen.com/Plots/Lahore-1-' + str(i) + '.html'
    urls.append(link)
```

In [72]:

```
for p in urls:
    try:
        headers = {'User-Agent':'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/53
        response = get(p, headers=headers)
        soup = bs(response.content, 'html.parser')
        jscript = soup.find_all('script')[5].text.strip()[15:-1].split('}};')[0] + '}}'
        script = json.loads(jscript)
        short = script["algolia"]["content"]["hits"]

        for index in range(0, len(short)):

            if short[index]['geography']['lat']:
                geolistlat.append(short[index]['geography']['lat'])
            else:
                geolistlat.append(None)

            if short[index]['geography']['lng']:
                geolistlng.append(short[index]['geography']['lng'])
            else:
                geolistlng.append(None)

            if short[index]['price']:
                pricelist.append(short[index]['price'])
            else:
                pricelist.append(None)

            if short[index]['area']:
                arealist.append(short[index]['area'])
            else:
                arealist.append(None)

            if short[index]['externalID']:
                zameenid.append(short[index]['externalID'])
            else:
                zameenid.append(None)

            if short[index]['title']:
                title.append(short[index]['title'])
            else:
                title.append(None)

            if short[index]['purpose']:
                purpose.append(short[index]['purpose'])
            else:
                purpose.append(None)

            if short[index]['photoCount']:
                photoCount.append(short[index]['photoCount'])
            else:
                photoCount.append(None)

            if short[index]['videoCount']:
                videoCount.append(short[index]['videoCount'])
            else:
                videoCount.append(None)

            if short[index]["agency"]['externalID']:
                agencyID.append(short[index]["agency"]['externalID'])
            else:
```

```
                agencyID.append(None)

            if short[index]["agency"]['product']:
                agencyproduct.append(short[index]["agency"]['product'])
            else:
                agencyproduct.append(None)

            if short[index]['isVerified']:
                isVerified.append(short[index]['isVerified'])
            else:
                isVerified.append(None)

            if short[index]["agency"]['productScore']:
                productScore.append(short[index]["agency"]['productScore'])
            else:
                productScore.append(None)

            if short[index]['location'][-1]['slug']:
                locationlist.append(short[index]['location'][-1]['slug'])
            else:
                locationlist.append(None)

            if short[index]['category'][-1]['nameSingular']:
                typelist.append(short[index]['category'][-1]['nameSingular'])
            else:
                typelist.append(None)


    except:
        print('Exception for: ' + p)
```

```
Exception for: https://www.zameen.com/Plots/Lahore-1-1.html (https://ww
w.zameen.com/Plots/Lahore-1-1.html)
Exception for: https://www.zameen.com/Plots/Lahore-1-8.html (https://ww
w.zameen.com/Plots/Lahore-1-8.html)
Exception for: https://www.zameen.com/Plots/Lahore-1-9.html (https://ww
w.zameen.com/Plots/Lahore-1-9.html)
Exception for: https://www.zameen.com/Plots/Lahore-1-11.html (https://ww
w.zameen.com/Plots/Lahore-1-11.html)
Exception for: https://www.zameen.com/Plots/Lahore-1-12.html (https://ww
w.zameen.com/Plots/Lahore-1-12.html)
Exception for: https://www.zameen.com/Plots/Lahore-1-13.html (https://ww
w.zameen.com/Plots/Lahore-1-13.html)
Exception for: https://www.zameen.com/Plots/Lahore-1-15.html (https://ww
w.zameen.com/Plots/Lahore-1-15.html)
Exception for: https://www.zameen.com/Plots/Lahore-1-18.html (https://ww
w.zameen.com/Plots/Lahore-1-18.html)
Exception for: https://www.zameen.com/Plots/Lahore-1-19.html (https://ww
w.zameen.com/Plots/Lahore-1-19.html)
Exception for: https://www.zameen.com/Plots/Lahore-1-20.html (https://ww
```

In [80]:

```
all_pages= pd.DataFrame(list(zip(geolistlat,geolistlng,pricelist,arealist,zameenid,title,
                                 agencyproduct,isVerified,productScore,locationlist,typel
                        columns=['Latitude','Longitude','Price','Area SqM','Zameen ID','T
                                 'VideoCount','Agency ID','Agency Product','Is Verified',
```

In [81]:

```
all_pages
```

Out[81]:

	Latitude	Longitude	Price	Area SqM	Zameen ID	Title	Purpose	PhotoCount	VideoCount
0	31.343674	74.230522	1500000	62.709552	45333940	Grab Your Chance to Invest in a Prime Location...	for-sale	14.0	NaN
1	31.343674	74.230522	2500000	104.515920	45338530	A Golden Opportunity To Upgrade Your Investmen...	for-sale	17.0	NaN
2	31.346332	74.248640	7000000	104.515920	44685399	Ideal Location, 5 Marla Residential Plot Is Av...	for-sale	24.0	1.0

In [1]:

```
all_pages.to_csv('zameen_properties_scrape.csv')
```

```
---------------------------------------------------------------------------
-
NameError                                 Traceback (most recent call las
t)
~\AppData\Local\Temp/ipykernel_12916/1261810286.py in <module>
----> 1 all_pages.to_csv('zameen_properties_scrape.csv')

NameError: name 'all_pages' is not defined
```

In [ ]: