***Actual dataset:***

<https://www.propertypriceregister.ie/website/npsra/pprweb.nsf/page/ppr-home-en>

***Ethical considerations:***

<http://psr.ie/en/PSRA/Pages/Re-Use_of_Information>

***Additional data to be added to the dataset:***

* Province data: <https://www.irelandtownslist.com/>
* Latitude and longitude:

1. Add county name to each address.
2. Google API <https://developers.google.com/maps/documentation/geocoding/start>

<https://towardsdatascience.com/an-analysis-of-property-prices-in-ireland-6fc34a56ac87>

1. Python library GeoPy <https://www.thepythoncode.com/article/get-geolocation-in-python>

<https://www.askpython.com/python/python-geopy-to-find-geocode-of-an-address>

* Unemployment data

16/03/2022

1. Merging province data
2. Finding dataset used in the previous works
3. Latitude and longitude coding using Google API
4. Find previous studies related to Ireland or small countries like that.

<https://towardsdatascience.com/predicting-house-prices-with-machine-learning-62d5bcd0d68f>

<https://towardsdatascience.com/predict-house-prices-with-machine-learning-5b475db4e1e>

**import** pandas **as** pd

**import** numpy **as** np

In [7]:

df **=** pd**.**read\_csv('PPR-ALL.csv', encoding**=**'latin-1')

In [8]:

df**.**head()

Out[8]:

|  | **Date of Sale (dd/mm/yyyy)** | **Address** | **Postal Code** | **County** | **Price (Û)** | **Not Full Market Price** | **VAT Exclusive** | **Description of Property** | **Property Size Description** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1/1/10 | 5 Braemor Drive, Churchtown, Co.Dublin | NaN | Dublin | 343000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |
| **1** | 3/1/10 | 134 Ashewood Walk, Summerhill Lane, Portlaoise | NaN | Laois | 185000.0 | No | Yes | New Dwelling house /Apartment | greater than or equal to 38 sq metres and less... |
| **2** | 4/1/10 | 1 Meadow Avenue, Dundrum, Dublin 14 | NaN | Dublin | 438500.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |
| **3** | 4/1/10 | 1 The Haven, Mornington | NaN | Meath | 400000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |
| **4** | 4/1/10 | 11 Melville Heights, Kilkenny | NaN | Kilkenny | 160000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |

In [9]:

df**.**tail()

Out[9]:

|  | **Date of Sale (dd/mm/yyyy)** | **Address** | **Postal Code** | **County** | **Price (Û)** | **Not Full Market Price** | **VAT Exclusive** | **Description of Property** | **Property Size Description** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **342502** | 14/12/2018 | MOUNTAIN CHASE, KILMIHILL, KILMALLOCK | NaN | Limerick | 120000.0 | Yes | No | Second-Hand Dwelling house /Apartment | NaN |
| **342503** | 14/12/2018 | NO. 63 WOODSTREAM, COOLANEY, SLIGO | NaN | Sligo | 91000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |
| **342504** | 14/12/2018 | SCAR, DUNCORMICK, WEXFORD | NaN | Wexford | 360000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |
| **342505** | 14/12/2018 | THE DIAMOND, CLONES, MONAGHAN | NaN | Monaghan | 45100.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |
| **342506** | 14/12/2018 | TORC, BALLYKEA, SKERRIES | NaN | Dublin | 585000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |

In [11]:

counties **=** df['County']**.**value\_counts()

In [13]:

counties

Out[13]:

Dublin 112072

Cork 37298

Galway 17625

Kildare 16289

Meath 12927

Limerick 12788

Wexford 11543

Wicklow 10367

Kerry 10062

Donegal 9459

Tipperary 8970

Waterford 8951

Louth 8931

Mayo 8151

Clare 7965

Westmeath 6558

Cavan 5441

Kilkenny 5378

Laois 5158

Sligo 5050

Roscommon 4899

Offaly 4060

Carlow 3694

Leitrim 3179

Longford 2973

Monaghan 2719

Name: County, dtype: int64

In [17]:

**import** requests

In [37]:

response**=**requests**.**get('https://maps.googleapis.com/maps/api/geocode/json?address=1600+Amphitheatre+Parkway,+Mountain+View,+CA&key=AIzaSyDHNguaHa1brL6IKKr2EkDDnym32RfaX8E')

In [38]:

print(response)

<Response [200]>

In [39]:

response**.**json()

Out[39]:

{'results': [{'address\_components': [{'long\_name': '1600',

'short\_name': '1600',

'types': ['street\_number']},

{'long\_name': 'Amphitheatre Parkway',

'short\_name': 'Amphitheatre Pkwy',

'types': ['route']},

{'long\_name': 'Mountain View',

'short\_name': 'Mountain View',

'types': ['locality', 'political']},

{'long\_name': 'Santa Clara County',

'short\_name': 'Santa Clara County',

'types': ['administrative\_area\_level\_2', 'political']},

{'long\_name': 'California',

'short\_name': 'CA',

'types': ['administrative\_area\_level\_1', 'political']},

{'long\_name': 'United States',

'short\_name': 'US',

'types': ['country', 'political']},

{'long\_name': '94043', 'short\_name': '94043', 'types': ['postal\_code']}],

'formatted\_address': '1600 Amphitheatre Pkwy, Mountain View, CA 94043, USA',

'geometry': {'location': {'lat': 37.4230414, 'lng': -122.0840448},

'location\_type': 'ROOFTOP',

'viewport': {'northeast': {'lat': 37.42439038029149,

'lng': -122.0826958197085},

'southwest': {'lat': 37.42169241970849, 'lng': -122.0853937802915}}},

'place\_id': 'ChIJhehRjJ-5j4ARKFssUSrwnhY',

'plus\_code': {'compound\_code': 'CWF7+7J Mountain View, California, United States',

'global\_code': '849VCWF7+7J'},

'types': ['street\_address']}],

'status': 'OK'}

In [40]:

response**.**json()['results'][0]['geometry']['location']

Out[40]:

{'lat': 37.4230414, 'lng': -122.0840448}

In [41]:

df**.**head()

Out[41]:

|  | **Date of Sale (dd/mm/yyyy)** | **Address** | **Postal Code** | **County** | **Price (Û)** | **Not Full Market Price** | **VAT Exclusive** | **Description of Property** | **Property Size Description** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1/1/10 | 5 Braemor Drive, Churchtown, Co.Dublin | NaN | Dublin | 343000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |
| **1** | 3/1/10 | 134 Ashewood Walk, Summerhill Lane, Portlaoise | NaN | Laois | 185000.0 | No | Yes | New Dwelling house /Apartment | greater than or equal to 38 sq metres and less... |
| **2** | 4/1/10 | 1 Meadow Avenue, Dundrum, Dublin 14 | NaN | Dublin | 438500.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |
| **3** | 4/1/10 | 1 The Haven, Mornington | NaN | Meath | 400000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |
| **4** | 4/1/10 | 11 Melville Heights, Kilkenny | NaN | Kilkenny | 160000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN |

In [42]:

df['Address'][0]**.**split()

Out[42]:

['5', 'Braemor', 'Drive,', 'Churchtown,', 'Co.Dublin']

In [ ]:

url**=**'https://maps.googleapis.com/maps/api/geocode/json?address=1600+Amphitheatre+Parkway,+Mountain+View,+CA&key=AIzaSyDHNguaHa1brL6IKKr2EkDDnym32RfaX8E'

In [43]:

url**=**'https://maps.googleapis.com/maps/api/geocode/json?address='

api\_key**=**'&key=AIzaSyDHNguaHa1brL6IKKr2EkDDnym32RfaX8E'

In [44]:

df['Address'][0]**.**split()

Out[44]:

['5', 'Braemor', 'Drive,', 'Churchtown,', 'Co.Dublin']

In [56]:

a**=**[x **+** '+' **for** x **in** df['Address'][0]**.**split()[:**-**1]]

In [57]:

a**.**append(df['Address'][0]**.**split()[**-**1])

In [58]:

a

Out[58]:

['5+', 'Braemor+', 'Drive,+', 'Churchtown,+', 'Co.Dublin']

In [66]:

address**=**""**.**join(a)

In [67]:

full\_path**=**url**+**address**+**api\_key

In [68]:

print(full\_path)

https://maps.googleapis.com/maps/api/geocode/json?address=5+Braemor+Drive,+Churchtown,+Co.Dublin&key=AIzaSyDHNguaHa1brL6IKKr2EkDDnym32RfaX8E

In [69]:

requests**.**get(full\_path)**.**json()

Out[69]:

{'results': [{'address\_components': [{'long\_name': '5',

'short\_name': '5',

'types': ['street\_number']},

{'long\_name': 'Braemor Drive',

'short\_name': 'Braemor Dr',

'types': ['route']},

{'long\_name': 'Newtown Little',

'short\_name': 'Newtown Little',

'types': ['neighborhood', 'political']},

{'long\_name': 'Dublin 14',

'short\_name': 'Dublin 14',

'types': ['postal\_town']},

{'long\_name': 'County Dublin',

'short\_name': 'County Dublin',

'types': ['administrative\_area\_level\_1', 'political']},

{'long\_name': 'Ireland',

'short\_name': 'IE',

'types': ['country', 'political']},

{'long\_name': 'D14 NX40',

'short\_name': 'D14 NX40',

'types': ['postal\_code']}],

'formatted\_address': '5 Braemor Dr, Newtown Little, Dublin 14, D14 NX40, Ireland',

'geometry': {'location': {'lat': 53.302391, 'lng': -6.263783099999999},

'location\_type': 'ROOFTOP',

'viewport': {'northeast': {'lat': 53.3037399802915,

'lng': -6.262434119708497},

'southwest': {'lat': 53.3010420197085, 'lng': -6.265132080291502}}},

'place\_id': 'ChIJ2V8XtvwLZ0gRpl6rLdYnd3Y',

'plus\_code': {'compound\_code': '8P2P+XF Dublin, County Dublin, Ireland',

'global\_code': '9C5M8P2P+XF'},

'types': ['street\_address']}],

'status': 'OK'}

In [91]:

**def** url\_creator(address, url**=**'https://maps.googleapis.com/maps/api/geocode/json?address=', api\_key**=**'&key=AIzaSyDHNguaHa1brL6IKKr2EkDDnym32RfaX8E'):

add\_list **=** address**.**split()

add\_list **=** [line **+** '+' **for** line **in** add\_list[:**-**1]]

add\_list**.**append(address**.**split()[**-**1])

add\_url**=**""**.**join(add\_list)

**return** url**+**add\_url**+**api\_key

In [85]:

url\_creator(df['Address'][1])

Out[85]:

'https://maps.googleapis.com/maps/api/geocode/json?address=134+Ashewood+Walk,+Summerhill+Lane,+Portlaoise&key=AIzaSyDHNguaHa1brL6IKKr2EkDDnym32RfaX8E'

In [87]:

df['api\_url']**=**df['Address']**.**apply(url\_creator)

In [88]:

df**.**head()

Out[88]:

|  | **Date of Sale (dd/mm/yyyy)** | **Address** | **Postal Code** | **County** | **Price (Û)** | **Not Full Market Price** | **VAT Exclusive** | **Description of Property** | **Property Size Description** | **api\_url** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1/1/10 | 5 Braemor Drive, Churchtown, Co.Dublin | NaN | Dublin | 343000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN | https://maps.googleapis.com/maps/api/geocode/j... |
| **1** | 3/1/10 | 134 Ashewood Walk, Summerhill Lane, Portlaoise | NaN | Laois | 185000.0 | No | Yes | New Dwelling house /Apartment | greater than or equal to 38 sq metres and less... | https://maps.googleapis.com/maps/api/geocode/j... |
| **2** | 4/1/10 | 1 Meadow Avenue, Dundrum, Dublin 14 | NaN | Dublin | 438500.0 | No | No | Second-Hand Dwelling house /Apartment | NaN | https://maps.googleapis.com/maps/api/geocode/j... |
| **3** | 4/1/10 | 1 The Haven, Mornington | NaN | Meath | 400000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN | https://maps.googleapis.com/maps/api/geocode/j... |
| **4** | 4/1/10 | 11 Melville Heights, Kilkenny | NaN | Kilkenny | 160000.0 | No | No | Second-Hand Dwelling house /Apartment | NaN | https://maps.googleapis.com/maps/api/geocode/j... |

In [129]:

**def** lat\_lng(api\_url):

**try**:

response **=** requests**.**get(api\_url)

**except** requests**.**exceptions**.**RequestException **as** e:

print(e)

**raise**

response**.**raise\_for\_status()

results **=** response**.**json()['results']

**if** len(results) **!=** 0:

**return** results[0]['geometry']['location']['lat'], results[0]['geometry']['location']['lng']

**else**:

**return** np**.**nan, np**.**nan

In [130]:

lat\_lng(url\_creator(df['Address'][7]))

Out[130]:

(52.82499970000001, -6.895761100000001)

In [131]:

df['lat'], df['lng'] **=** df['api\_url']**.**apply(lat\_lng)

---------------------------------------------------------------------------

TypeError Traceback (most recent call last)

/anaconda3/envs/housing/lib/python3.7/site-packages/urllib3/connectionpool.py in \_make\_request(self, conn, method, url, timeout, chunked, \*\*httplib\_request\_kw)

**376** try: # Python 2.7, use buffering of HTTP responses

--> 377 httplib\_response = conn.getresponse(buffering=True)

**378** except TypeError: # Python 3

TypeError: getresponse() got an unexpected keyword argument 'buffering'

During handling of the above exception, another exception occurred:

KeyboardInterrupt Traceback (most recent call last)

<ipython-input-131-a33d1772e4e9> in <module>

----> 1 df['lat'], df['lng'] = df['api\_url'].apply(lat\_lng)

/anaconda3/envs/housing/lib/python3.7/site-packages/pandas/core/series.py in apply(self, func, convert\_dtype, args, \*\*kwds)

**3192** else:

**3193** values = self.astype(object).values

-> 3194 mapped = lib.map\_infer(values, f, convert=convert\_dtype)

**3195**

**3196** if len(mapped) and isinstance(mapped[0], Series):

pandas/\_libs/src/inference.pyx in pandas.\_libs.lib.map\_infer()

<ipython-input-129-2c6b468abff9> in lat\_lng(api\_url)

**1** def lat\_lng(api\_url):

**2** try:

----> 3 response = requests.get(api\_url)

**4** except requests.exceptions.RequestException as e:

**5** print(e)

/anaconda3/envs/housing/lib/python3.7/site-packages/requests/api.py in get(url, params, \*\*kwargs)

**73**

**74** kwargs.setdefault('allow\_redirects', True)

---> 75 return request('get', url, params=params, \*\*kwargs)

**76**

**77**

/anaconda3/envs/housing/lib/python3.7/site-packages/requests/api.py in request(method, url, \*\*kwargs)

**58** # cases, and look like a memory leak in others.

**59** with sessions.Session() as session:

---> 60 return session.request(method=method, url=url, \*\*kwargs)

**61**

**62**

/anaconda3/envs/housing/lib/python3.7/site-packages/requests/sessions.py in request(self, method, url, params, data, headers, cookies, files, auth, timeout, allow\_redirects, proxies, hooks, stream, verify, cert, json)

**531** }

**532** send\_kwargs.update(settings)

--> 533 resp = self.send(prep, \*\*send\_kwargs)

**534**

**535** return resp

/anaconda3/envs/housing/lib/python3.7/site-packages/requests/sessions.py in send(self, request, \*\*kwargs)

**644**

**645** # Send the request

--> 646 r = adapter.send(request, \*\*kwargs)

**647**

**648** # Total elapsed time of the request (approximately)

/anaconda3/envs/housing/lib/python3.7/site-packages/requests/adapters.py in send(self, request, stream, timeout, verify, cert, proxies)

**447** decode\_content=False,

**448** retries=self.max\_retries,

--> 449 timeout=timeout

**450** )

**451**

/anaconda3/envs/housing/lib/python3.7/site-packages/urllib3/connectionpool.py in urlopen(self, method, url, body, headers, retries, redirect, assert\_same\_host, timeout, pool\_timeout, release\_conn, chunked, body\_pos, \*\*response\_kw)

**598** timeout=timeout\_obj,

**599** body=body, headers=headers,

--> 600 chunked=chunked)

**601**

**602** # If we're going to release the connection in ``finally:``, then

/anaconda3/envs/housing/lib/python3.7/site-packages/urllib3/connectionpool.py in \_make\_request(self, conn, method, url, timeout, chunked, \*\*httplib\_request\_kw)

**378** except TypeError: # Python 3

**379** try:

--> 380 httplib\_response = conn.getresponse()

**381** except Exception as e:

**382** # Remove the TypeError from the exception chain in Python 3;

/anaconda3/envs/housing/lib/python3.7/http/client.py in getresponse(self)

**1319** try:

**1320** try:

-> 1321 response.begin()

**1322** except ConnectionError:

**1323** self.close()

/anaconda3/envs/housing/lib/python3.7/http/client.py in begin(self)

**294** # read until we get a non-100 response

**295** while True:

--> 296 version, status, reason = self.\_read\_status()

**297** if status != CONTINUE:

**298** break

/anaconda3/envs/housing/lib/python3.7/http/client.py in \_read\_status(self)

**255**

**256** def \_read\_status(self):

--> 257 line = str(self.fp.readline(\_MAXLINE + 1), "iso-8859-1")

**258** if len(line) > \_MAXLINE:

**259** raise LineTooLong("status line")

/anaconda3/envs/housing/lib/python3.7/socket.py in readinto(self, b)

**587** while True:

**588** try:

--> 589 return self.\_sock.recv\_into(b)

**590** except timeout:

**591** self.\_timeout\_occurred = True

/anaconda3/envs/housing/lib/python3.7/site-packages/urllib3/contrib/pyopenssl.py in recv\_into(self, \*args, \*\*kwargs)

**292** def recv\_into(self, \*args, \*\*kwargs):

**293** try:

--> 294 return self.connection.recv\_into(\*args, \*\*kwargs)

**295** except OpenSSL.SSL.SysCallError as e:

**296** if self.suppress\_ragged\_eofs and e.args == (-1, 'Unexpected EOF'):

/anaconda3/envs/housing/lib/python3.7/site-packages/OpenSSL/SSL.py in recv\_into(self, buffer, nbytes, flags)

**1811** result = \_lib.SSL\_peek(self.\_ssl, buf, nbytes)

**1812** else:

-> 1813 result = \_lib.SSL\_read(self.\_ssl, buf, nbytes)

**1814** self.\_raise\_ssl\_error(self.\_ssl, result)

**1815**

KeyboardInterrupt:

In [ ]: