

Exercises

This notebook focuses on **data understanding and preparation** for a particular problem and dataset. The data comes from the Residential Property Price Register (RPPR) (<https://www.propertypriceregister.ie/website/npsra/prweb.nsf/page/ppr-home-en>). The Residential Property Price Register is produced by the Property Services Regulatory Authority (PSRA). It includes Date of Sale, Price and Address of all residential properties purchased in Ireland since the 1st January 2010, until 5th May 2023, as declared to the Revenue Commissioners for stamp duty purposes. In this analysis, **we focus on using the data collected by RPPR to build a data analytics solution for residential property price prediction.**

We will work with the full set of the data collected in a CSV file. There are 4 parts in this notebook:

- (1). Prepare a data quality report for your CSV file.
- (2). Prepare a data quality plan for the cleaned CSV file.
- (3). Exploring relationships between feature pairs.
- (4). Create a few new features (at least 3) with the aim to better capture the problem domain and the target outcome.

```
In [1]: # import packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
from matplotlib.backends.backend_pdf import PdfPages
```

```
In [2]: # load dataset
df = pd.read_csv('PPR-ALL.csv', encoding = 'unicode_escape')

/var/folders/bn/6x11zkt6n14vfwybjymgt00000gp/T/ipykernel_31017/4033051617.py:2: DtypeWarning: Columns (3,8) have mixed types. Specify dtype option on import or set low_memory=False.
df = pd.read_csv('PPR-ALL.csv', encoding = 'unicode_escape')
```

```
In [3]: # show dataset details
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 596528 entries, 0 to 596527
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
0   Date of Sale (dd/mm/yyyy)  596528 non-null   object 
1   Address          596528 non-null   object 
2   County           596528 non-null   object 
3   Eircode          91579 non-null   object 
4   Price ()         596528 non-null   object 
5   Not Full Market Price  596528 non-null   object 
6   VAT Exclusive    596528 non-null   object 
7   Description of Property 596528 non-null   object 
8   Property Size Description 52816 non-null   object 
dtypes: object(9)
memory usage: 41.0+ MB
```

1 - Prepare a data quality report for your CSV file.

- Check how many rows and columns your CSV has.

```
In [4]: # Check how many rows and columns this dataframe has
df.shape
```

```
Out[4]: (596528, 9)
```

- Print the first and the last 5 rows.

```
In [5]: # check first 5 rows
df.head(5)
```

	Date of Sale (dd/mm/yyyy)	Address	County	Eircode	Price ()	Not Full Market Price	VAT Exclusive	Description of Property	Property Size Description
0	01/01/2010	5 Braemor Drive, Churchtown, Co.Dublin	Dublin	Nan	343,000.00	No	No	Second-Hand Dwelling house /Apartment	NaN
1	03/01/2010	134 Ashewood Walk, Summerhill Lane, Portlaoise	Laois	Nan	185,000.00	No	Yes	New Dwelling house /Apartment	greater than or equal to 38 sq metres and less...
2	04/01/2010	1 Meadow Avenue, Dundrum, Dublin 14	Dublin	Nan	438,500.00	No	No	Second-Hand Dwelling house /Apartment	NaN
3	04/01/2010	1 The Haven, Mornington	Meath	Nan	400,000.00	No	No	Second-Hand Dwelling house /Apartment	NaN
4	04/01/2010	11 Melville Heights, Kilkenny	Kilkenny	Nan	160,000.00	No	No	Second-Hand Dwelling house /Apartment	NaN

```
In [6]: # check last 5 rows
df.tail(5)
```

	Date of Sale (dd/mm/yyyy)	Address	County	Eircode	Price ()	Not Full Market Price	VAT Exclusive	Description of Property	Property Size Description
596523	05/05/2023	Palmersgate, Kennelsfort Road Lower, Palmerstown	Dublin	Nan	2,116,150.75	No	No	New Dwelling house /Apartment	NaN
596524	05/05/2023	PARK RD, SWINFORD, MAYO	Mayo	F12F992	190,000.00	No	No	Second- Hand Dwelling house /Apartment	NaN
596525	05/05/2023	SPRING COTTAGE, SHALLON, THE WARD	Dublin	D11P582	740,000.00	No	No	Second- Hand Dwelling house /Apartment	NaN
596526	05/05/2023	THE ORCHARD, GARRYHILL, CARLOW	Carlow	R21H683	375,000.00	No	No	Second- Hand Dwelling house /Apartment	NaN
596527	05/05/2023	TULLYCROSS MOYDRUM, ATHLONE, WESTMEATH	Westmeath	N37TY28	381,000.00	No	No	Second- Hand Dwelling house /Apartment	NaN

- Convert the features to their appropriate data types (e.g., decide which features are more appropriate as continuous and which ones as categorical types).

```
In [7]: # Firstly look at column names
df.columns
```

```
Out[7]: Index(['Date of Sale (dd/mm/yyyy)', 'Address', 'County', 'Eircode',
       'Price ()', 'Not Full Market Price', 'VAT Exclusive',
       'Description of Property', 'Property Size Description'],
       dtype='object')
```

```
In [8]: # Clean the column names to remove white space after the name or in the name
df.columns = df.columns.str.replace(' ', '')

# Look at column names
df.columns
```

```
Out[8]: Index(['DateofSale(dd/mm/yyyy)', 'Address', 'County', 'Eircode', 'Price()', 
       'NotFullMarketPrice', 'VATEExclusive', 'DescriptionofProperty',
       'PropertySizeDescription'],
       dtype='object')
```

```
In [9]: # check type of each feature
df.dtypes
```

```
Out[9]: DateofSale(dd/mm/yyyy)      object
Address          object
County           object
Eircode          object
Price()          object
NotFullMarketPrice    object
VATExclusive     object
DescriptionofProperty   object
PropertySizeDescription  object
dtype: object
```

Date of Sale (dd/mm/yyyy) should have datetime data. I might create two separate columns later that identify both the date and month, in which case it will be useful to have datetime data. Datetime data will also make it possible to plot this feature on a histogram.

```
In [10]: # convert date of sale data from string to datetime data type
df['DateofSale(dd/mm/yyyy)'] = pd.to_datetime(df['DateofSale(dd/mm/yyyy)'], format="%d/%m/%Y")
```

- I will rename the date and price columns to remove special characters, making them easier to manipulate.

```
In [11]: # rename columns
df = df.rename({'Price()': 'Price', 'DateofSale(dd/mm/yyyy)': 'DateofSale'}, axis='columns')
df.columns
```

```
Out[11]: Index(['DateofSale', 'Address', 'County', 'Eircode', 'Price',
       'NotFullMarketPrice', 'VATExclusive', 'DescriptionofProperty',
       'PropertySizeDescription'],
      dtype='object')
```

- Price data is currently string data when it should be of type float in order to be considered continuous data. To convert this feature's data type, I will remove '€' from each row in the Price column.

```
In [12]: # remove '€' and ',' characters from each row of Price and convert its data type to float
df['Price'] = df['Price'].replace('[\,\,]', '', regex=True)
df['Price'] = df['Price'].astype('float64')
```

```
In [13]: # check first 5 rows
df.head(5)
```

	DateofSale	Address	County	Eircode	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescr
0	2010-01-01	5 Braemor Drive, Churchtown, Co.Dublin	Dublin	NaN	343000.0	No	No	Second-Hand Dwelling house /Apartment	
1	2010-01-03	134 Ashewood Walk, Summerhill Lane, Portlaoise	Laois	NaN	185000.0	No	Yes	New Dwelling house /Apartment	greater than or equal to 38 sq metres and
2	2010-01-04	1 Meadow Avenue, Dundrum, Dublin 14	Dublin	NaN	438500.0	No	No	Second-Hand Dwelling house /Apartment	
3	2010-01-04	1 The Haven, Mornington	Meath	NaN	400000.0	No	No	Second-Hand Dwelling house /Apartment	
4	2010-01-04	11 Melville Heights, Kilkenny	Kilkenny	NaN	160000.0	No	No	Second-Hand Dwelling house /Apartment	

```
In [14]: # Check type of each feature
df.dtypes
```

```
Out[14]: DateofSale            datetime64[ns]
Address             object
County              object
Eircode             object
Price               float64
NotFullMarketPrice    object
VATExclusive        object
DescriptionofProperty   object
PropertySizeDescription  object
dtype: object
```

- Look for duplicate rows and columns. Consider whether it makes sense to keep them or drop them.

```
# Print the number of duplicates, without the original rows that were duplicated
print('Number of duplicate (excluding first) rows in the table is: ', df.duplicated().sum())

# Check for duplicate rows.
# Use "keep=False" to mark all duplicates as true, including the original rows that were duplicated.
print('Number of duplicate rows (including first) in the table is:', df[df.duplicated(keep=False)].shape[0])
```

Number of duplicate (excluding first) rows in the table is: 850
 Number of duplicate rows (including first) in the table is: 1597

- There appears to be 850 duplicated rows. We will check duplicated data below and make sure this is correct and we can drop duplicated rows.

```
In [16]: # show rows that have been duplicated and sort by address
df[df.duplicated(keep=False)].sort_values(by=['Address'])
```

		DateofSale	Address	County	Eircode	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Pr
246466	2017-03-16	1 AN BRUACH, BALLYDESMOND, MALLOW		Cork	NaN	96500.0		Yes	No	Second-Hand Dwelling house /Apartment
246467	2017-03-16	1 AN BRUACH, BALLYDESMOND, MALLOW		Cork	NaN	96500.0		Yes	No	Second-Hand Dwelling house /Apartment
418518	2020-03-18	1 Belcamp Manor, Malahide Road, Balgriffin		Dublin	NaN	337500.0		Yes	Yes	New Dwelling house /Apartment
418519	2020-03-18	1 Belcamp Manor, Malahide Road, Balgriffin		Dublin	NaN	337500.0		Yes	Yes	New Dwelling house /Apartment
442381	2020-10-15	1 COURT DEVENISH, ATHLONE, WESTMEATH		Westmeath	N37YC98	134000.0		No	No	Second-Hand Dwelling house /Apartment
...
2064	2010-02-18	Wyanstown, Oldtown		Dublin	NaN	6350.0		No	No	Second-Hand Dwelling house /Apartment
40242	2012-01-23	caherduggan, leamlara		Cork	NaN	300000.0		No	No	Second-Hand Dwelling house /Apartment
40244	2012-01-23	caherduggan, leamlara		Cork	NaN	300000.0		No	No	Second-Hand Dwelling house /Apartment
10496	2010-07-16	drumboy, mohill		Leitrim	NaN	46000.0		No	No	Second-Hand Dwelling house /Apartment
10495	2010-07-16	drumboy, mohill		Leitrim	NaN	46000.0		No	No	Second-Hand Dwelling house /Apartment

1597 rows × 9 columns

- Clearly data has been entered more than once in error since it is highly unlikely that a property has been sold more than once in a single day. We can drop these duplicates.

```
In [17]: # drop duplicate rows
df = df.drop_duplicates()
```

```
In [18]: #Print the number of duplicates, without the original rows that were duplicated
print('Number of duplicate (excluding first) rows in the table is: ', df.duplicated().sum())

# Check for duplicate rows.
# Use "keep=False" to mark all duplicates as true, including the original rows that were duplicated.
print('Number of duplicate rows (including first) in the table is:', df[df.duplicated(keep=False)].shape[0])
```

Number of duplicate (excluding first) rows in the table is: 0
 Number of duplicate rows (including first) in the table is: 0

```
In [19]: # show new number of rows and columns after dropping duplicates
df.shape
```

Out[19]: (595678, 9)

```
In [20]: # check cardinality of each feature
df.unique()
```

```
Out[20]: DateofSale      4605
Address          542347
County           26
Eircode          89876
Price            35166
NotFullMarketPrice    2
VATExclusive      2
DescriptionofProperty 5
PropertySizeDescription 6
dtype: int64
```

- Two columns, 'NotFullMarketPrice' and 'VATExclusive', appear to have the same cardinality and so it is possible that these could be a copy of each other. I will check whether these are duplicate columns.

```
In [21]: # create a series with a boolean value stating whether each row in the NotFullMarketPrice and VATExclusive column
# and sum the results to give a number which indicates the amount of duplicate rows
(df['NotFullMarketPrice'] == df['VATExclusive']).sum()
```

```
Out[21]: 479604
```

```
In [22]: #check number of empty rows per feature
df.isnull().sum()
```

```
Out[22]: DateofSale      0
Address          0
County           0
Eircode          504162
Price            0
NotFullMarketPrice    0
VATExclusive      0
DescriptionofProperty 0
PropertySizeDescription 543099
dtype: int64
```

- 'NotFullMarketPrice' and 'VATExclusive' have 479,604 rows that have duplicate information. This means they are not duplicate columns since they both have 595,678 non-empty rows.

- Look for constant columns. Consider whether it makes sense to keep them or drop them.

- As seen above, each feature has a cardinality greater than 1. This means there are no constant columns.

- Logic check

```
In [23]: # check first 5 rows
df.head(5)
```

	DateofSale	Address	County	Eircode	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescri
0	2010-01-01	5 Braemor Drive, Churchtown, Co.Dublin	Dublin	NaN	343000.0	No	No	Second-Hand Dwelling house /Apartment	
1	2010-01-03	134 Ashewood Walk, Summerhill Lane, Portlaoise	Laois	NaN	185000.0	No	Yes	New Dwelling house /Apartment	greater than or equal 38 sq metres and
2	2010-01-04	1 Meadow Avenue, Dundrum, Dublin 14	Dublin	NaN	438500.0	No	No	Second-Hand Dwelling house /Apartment	
3	2010-01-04	1 The Haven, Mornington	Meath	NaN	400000.0	No	No	Second-Hand Dwelling house /Apartment	
4	2010-01-04	11 Melville Heights, Kilkenny	Kilkenny	NaN	160000.0	No	No	Second-Hand Dwelling house /Apartment	

- The first logical check I will make is that the addresses are within the range of 1/1/2010 until 5/5/2023.

```
In [24]: # check first date of data entered
pd.DatetimeIndex(df['DateofSale']).min()
```

```
Out[24]: Timestamp('2010-01-01 00:00:00')
```

```
In [25]: # check last date of data entered
pd.DatetimeIndex(df['DateofSale']).max()
```

```
Out[25]: Timestamp('2023-05-05 00:00:00')
```

```
In [26]: # Double check row with latest date.
df[df['DateofSale'] == pd.DatetimeIndex(df['DateofSale']).max()]
```

	DateofSale	Address	County	Eircode	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty
596485	2023-05-05	1 GRANGE VIEW DR, CLONDALKIN, DUBLIN 22	Dublin	D22TE24	300000.00	No	No	Second-Hand Dwelling house /Apartment
596486	2023-05-05	1 OAKS AVE, CARRAIG AN AIRD, SIX CROSS ROADS	Waterford	X91HEE7	249000.00	No	No	Second-Hand Dwelling house /Apartment
596487	2023-05-05	10 BARN ELMS, UPPER CHURCHTOWN ROAD, CHURCHTOWN	Dublin	NaN	995000.00	No	Yes	New Dwelling house /Apartment
596488	2023-05-05	114 SCEILG ARD, HEADFORD RD, GALWAY	Galway	H91CP7X	295000.00	No	No	Second-Hand Dwelling house /Apartment
596489	2023-05-05	147 Station Walk, Rickardstown, Newbridge	Kildare	NaN	352423.00	No	Yes	New Dwelling house /Apartment
596490	2023-05-05	15 LAVEY MANOR DUBLIN RD, CHARLESTOWN, MAYO	Mayo	F12TK38	165000.00	No	No	Second-Hand Dwelling house /Apartment
596491	2023-05-05	15 Rinn na Mara, Kinvara	Galway	H91A9WH	365638.76	No	Yes	New Dwelling house /Apartment
596492	2023-05-05	154 MEADOW GROVE, DUNDRUM, DUBLIN 16	Dublin	D16YW59	645000.00	No	No	Second-Hand Dwelling house /Apartment
596493	2023-05-05	16 FAIRYFIELD, PARTEEN, CLARE	Clare	V94A9N9	385000.00	No	No	Second-Hand Dwelling house /Apartment
596494	2023-05-05	18 REDINGTON WOODS, CLARINBRIDGE, GALWAY	Galway	H91W29T	770000.00	No	No	Second-Hand Dwelling house /Apartment
596495	2023-05-05	2 MOORE STREET, CAPPAMORE, COUNTY LIMERICK	Limerick	NaN	20000.00	No	No	Second-Hand Dwelling house /Apartment
596496	2023-05-05	20 PINE GROVE, ATHLUMNEY WOOD, NAVAN	Meath	C15FA3F	240000.00	No	No	Second-Hand Dwelling house /Apartment
596497	2023-05-05	22 ARD NA GREINE, CHURCH RD, TULLAMORE	Offaly	R35XE95	320000.00	No	No	Second-Hand Dwelling house /Apartment
596498	2023-05-05	23 LLEWELLYN CLOSE, GRANGE VALLEY, RATHFARNHAM	Dublin	D16VX74	460000.00	No	No	Second-Hand Dwelling house /Apartment
596499	2023-05-05	24 WOODFORD WAY, CLONDALKIN, DUBLIN 22	Dublin	D22RK59	352000.00	No	No	Second-Hand Dwelling house /Apartment
596500	2023-05-05	25 THE GREEN, NORBURY WOODS, TULLAMOER	Offaly	R35A2F6	220000.00	No	No	Second-Hand Dwelling house /Apartment
596501	2023-05-05	27 EDMUND RICE PARK, CALLAN, KILKENNY	Kilkenny	R95W3Y6	175000.00	No	No	Second-Hand Dwelling house /Apartment
596502	2023-05-05	275 SUNDAYS WELL, NAAS, KILDARE	Kildare	W91T2HV	449000.00	No	No	Second-Hand Dwelling house /Apartment
596503	2023-05-05	3 HAROLDVILLE AVE, CIRCULAR RD SOUTH, RIALTO D...	Dublin	D08F9W8	625000.00	No	No	Second-Hand Dwelling house /Apartment

- There appears to be no problem with the dates entered.

```
In [27]: # check cardinality of each feature
df.unique()
```

```
Out[27]: DateofSale      4605
Address        542347
County          26
Eircode        89876
Price          35166
NotFullMarketPrice    2
VATExclusive     2
DescriptionofProperty 5
PropertySizeDescription 6
dtype: int64
```

```
In [28]: # change address values to upper case for readability, easier string identification and consistent formatting
df['Address'] = df['Address'].str.upper()
df.head(5)
```

	DateofSale	Address	County	Eircode	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescription
0	2010-01-01	5 BRAEMOR DRIVE, CHURCHTOWN, CO.DUBLIN	Dublin	NaN	343000.0	No	No	Second-Hand Dwelling house /Apartment	
1	2010-01-03	134 ASHEWOOD WALK, SUMMERHILL LANE, PORTLAOISE	Laois	NaN	185000.0	No	Yes	New Dwelling house /Apartment	greater than 38 sq metres
2	2010-01-04	1 MEADOW AVENUE, DUNDRUM, DUBLIN 14	Dublin	NaN	438500.0	No	No	Second-Hand Dwelling house /Apartment	
3	2010-01-04	1 THE HAVEN, MORNINGTON	Meath	NaN	400000.0	No	No	Second-Hand Dwelling house /Apartment	
4	2010-01-04	11 MELVILLE HEIGHTS, KILKENNY	Kilkenny	NaN	160000.0	No	No	Second-Hand Dwelling house /Apartment	

- I will now check the unique values of other features with low cardinality and make sure they are logical.

```
In [29]: # check unique values for County feature
county_array = []
for i in df['County'].unique():
    county_array += [i]
print(county_array)
```

```
['Dublin', 'Laois', 'Meath', 'Kilkenny', 'Limerick', 'Carlow', 'Cork', 'Clare', 'Sligo', 'Cavan', 'Tipperary', 'Wicklow', 'Roscommon', 'Wexford', 'Mayo', 'Donegal', 'Longford', 'Galway', 'Offaly', 'Kildare', 'Waterford', 'Louth', 'Kerry', 'Westmeath', 'Monaghan', 'Leitrim']
```

- County values are entered correctly, with all 26 counties in the Republic of Ireland represented.

```
In [30]: # check unique values for NotFullMarketPrice feature
df['NotFullMarketPrice'].unique()
```

```
Out[30]: array(['No', 'Yes'], dtype=object)
```

```
In [31]: # check unique values for VATExclusive feature
df['VATExclusive'].unique()
```

```
Out[31]: array(['No', 'Yes'], dtype=object)
```

```
In [32]: # check unique values for DescriptionofProperty feature
df['DescriptionofProperty'].unique()
```

```
Out[32]: array(['Second-Hand Dwelling house /Apartment',
 'New Dwelling house /Apartment', 'Teach/Árasán Cónaithe Atháimhe',
 'Teach/Árasán Cónaithe Nua', 'Teach/?ras?n C?naithe Nua'],
 dtype=object)
```

- There are three categories in Irish for this feature, I will change them to the English translation which are the same as 'Second-Hand Dwelling house /Apartment' and 'New Dwelling house /Apartment'.

```
In [33]: # Change value to english translation
df['DescriptionofProperty'] = df['DescriptionofProperty'].replace('Teach/Árasán Cónaithé Atháimhe',
                                                               'Second-Hand Dwelling house /Apartment', regex=True)
df['DescriptionofProperty'] = df['DescriptionofProperty'].replace('Teach/Árasán Cónaithé Nua',
                                                               'New Dwelling house /Apartment', regex=True)
```

```
In [34]: # show property descriptions that are still in Irish
df[(df['DescriptionofProperty'] != 'Second-Hand Dwelling house /Apartment') & (df['DescriptionofProperty'] != 'New Dwelling house /Apartment')]
```

```
Out[34]:
```

	DateofSale	Address	County	Eircode	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescription
39803	2012-01-12	APARTMENT 12 BLOCK B, COROFIN HOUSE CLARE VI...	Dublin	NaN	115045.0	No	Yes	Teach/?ras?n C?naithe Nua	n?os l? n?

```
In [35]: # change one row to English description
df.at[39803, 'DescriptionofProperty'] = 'New Dwelling house /Apartment'
```

```
In [36]: # check cardinality of feature
df['DescriptionofProperty'].nunique()
```

```
Out[36]: 2
```

```
In [37]: # check unique values for PropertySizeDescription feature
df['PropertySizeDescription'].unique()
```

```
Out[37]: array([nan,
                'greater than or equal to 38 sq metres and less than 125 sq metres',
                'greater than 125 sq metres', 'less than 38 sq metres',
                'greater than or equal to 125 sq metres',
                'níos mó ná nó cothrom le 38 méadar cearnach agus níos lú ná 125 méadar cearnach',
                'n?os l? n? 38 m?adar cearnach'], dtype=object)
```

- Two values, 'greater than 125 sq metres' and 'greater than or equal to 125 sq metres', should fall under the same category so I will combine the two. 'níos mó ná nó cothrom le 38 méadar cearnach agus níos lú ná 125 méadar cearnach' and 'n?os l? n? 38 m?adar cearnach' should also be translated to English.

```
In [38]: # replace value to reduce cardinality
df['PropertySizeDescription'] = df['PropertySizeDescription'].replace('greater than 125 sq metres',
                                                               'greater than or equal to 125 sq metres', regex=True)
```

```
In [39]: # replace to English translations
df['PropertySizeDescription'] = df['PropertySizeDescription'].replace('níos mó ná nó cothrom le 38 méadar cearnach',
                                                               'greater than or equal to 38 sq metres and less than 125 sq metres', regex=True)
df['PropertySizeDescription'] = df['PropertySizeDescription'].replace('n?os l? n? 38 m?adar cearnach',
                                                               'less than 38 sq metres', regex=True)
df['PropertySizeDescription'].unique()
```

```
Out[39]: array([nan,
                'greater than or equal to 38 sq metres and less than 125 sq metres',
                'greater than or equal to 125 sq metres', 'less than 38 sq metres',
                'n?os l? n? 38 m?adar cearnach'], dtype=object)
```

```
In [40]: # show property descriptions that are still in Irish
df[(df['PropertySizeDescription'] != 'greater than or equal to 38 sq metres and less than 125 sq metres') &
   (df['PropertySizeDescription'] != 'less than 38 sq metres') & (df['PropertySizeDescription'].notnull())]
```

```
Out[40]:
```

	DateofSale	Address	County	Eircode	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescription
39803	2012-01-12	APARTMENT 12 BLOCK B, COROFIN HOUSE CLARE VI...	Dublin	NaN	115045.0	No	Yes	New Dwelling house /Apartment	n?os l? n?

```
In [41]: # change one row to English description
df.at[39803, 'PropertySizeDescription'] = 'less than 38 sq metres'
```

```
In [42]: # check cardinality of feature
df['PropertySizeDescription'].nunique()
```

```
Out[42]: 3
```

```
In [43]: df['PropertySizeDescription'].unique()
```

```
Out[43]: array([nan,  
   'greater than or equal to 38 sq metres and less than 125 sq metres',  
   'greater than or equal to 125 sq metres', 'less than 38 sq metres'],  
  dtype=object)
```

```
In [44]: # check types  
df.dtypes
```

```
Out[44]: DateofSale          datetime64[ns]  
Address            object  
County             object  
Eircode            object  
Price              float64  
NotFullMarketPrice    object  
VATExclusive        object  
DescriptionofProperty    object  
PropertySizeDescription    object  
dtype: object
```

- Both EirCode and PropertySizeDescription still have a very high amount of empty rows. I will address this in the data quality plan.

- Save your updated/cleaned data frame to a new csv file.

```
In [45]: # saving cleaned dataframe to a csv  
df.to_csv("PPR-ALL_cleaned_section1.csv")
```

```
In [46]: # convert all object data into categorical columns, except for address column  
object_columns = list(df.select_dtypes(['object']).columns)  
object_columns.remove('Address')  
  
for column in object_columns:  
    df[column] = df[column].astype('category')  
df.dtypes
```

```
Out[46]: DateofSale          datetime64[ns]  
Address            object  
County             category  
Eircode            category  
Price              float64  
NotFullMarketPrice    category  
VATExclusive        category  
DescriptionofProperty    category  
PropertySizeDescription    category  
dtype: object
```

- Prepare a table with descriptive statistics for all the continuous features.

```
In [47]: # I will need separate tables for the numeric and date columns as the nature of the descriptive statistics will  
# I will start with the numeric columns  
numeric_columns = df.select_dtypes(['float64', 'int64']).columns  
numeric_columns
```

```
Out[47]: Index(['Price'], dtype='object')
```

```
In [48]: df[numeric_columns].describe().T
```

```
Out[48]:
```

	count	mean	std	min	25%	50%	75%	max
Price	595678.0	281362.668497	1.013935e+06	5001.0	126000.0	213656.39	322000.0	225000000.0

```
In [49]: # Add cardinality column  
numeric_columns_card = df[numeric_columns].nunique()  
  
# store the values in a dataframe  
df_numeric_card = pd.DataFrame(numeric_columns_card, columns=['card'])  
  
df_numeric_card
```

```
Out[49]:
```

card
Price 35166

```
In [50]: # I will now create the table with descriptive statistics for the datetime columns  
date_columns = df.select_dtypes(['datetime64[ns]']).columns  
date_columns
```

```
Out[50]: Index(['DateofSale'], dtype='object')
```

```
In [51]: df[date_columns].describe().T
```

```
Out[51]:
```

	count	mean	min	25%	50%	75%	max
DateofSale	595678	2017-11-09 00:34:28.338934784	2010-01-01	2015-03-31	2018-02-16	2020-11-12	2023-05-05

```
In [52]: # Add cardinality column  
date_columns_card = df[date_columns].nunique()
```

```
# store the values in a dataframe  
df_date_card = pd.DataFrame(date_columns_card, columns=['card'])  
  
df_date_card
```

```
Out[52]:
```

card	
DateofSale	4605

```
In [53]: # Prepare %missing column  
numeric_columns_missing = 100 * (df[numeric_columns].isnull().sum()/df.shape[0])  
date_columns_missing = 100 * (df[date_columns].isnull().sum()/df.shape[0])  
  
# store the values in a dataframe  
df_numeric_missing = pd.DataFrame(numeric_columns_missing, columns=['%missing'])  
df_date_missing = pd.DataFrame(date_columns_missing, columns=['%missing'])
```

```
In [54]: # Add missing values column  
df_table_numeric = df[numeric_columns].describe().T  
  
# Put the columns together to prepare the final table for numeric_columns  
df_numeric_columns_data_quality_report_table = pd.concat([df_table_numeric, df_numeric_missing, df_numeric_card])  
  
df_numeric_columns_data_quality_report_table
```

```
Out[54]:
```

	count	mean	std	min	25%	50%	75%	max	%missing	card
Price	595678.0	281362.668497	1.013935e+06	5001.0	126000.0	213656.39	322000.0	225000000.0	0.0	35166

```
In [55]: # Add missing values column  
df_table_date = df[date_columns].describe().T  
  
# Put the columns together to prepare the final table for numeric_columns  
df_date_columns_data_quality_report_table = pd.concat([df_table_date, df_date_missing, df_date_card], axis=1)  
  
df_date_columns_data_quality_report_table
```

```
Out[55]:
```

	count	mean	min	25%	50%	75%	max	%missing	card
DateofSale	595678	2017-11-09 00:34:28.338934784	2010-01-01	2015-03-31	2018-02-16	2020-11-12	2023-05-05	0.0	4605

- Prepare a table with descriptive statistics for all the categorical features.

```
In [56]: category_columns = df.select_dtypes(['category']).columns  
category_columns
```

```
Out[56]: Index(['County', 'Eircode', 'NotFullMarketPrice', 'VATExclusive',  
               'DescriptionofProperty', 'PropertySizeDescription'],  
               dtype='object')
```

```
In [57]: # Add cardinality column  
# A simpler way to get the number of unique values per feature  
category_columns_card = df[category_columns].nunique()  
# store the values in a dataframe  
df_category_card = pd.DataFrame(category_columns_card, columns=['card'])  
df_category_card
```

```
Out[57]:
```

	card
County	26
Eircode	89876
NotFullMarketPrice	2
VATExclusive	2
DescriptionofProperty	2
PropertySizeDescription	3

```
In [58]: # Prepare %missing column
category_columns_missing = 100 * (df[category_columns].isnull().sum()/df.shape[0])

# store the values in a dataframe
df_category_missing = pd.DataFrame(category_columns_missing, columns=['%missing'])
df_category_missing
```

```
Out[58]: %missing
```

	%missing
County	0.000000
Eircode	84.636666
NotFullMarketPrice	0.000000
VATExclusive	0.000000
DescriptionofProperty	0.000000
PropertySizeDescription	91.173251

```
In [59]: df[category_columns].describe().T
```

```
Out[59]: count unique top freq
```

	count	unique	top	freq
County	595678	26	Dublin	187661
Eircode	91516	89876	A96WV79	8
NotFullMarketPrice	595678	2	No	567023
VATExclusive	595678	2	No	499505
DescriptionofProperty	595678	2	Second-Hand Dwelling house /Apartment	497608
PropertySizeDescription	52579	3	greater than or equal to 38 sq metres and less...	37891

```
In [60]: # Add missing values column
df_table_category = df[category_columns].describe().T

# Put the columns together to prepare the final table for numeric_columns
df_category_columns_data_quality_report_table = pd.concat([df_table_category, df_category_missing, df_category_]

df_category_columns_data_quality_report_table
```

```
Out[60]: count unique top freq %missing card
```

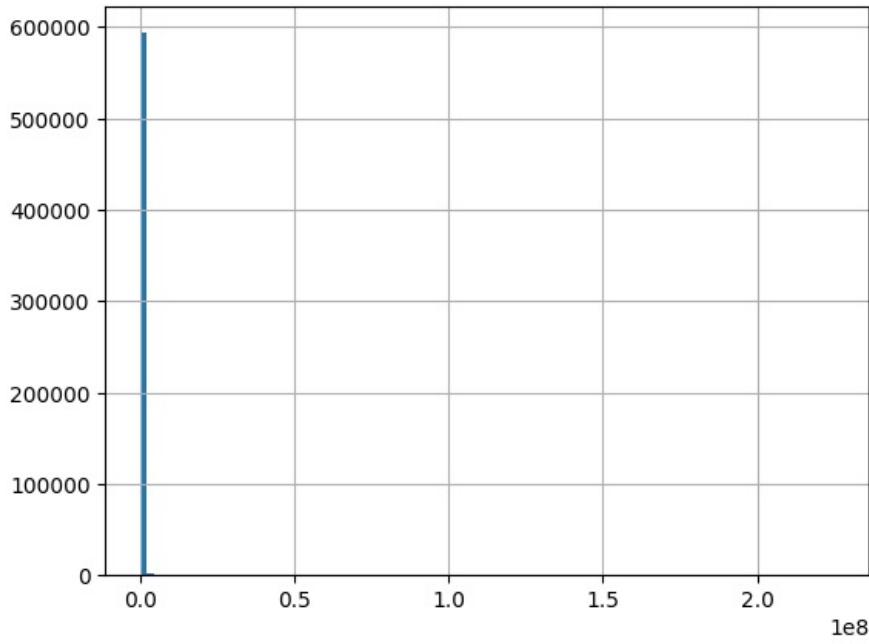
	count	unique	top	freq	%missing	card
County	595678	26	Dublin	187661	0.000000	26
Eircode	91516	89876	A96WV79	8	84.636666	89876
NotFullMarketPrice	595678	2	No	567023	0.000000	2
VATExclusive	595678	2	No	499505	0.000000	2
DescriptionofProperty	595678	2	Second-Hand Dwelling house /Apartment	497608	0.000000	2
PropertySizeDescription	52579	3	greater than or equal to 38 sq metres and less...	37891	91.173251	3

- Plot histograms for all the continuous features.

- I will start by creating the histograms the numeric and datetime features.

```
In [61]: # Plot histogram for Price
df['Price'].hist(bins=100)
```

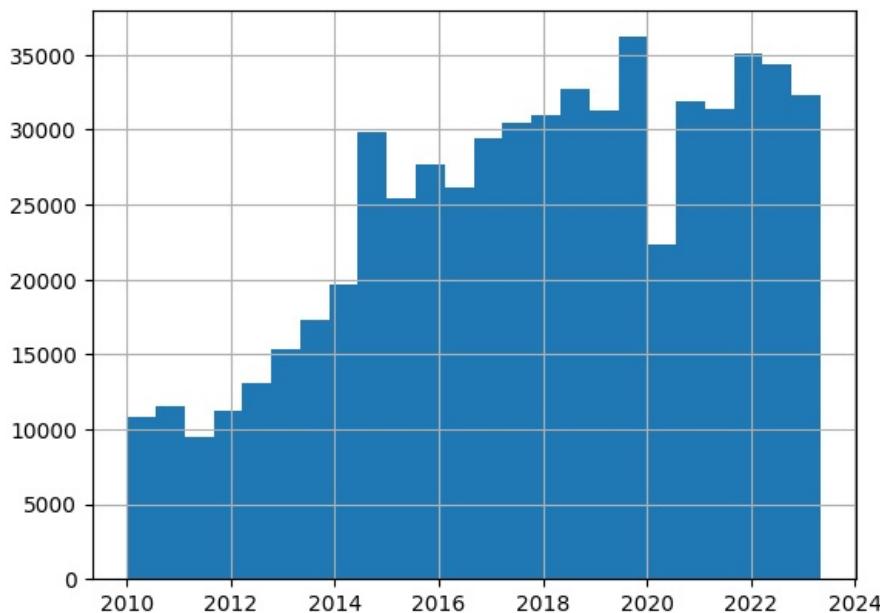
```
Out[61]: <Axes: >
```



- The histogram for Price shows that most of the data is below €50,000,000. So much so that only one bin is visible which does not tell us much about the distribution of the data. This could suggest that there are a number of outliers spread out across a range of roughly €50,000,000 - €250,000,000.

```
In [62]: #Plot histogram for date of sale  
df['DateofSale'].hist(bins=24)
```

```
Out[62]: <Axes: >
```

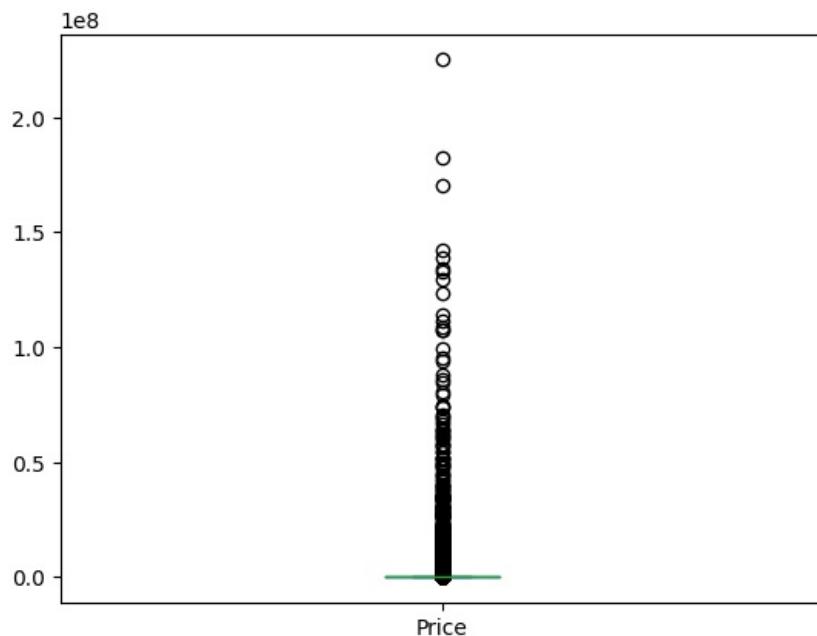


- The DateofSale histogram shows most sales occurred at the end of 2019, while the number of sales generally trended upwards from 2010 to 2023.

- Plot box plots for all the continuous features.

```
In [63]: # Plot boxplot for Price
df['Price'].plot(kind='box')
```

```
Out[63]: <Axes: >
```

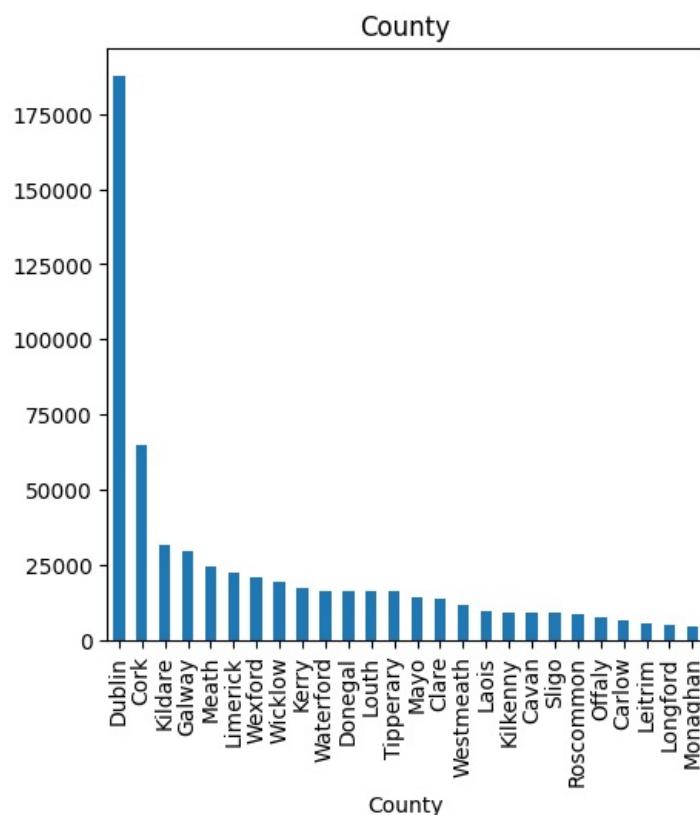


- The boxplot of Price shows a large number of outliers with a large price range.

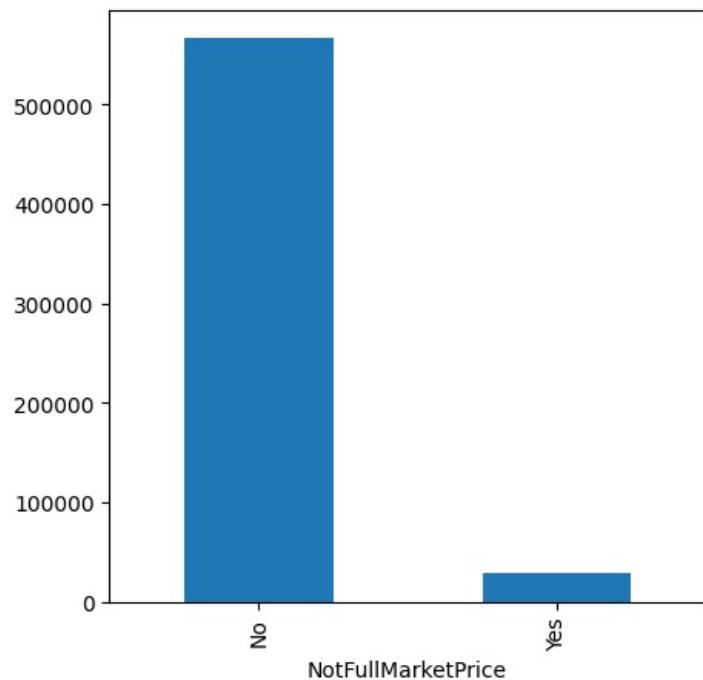
- Plot bar plots for all the categorical features.

```
In [64]: categorical_columns = list(df.select_dtypes(['category']).columns)
categorical_columns.remove('Eircode')

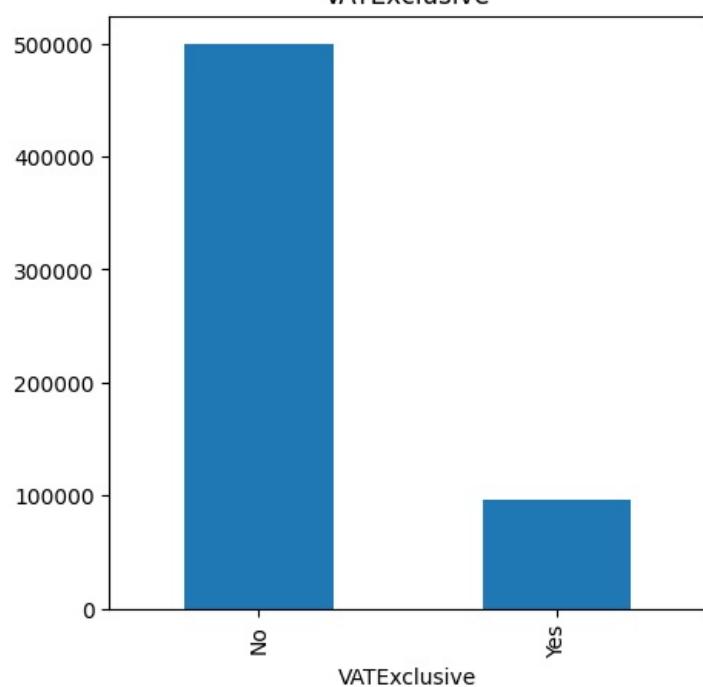
# loop over categorical columns and print a bar chart for each
for column in categorical_columns:
    f = df[column].value_counts(dropna=False).plot(kind='bar', title=column, figsize=(5,5))
    plt.show()
```

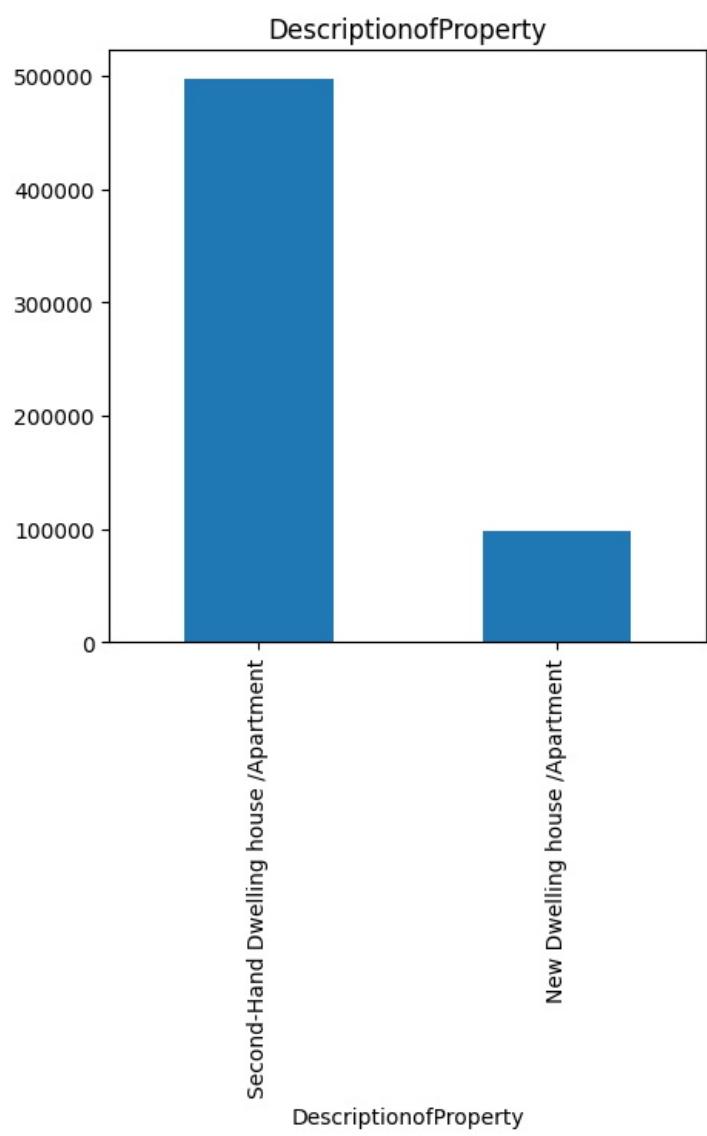


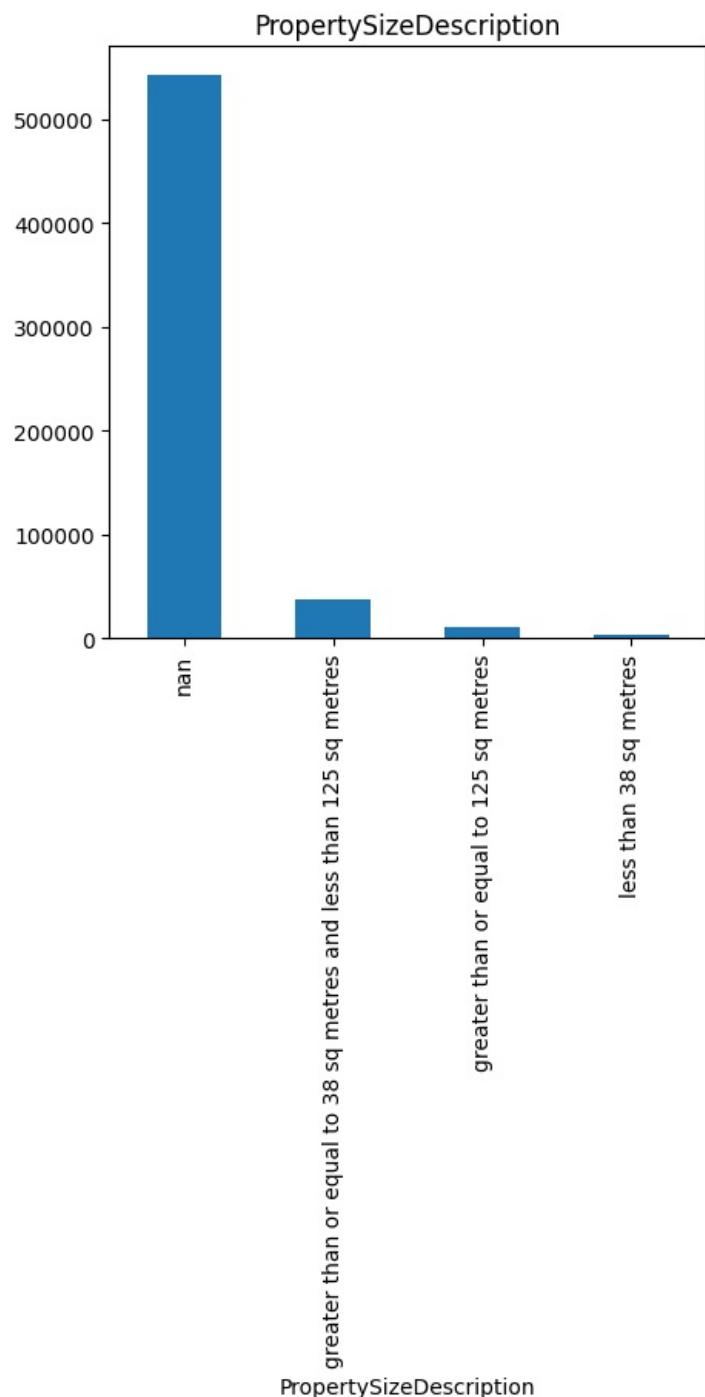
NotFullMarketPrice



VATExclusive







- The significant bar charts above are the following:

County

- Here we can clearly see that the most common addresses are in Dublin, with the next most common is Cork. These are the two most populous counties in Ireland and it would make sense for them to have the most properties sold, although it is interesting that the number of Dublin sales are over double the amount of Cork sales.

DescriptionofProperty

- It appears over 80% of properties sold are a Second-Hand Dwelling house/appartment. This would suggest that the number of new homes being built is quite low compared to available housing.

PropertySizeDescription

- The most common properties have a range of size between 38 sq metres and less than 125 sq metres. However, most of the data is missing from this feature, which will be addressed in the data quality report.

Detailed report of initial findings in accompanying file: *Data_Quality_Report_Initial_Findings.pdf*

2 - Prepare a data quality plan for the cleaned CSV file.

```
In [65]: df = pd.read_csv("PPR-ALL_cleaned_section1.csv", index_col=0)
df.shape
```

```
/var/folders/bn/6x11zkt6n14vfwyb0jymgt00000gp/T/ipykernel_31017/1354650328.py:1: DtypeWarning: Columns (4,9) have mixed types. Specify dtype option on import or set low_memory=False.
df = pd.read_csv("PPR-ALL_cleaned_section1.csv", index_col=0)
```

```
Out[65]: (595678, 9)
```

- Mark down all the features where there are potential problems or data quality issues.

As outlined in the accompanying pdf containing the data quality report the following problems have been identified with the data:

- Missing values in Eircode column
- Missing values in PropertySizeDescription column
- Outliers in Price data

- Propose solutions to deal with the problems identified. Explain why did you choose one solution over potentially many other. It is very important to provide justification for your thinking in this part and to list potential solutions, including the solution that will be implemented to clean the data.

1. Missing values in Eircode column

Upon first inspection, there does not appear to be Eircode data within Address data of some rows. I will inspect the Address data further and try to extract Eircode data from the Address column.

Initially I considered dropping this column as Eircode data is missing from over 84% of rows. I have decided not to drop it since it could provide useful information on property location such as latitude and longitude, however I will drop it if I am unable to extract a significant number of Eircodes.

2. Missing values in PropertySizeDescription column

This data is missing over 91% of data and could be considered reasonable to drop this column altogether. However, I believe the count of non-empty rows of PropertySizeDescription is close to the frequency of 'New Dwelling house /Apartment'. If this is true, it could contain useful information about newly built properties, and so I will investigate and see if there is a correlation between rows with 'New Dwelling house /Apartment' and non-empty PropertySizeDescription to see if any action needs to be taken.

3. Outliers in Price data

I will investigate the outliers and see if any action needs to be taken.

- Apply your solutions to obtain a new CSV file where the identified data quality issues were addressed.

1. Missing values in Eircode column

```
In [4]: # check amount of addresses with a description of the property size
df['Eircode'].isnull().sum()
```

```
Out[4]: 504162
```

```
In [5]: # replace NaN values as 'other'
df['Eircode'] = df['Eircode'].astype('object')
df.Eircode = df.Eircode.fillna('other')

df.Eircode.isnull().sum()
```

```
Out[5]: 0
```

- Here I will use the Address column to look for Eircodes to fill in missing Eircode data. Eircodes are quite structured, so there

shouldn't be a need to normalise the address before extraction. There are 139 eircode routing keys, so we will look for one of these within each address.

```
In [6]: # list of eircode routing keys found at
# https://www.autoaddress.ie/blog/autoaddressblog/2016/09/21/eircode-routing-keys
eircode_routing_keys = [
    'a41', 'a42', 'a45', 'a63', 'a67', 'a75', 'a81', 'a82', 'a83', 'a84', 'a85', 'a86', 'a91', 'a92', 'a94',
    'a98', 'c15', 'd01', 'd02', 'd03', 'd04', 'd05', 'd06', 'd6w', 'd07', 'd08', 'd09', 'd10', 'd11', 'd12',
    'd14', 'd15', 'd16', 'd17', 'd18', 'd20', 'd22', 'd24', 'e21', 'e25', 'e32', 'e34', 'e41', 'e45', 'e53',
    'f12', 'f23', 'f26', 'f28', 'f31', 'f35', 'f42', 'f45', 'f52', 'f56', 'f91', 'f92', 'f93', 'f94', 'h12',
    'h16', 'h18', 'h23', 'h53', 'h54', 'h62', 'h65', 'h71', 'h91', 'k32', 'k34', 'k36', 'k45', 'k56', 'k67',
    'n37', 'n39', 'n41', 'n91', 'p12', 'p14', 'p17', 'p24', 'p25', 'p31', 'p32', 'p36', 'p43', 'p47', 'p51',
    'p61', 'p67', 'p72', 'p75', 'p81', 'p85', 'r14', 'r21', 'r32', 'r35', 'r42', 'r45', 'r51', 'r56', 'r93',
    't12', 't23', 't34', 't45', 't56', 'v14', 'v15', 'v23', 'v31', 'v35', 'v42', 'v92', 'v93', 'v94', 'v95',
    'w23', 'w34', 'w91', 'x35', 'x42', 'x91', 'y14', 'y21', 'y25', 'y34', 'y35'
]
```

```
In [7]: addressRow = []

# for each address, check if one of the routing keys are present
for row in range(0, len(df)):
    address = df['Address'][row]
    for key in eircode_routing_keys:
        if key.upper() in address:
            addressRow += [row]

len(addressRow)
```

```
Out[7]: 101
```

- We can see that there are 101 rows with Eircode routing keys present in the address data. We'll examine these rows below.

```
In [8]: pd.set_option('display.max_rows', len(df))

# show rows with Eircode key in Address and no Eircode listed
df.iloc[addressRow][df['Eircode'] == 'other'][['Address']]
```

```
/var/folders/bn/6x11zkt6n14vfwyb0jymgt00000gp/T/ipykernel_23318/2817090869.py:4: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
df.iloc[addressRow][df['Eircode'] == 'other'][['Address']]
```

```
Out[8]: 3046      APARTMENT 146 WYCKHAM POINT, WYCKHAM WAY, DUND...
5529      APARTMENT NO. A42/42A, DRUMBEAR WOOD, MONAGHAN
17925     25 OAK HOUSE, CARRICKMINES GREEN, GLENAMUCK C...
19496     25 CASTAHENY, CLONEE, D15
20158     34 LUTTRELSTOWN RISE, LUTTREYSTOWN GLEN, CASTLE...
22935     26 ORCHARD WAY GREENWOOD WALK, AYREFIELD MALAH...
23581     APT. C15 CITYGATE, SLIGO
32322     DOOR 16 (SITE34), THE BIRCHES LOUGH NA GLACK, ...
37038     19 ALLENTON DRIVE, BALLYCRAGH, TALLAGHT D24
47840     APT 603 VANTAGE EAST, CENTRAL PARK, LEOPARDSTO...
56932     14 DELHURST TERRACE ONGAR D15, ONGAR
58359     47 KILCARRIG CRESCENT, FETTERCAIRN, TALLAGHT D24
58906     25 LATCHFORD PARK CLONEE D15, D15
81794     APT12 CHURCHFIELDS, ASHBOURNE
82603     THE STUDIO, WOODTOWN, RATHFARNHAM D16
95888     37 THE COURTYARD, COLLEGE DRIVE FORTFIELD, S...
106072    H160 BRU NA SIONNA, SHANNON, CLARE
106073    H161 BRU NA SIONNA, SHANNON, CLARE
106074    H162 BRU NA SIONNA, SHANNON, CLARE
106075    H163 BRU NA SIONNA, SHANNON, CLARE
106076    H164 BRU NA SIONNA, SHANNON, CLARE
112195    H142 BRU NA SIONNA, SHANNON, CLARE
114059    H141 BRU NA SIONNA, SHANNON, CLARE
118129    H140 BRU NA SIONNA, SHANNON, CLARE
135035    UNIT F12, SPANISH COVE, KILKEE
137357    D02 EDEN HALL, MODEL FARM RD, CARRIGROHANE
138040    MAIN STREET, CASTLEDERMOT, KE41963F
139554    H143 BRU NA SIONNA, SHANNON, CLARE
139908    345 ORWELL PARK, CLOSE, TEMPLEogue D6W
161879    2 SLANELOUGH, ASHFORD, CO WICKLOW A67 ED73
164884    PARSONS GATES, BELMONT ROAD, CASTLECONNELL V94...
168346    7 MONTPELLIER TERRACE, GALWAY, H91 C2DN
170456    21 HARBOUR VIEW, THE SQUARE, CLIFDEN H71 AH59
171861    APARTMENT 19 DESART COURT, NEW STREET, KILKENN...
172622    31 RATH GAILINE, DUBLIN ROAD, PORTLAOISE R32 W24X
174138    8 HARBOUR VIEW, THE SQUARE, CLIFDEN H71 X957
174365    56 COLTHURST CRESCENT, LUCAN, COUNTY DUBLIN K7...
175063    3 HARBOUR VIEW, MARKET SQUARE, CLIFDEN H71 AW98
181114    107 SAUNDERS LANE, RATHNEW, WICKLOW A67 NY01
181834    7 NEW STREET, KILLALOE, V94 KN8C
```

182120	19 OLDTOWN GARDENS, NAAS, W91 EKR9
183632	321 CHALESLAND PK, GREYSTONES, A63 HK31
184384	12 LONGLANDS, SWORDS, K67 KR86
186763	98 LANDSCAPE PK, CHURCHTOWN, DUBLIN D14 X654
186773	APTMENT 22 NUTGROVE MEWS, MARYBOROUGH WOODS,...
188977	12 AN TSEAN SCOIL, CLIFDEN, H71 VX59
189680	76 ROSEHILL, WICKLOW, A67 HK19
192819	14 GREENHILLS, BROWNSHILL ROAD, CARLOW R93 X3T1
202972	7 TRAVERSLEA WOODS, LOWER GLENAGEARY ROAD, A96...
227026	APTMENT121, 109 PARNELL ST, DUBLIN 1
242291	APT12 ABBOT COURT CUALANOR, UPPER GLENAGEARY ...
242302	APT34 ABBOT COURT CUALANOR, UPPER GLENAGEARY ...
251372	APT34 OLD DISTILLERY, BERESFORD ST, DUBLIN 7
265503	H163 BRU NA SIONNA, SHANNON, CO CLARE
269757	37A SARSFIELD STREET, NENAGH, E45 VK33
270059	'MOUNT ALVERNO', 17 NEWCASTLE ROAD, GALWAY H91...
288723	APTMENT122, NORTHBANK, CASTLEFORBES ROAD UPP...
295438	MAIN STREET, BALLYDEHOB, P81 H504
300521	D10 EDEN HALL, MODEL FARM ROAD, CORK
319419	EDEN HALL, MODEL FARM RD, APT D07 STUDENT ACCOM
319420	EDEN HALL, MODEL FARM RD, APT D08 STUDENT ACCOM
328257	APT126 BLOCK 3, ISLAND KEY, EAST WALL DUBLIN 3
336205	NO 85 NEW BANCROFT HALL, TALLAGHT, D24
344168	C15 CAHERMORE HOLIDAY VILLAGE, ENNISCRONE, CO ...
344451	D02 EDEN HALL, MODEL FARM RD, CARRIGROHANE
346208	DASS 117 FOXROCK GROVE, DULBIN 18, D18 RD82
353919	70 ROANMORE PARK, WATERFORD, X91YN5K
356224	UNIT12 GRACE PARK ROW, GRACEPARK WOOD, GRACE P...
362279	11 BANK PLACE, TIPPERARY, E34AP80
373449	1 FITZWILLIAM SQUARE, DUBLIN 2, D02XY95
395469	UNIT12, SYCAMORE GROVE, GROVE ROAD
404389	APT W12, THE COBBLES, 28 WELLINGTON QUAY DUBLIN 2
411178	SKREEN, DONEGAL, F94P5V6
426178	1A JOHNS RD, WEXFORD, Y35RYW5
437147	CORRANEARY, AUGHAVAS, H12TD61
438194	127 DRAKES POINT, CROSSHAVEN, P43 NN67
439561	LYNN, MULLINGAR, N91 E9W6
441418	KILCOONEY, BALLINAMULT, E91 K797
447010	1 FAIRWAYS GARDENS, DUNDALK, A91 W5KD
449909	NO 2 BROOKFIELD, MULLINGAR, N91 V2PV
461287	APT23 BLOCK H THE CAMPION, MARINA VILLAGE, GRE...
463617	APT W12, THE COBBLES, 28 WELLINGTON QUAY DUBLIN 2
469807	BALLINAD01, LEITRIM PO, CARRICK ON SHANNON
503924	MOUNT SHANNON, CO CLARE, V94 K2CC
529662	89 BARNWELL PARK, HANSFIELD, D15
533471	87 BARNWELL PARK, HANSFIELD, D15
537937	96 BARNWELL PARK, ONGAR, D15
540291	APT YV42, YEATS VILLAGE, BALLINODE
542934	104 BARNWELL PARK, HANSFIELD, D15
564297	51 BARNWELL GROVE, HANSFIELD, D15
568695	29 MERCERS ROAD MERCER VALE, CHERRYWOOD, D18

Name: Address, dtype: object

- Some of the addresses that have the keys present are not actually used for the Eircode, meaning we will only be able to extract less than 101 new Eircodes.

```
In [9]: # properties with no Eircode and have a routing key in the address
no_code = df.iloc[addressRow][df['Eircode'] == 'other']['Address'].index

# extract Eircode if in address
for row in no_code:
    address = df['Address'][row].split(',')[-1].split(' ')
    for key in eircode_routing_keys:
        if (key.upper() in address[-1]) & (len(address[-1]) > 3):
            df.at[row, 'Eircode'] = address[-1]
        elif (key.upper() in address[-2]):
            df.at[row, 'Eircode'] = address[-2] + address[-1]

# new number of empty Eircodes
df[df['Eircode'] == 'other'].count()
```

```
/var/folders/bn/6x11zkt6n14vfwyb0jymgt00000gp/T/ipykernel_23318/1464443346.py:2: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
no_code = df.iloc[addressRow][df['Eircode'] == 'other']['Address'].index
```

```
Out[9]: DateofSale      504125
Address        504125
County         504125
Eircode        504125
Price          504125
NotFullMarketPrice 504125
VATExclusive   504125
DescriptionofProperty 504125
PropertySizeDescription 52554
dtype: int64
```

- We were only able to fill in 37 new Eircodes. This is an insignificant amount so I will drop this feature.

```
In [10]: df = df.drop(columns='Eircode')
df.shape
```

```
Out[10]: (595678, 8)
```

2. Missing values in PropertySizeDescription column

- Here I will check if rows with nonempty PropertySizeDescription data are related to rows that are new addresses.

```
In [11]: # check amount of addresses with a description of the property size
df['PropertySizeDescription'].notnull().sum()
```

```
Out[11]: 52579
```

```
In [12]: # check amount of new addresses vs second hand addresses
df['DescriptionofProperty'].value_counts()
```

```
Out[12]: DescriptionofProperty
Second-Hand Dwelling house /Apartment    497608
New Dwelling house /Apartment            98070
Name: count, dtype: int64
```

```
In [13]: # check total amount of rows with a description of the property size that are new properties
((df['PropertySizeDescription'].notnull()) & (df['DescriptionofProperty'] == 'New Dwelling house /Apartment')).sum()
```

```
Out[13]: 52579
```

- We can see that the number of rows with both a nonempty PropertySizeDescription and DescriptionofProperty value of 'New Dwelling house /Apartment' is equal to the same number of rows that have a nonempty PropertySizeDescription. This means that all information regarding property size is in relation to new houses.

```
In [14]: # check total amount of rows with a description of the property size that are new properties
new_size_missing_sum = ((df['PropertySizeDescription'].isnull()) & (df['DescriptionofProperty'] == 'New Dwelling house /Apartment')).sum()

# prepare %missing column
new_house_size_missing = 100 * (new_size_missing_sum/df[(df['DescriptionofProperty'] == 'New Dwelling house /Apartment')].count())

# assign 2 variables as the number of new addresses with empty and nonempty data with regards to the property size
empty = [new_house_size_missing]
nonempty = [100 - new_house_size_missing]

# store the values in a dataframe
df_new_house_size_missing = pd.DataFrame(empty, columns=['%Size Description missing from New Dwelling house'])
df_new_house_size_not_missing = pd.DataFrame(nonempty, columns=['%Size Description not missing from New Dwelling house'])

#concatenate the two dataframes
df_new_house_size = pd.concat([df_new_house_size_missing, df_new_house_size_not_missing], axis = 1)

#display percentage of new addresses with a size description vs no size description
df_new_house_size
```

	%Size Description missing from New Dwelling house	%Size Description not missing from New Dwelling house
0	46.386255	53.613745

Over 53% of new homes contain information about the size of the property. This is may be useful in the analysis of newly built homes and so I will not drop the PropertySizeDescription column. We will fill in empty values with 'other' to make programming easier.

```
In [15]: # replace NaN values as 'other'
df['PropertySizeDescription'] = df['PropertySizeDescription'].astype('object')
df.PropertySizeDescription = df.PropertySizeDescription.fillna('other')
df['PropertySizeDescription'] = df['PropertySizeDescription'].astype('category')
```

```
df.PropertySizeDescription.isnull().sum()
```

```
Out[15]: 0
```

3. Outliers in Price data

I will investigate the outliers and see if any action needs to be taken.

```
In [16]: # show Price statistics  
df['Price'].describe().apply(lambda x: format(x, 'f'))
```

```
Out[16]: count      595678.000000  
mean       281362.668497  
std        1013935.021061  
min        5001.000000  
25%     126000.000000  
50%     213656.390000  
75%     322000.000000  
max    225000000.000000  
Name: Price, dtype: object
```

```
In [17]: # interquartile range  
p_25 = np.percentile(df['Price'], 25)  
p_75 = np.percentile(df['Price'], 75)  
  
IQR = p_75 - p_25  
IQR
```

```
Out[17]: 196000.0
```

- The maximum price is almost 700 times the price at the 75th percentile, while the minimum price is roughly 25 times lower than the price at the 25th percentile. The biggest concern about the data is the gap between the max price and the price at the 75th percentile. The IQR shows us that the middle 50% of property prices is spread across €196,000, which is quite small relative to the maximum property price.
- There are a number of high outliers in Price data. The majority of the data is below €50,000,000 and a number of outliers above this range increase the amount of bins in the histogram, making it difficult to make any inference from the data. It appears to be a small number of outliers with a high range.
- I believe it is better to take a percentile approach for identifying outliers rather than using the IQR proximity rule, since there is such a large range between the maximum price and the 3rd quartile. I will begin by checking the 99th percentile of data.

```
In [18]: price_data = df['Price']  
  
# check the price threshold of the highest 99% of prices and the number of rows  
p_99 = np.percentile(price_data, 99)  
p_99_rows = (df['Price'] < p_99).sum()  
p_1_rows = df['Price'].size - (df['Price'] < p_99).sum()  
  
# print the results  
print("99% of data is <= " + str(p_99) + ', containing ' + str(p_99_rows) + ' rows')  
print("Top 1% percent of data is <= " + str(p_99) + ', containing ' + str(p_1_rows) + ' rows')  
  
99% of data is <= 1275000.0, containing 589669 rows  
Top 1% percent of data is <= 1275000.0, containing 6009 rows
```

```
In [19]: # check range of top 1%  
price_data.max() - p_99
```

```
Out[19]: 223725000.0
```

- The top 1% of prices are in the range €1,275,000 - €225,000,000. This is a range of €223,725,000 which is quite large. I will check how many of the 6009 rows in the top 1% are above €2,000,000.

```
In [20]: # check number of rows above €2,000,000  
(price_data >= 2000000).sum()
```

```
Out[20]: 2448
```

- This means 4561 rows are in the price range of €1,275,000 - €2,000,000. I will look at this data more closely.

```
In [21]: # show most expensive property  
df[df['Price'] == df['Price'].max()]
```

Out[21]:	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescri
	584494	O'DEVANEY GARDENS, ARBOUR HILL, DUBLIN 7	Dublin	225000000.0	Yes	Yes	New Dwelling house /Apartment	

- We can see from the Address column that some of these properties are apartment blocks, including the most expensive property O'Devaney Gardens which, after a quick Google, I found is an apartment complex. It makes sense to drop these rows since they are significant outliers in the data and it does not make sense to count multiple residences as single properties. I will remove the O'DEVANEY GARDENS property and recheck the data.

```
In [22]: # drop row with max price
df = df.drop(index=584494)
df = df.reset_index(drop=True)
```

- Having multiple numbers present in the first line of the address may be an indication that multiple apartments were included in the property sale. I will use the function below to identify this criteria and remove apartment blocks from the data.

```
In [23]: import re

# function to check how many numbers in a string
def num_there(st):
    words = re.split(r"[^a-zA-Z]", st)
    count = 0
    for c in words:
        if c.isdigit():
            count += 1
    return count
```

```
In [24]: #numbers of rows that are apartments or multiple units
rows = []
# identify multiple apartments/units in data
for row in range(0, len(df)):
    address = df['Address'][row].split(',') [0]

    # after a quick glance, the BLOCK no. is included in some addresses and may be mistaken as an apartment no.
    if (num_there(address) > 1) & ('BLOCK' not in address):
        rows += [row]
df.iloc[rows]
```

Out[24]:	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	Description
	190	APT 8 1 GRANGE LODGE AVENUE, CLONGRIFFIN	Dublin	1.776860e+05	No	Yes	New Dwel /
	338	22 4 HERBERTON STREET, RIALTO, DUBLIN 8	Dublin	1.820000e+05	No	Yes	New Dwel /
	562	SITE NO. 2 (KNOWN AS NO. 4), LOUGHCURRA, KINVARA	Galway	3.083700e+05	No	Yes	New Dwel /
	1217	60/61 THE ELMS, MOUNT MERRION AVENUE, BLACKROCK	Dublin	6.850000e+05	No	No	Second-Han house /
	1318	58-65 KNOCKROE GLEN, KNOCKROE, MONAGHAN	Monaghan	1.654800e+06	Yes	Yes	New Dwel /
	2205	APT. NO. 15 86 REUBEN STREET, RIALTO, DUBLIN 8	Dublin	2.249000e+05	No	Yes	New Dwel /
	2416	127 129 AND 131 AND PART 133, LOWER RATHMINES ...	Dublin	6.010195e+06	No	No	Second-Han house /
	2486	6 & 7 CATHERINE PLACE, LIMERICK	Limerick	1.391333e+06	No	Yes	New Dwel /
	3198	APT 6 29 JAMES' WALK, RIALTO, DUBLIN 8	Dublin	1.880000e+05	No	Yes	New Dwel /
	3340	APT 14 HERBERTON 86 REUBEN STREET, ST JAMES'S ...	Dublin	1.720000e+05	No	Yes	New Dwel /
	3341	APT 21 29 ST JAMES'S WALK, RIALTO, DUBLIN 8	Dublin	1.830000e+05	No	Yes	New Dwel /
	3358	1 - 6 (INCL) 11 & 12 THE MEADOWS, CAHEREEN HE...	Kerry	1.378724e+06	No	Yes	New Dwel /
	3407	NO.8 (OTHERWISE SITE NO.3) DUNMASC, PORTLAOISE	Laois	1.096500e+05	Yes	No	Second-Han house /
	3768	11 B (42) MOORE BAY, KILKEE	Clare	1.330000e+05	No	No	Second-Han house /

445519	2020-11-18	6 APARTMENTS, SUNVILLE, CHURCH ROAD	Cork	1350000.00	No	No	New Dwelling house /Apartment
464347	2021-03-09	APT BLOCK B KINGSLODGE APARTMENTS &, APT BLOCK...	Kildare	7048458.00	No	Yes	New Dwelling house /Apartment
469958	2021-04-16	APARTMENTS, NAVENNY PLACE, BALLYBOFEY	Donegal	1450000.00	No	Yes	New Dwelling house /Apartment
491802	2021-09-06	26 APARTMENTS AT BLOCK D, THE PADDOCKS, STATIO...	Kildare	6529515.42	No	Yes	New Dwelling house /Apartment
491855	2021-09-06	61 APARTMENTS AND RETAIL, BALCURRIS ROAD, MAIN...	Dublin	44620000.00	No	Yes	New Dwelling house /Apartment
505793	2021-11-23	33 APARTMENTS AT BLOCK C, THE PADDOCKS, STATIO...	Kildare	8399118.94	No	Yes	New Dwelling house /Apartment
505970	2021-11-24	151 APARTMENTS, BEECH PARK, CABINTEELY	Dublin	57928634.00	No	Yes	New Dwelling house /Apartment
512077	2021-12-17	63 APARTMENTS, KNOCKRABO, MOUNT ANVILLE ROAD	Dublin	35186409.74	No	Yes	New Dwelling house /Apartment
526983	2022-03-30	26 APARTMENTS, THE MILLWORKS, ISLANDBRIDGE	Dublin	11548926.00	No	Yes	New Dwelling house /Apartment
526989	2022-03-30	3 APARTMENTS, THE MILLWORKS, ISLANDBRIDGE	Dublin	1332572.00	No	Yes	New Dwelling house /Apartment
533774	2022-05-13	BLOCK A PHASE 4, CONSISTING OF 39 APARTMENTS,...	Dublin	10779736.00	No	Yes	New Dwelling house /Apartment
533775	2022-05-13	BLOCK B PHASE 4, CONSISTING OF 68 APARTMENTS,...	Dublin	18466960.00	No	Yes	New Dwelling house /Apartment
538022	2022-06-09	46 APARTMENTS AT BLOCK B, THE PADDOCKS, STATIO...	Kildare	12437004.40	No	Yes	New Dwelling house /Apartment

- All these rows appear to be multiple apartments except for 2.

```
In [34]: rows = list(df[(df['Price'] >= p_99) & ((df['Address'].str.contains('APARTMENTS')))].index)
# APARTMENT 3, GROVE COURT APARTMENTS,
rows.remove(406992)
# HERBERTON BLOCK B APT 23, 2 HERBERTON STREET
rows.remove(414532)

# drop rows
df = df.drop(index=rows)
df = df.reset_index(drop=True)
df.shape
```

Out[34]: (592461, 8)

- Look for addresses with keyword 'BLOCK'.

```
In [35]: df[df['Address'].str.contains('BLOCK')]
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofPropert
189	2010-01-08	APT 25 BLOCK 2, THE VIEW NEWTOWN HALL,	Kildare	2.019500e+05		No	New Dwelling house /Apartment

562443	2022-10-26	1-78 THE SKELLIGS BLOCK D, AND 1-10 THE BLASK...	Cork	3.064493e+07	No	Yes	New Dwelling house /Apartment
562630	2022-10-26	APARTMENT 19, THE BURBRIDGE BLOCK 5, LANSDOWNE...	Dublin	1.740088e+06	No	Yes	New Dwelling house /Apartment
564601	2022-11-07	1-50 BLOCK A, THE PADDOCKS, STATION ROAD	Kildare	1.352952e+07	No	Yes	New Dwelling house /Apartment
564729	2022-11-07	APT 13 THE BURBRIDGE, BLOCK 5 LEVEL 3 LANDSD...	Dublin	1.519824e+06	No	Yes	New Dwelling house /Apartment
567734	2022-11-22	APT. 1, IVERAGH BLOCK 3, CAHIRCIVEEN	Kerry	1.800000e+06	No	No	Second-Hand Dwelling house /Apartment
568823	2022-11-28	2 THE BURBRIDGE, BLOCK 5, LANSDOWNE PLACE LANS...	Dublin	1.718062e+06	No	Yes	New Dwelling house /Apartment
570205	2022-12-02	APT 4 THE BURBRIDGE, BLOCK 5 LEVEL 1, LANSDOW...	Dublin	1.453745e+06	No	Yes	New Dwelling house /Apartment
575621	2022-12-21	FOLIO DN244346F, BLOCK 1A, ST EDMUNDS	Dublin	2.391598e+07	No	Yes	New Dwelling house /Apartment
588127	2023-03-27	BLOCK 1 (1-31 ARCHERS WOOD VIEW), BLOCK 2 (1-3...	Wicklow	2.293392e+07	No	Yes	New Dwelling house /Apartment

In [37]:

```
rows = list(df[(df['Price'] >= p_99) & ((df['Address'].str.contains('BLOCK')))].index)

single_apts = ['APT 11 BLOCK 1 THE LINKS, ELM PARK, MERRION ROAD',
               'APT. 1 BLOCK A, BLACKHALL PLACE, ARBOUR HILL', 'APT 36 BLOCK C, HAZLEWOOD, SANTRY', '16 BLOCK B',
               'APT 23 BLOCK L TEMPLEGROVE, CASTLETREASURE, DOUGLAS', 'APT 25, BLOCK 1, GALLERY QUAY',
               '13 ORWELL GATE, BLOCK A, MARIANELLA', 'APARTMENT 13, ORWELL HALL BLOCK C, MARIANELLA',
               'APARTMENT 16 BLOCK 3, THE TEMPLETON LANDSDOWNE PLACE, LANDSDOWNE ROAD',
               'APT 5 BLOCK 3 THE TEMPLETON, LANDSDOWNE PLACE, LANDSDOWNE ROAD',
               'APT 15 BLOCK 3, THE TEMPLETON, LANDSDOWNE PLACE',
               'APT 18 BLOCK 3 THE TEMPLETON, LANDSDOWNE PLACE, LANDSDOWNE ROAD BALLSBIDGE',
               'APARTMENT 14 BLOCK 3, THE TEMPLETON, LANDSDOWNE PLACE',
               'APARTMENT 9 BLOCK 3, THE TEMPLETON, LANDSDOWNE PLACE',
               'APARTMENT 2 THE TEMPLETON, BLOCK 3 LEVEL 1 LANDSDOWNE PLACE, LANDSDOWNE ROAD',
               'APARTMENT 20 BLOCK A, FARNHAM COURT, FARNHAM ROAD',
               'APT 22 THE BARRINGTON, BLOCK 6A LANDSDOWNE PLACE, BALLSBIDGE DUBLIN 4',
               '47 THE BARRINGTON, BLOCK 6B LEVEL 6, LANDSDOWNE PLACE',
               '27 THE BARRINGTON BLOCK 6A LEVEL 8, LANDSDOWNE PLACE BALLSBIDGE, FOLIO NO. DN211931F',
               '24 THE BARRINGTON, BLOCK 6A, LANDSDOWNE PLACE', 'APT 46 THE BARRINGTON, BLOCK 6B, LANDSDOWNE PL',
               '207 BLOCK A1, BEACON SOUTH QUARTER, SANDYFORD', '8 THE BLAKE, BLOCK 4 LEVEL 2, LANDSDOWNE PLA',
               'APT 11 THE BLAKE, BLOCK 4 LEVEL 3, LANDSDOWNE PLACE', 'APT 15 THE BLAKE, BLOCK 4 LEVEL 4',
               'APARTMENT 19 THE BLAKE BLOCK 4, LEVEL 5 LANDSDOWNE PLACE, LANDSDOWNE ROAD BALLSBIDGE',
               'APT 20 THE TEMPLETON, BLOCK 3 LEVEL 6, BALLSBIDGE', 'APT 21 THE BLAKE, BLOCK 4 LEVEL 6',
               'APARTMENT 302 BLOCK 2, THE ELM, TRIMBLESTON', '51 THE BARRINGTON, BLOCK 6B LEVEL 7, LANDSDOWNE I',
               'APT 14 THE BLAKE, BLOCK 4 LEVEL 3, LANDSDOWNE PLACE LANDSDOWNE ROAD',
               '18 DONNYBROOK GARDENS, BLOCK 1 DONNYBROOK GARDENS, GREENFIELDS',
               'APARTMENT 13 THE BARRINGTON, BLOCK 6A LEVEL 3, LANDSDOWNE PLACE',
               'APT.604 - BLOCK A3, BEACON SOUTH QUARTER, SANDYFORD',
               'APT 10 THE BLAKE, BLOCK 4 LEVEL 2 LANDSDOWNE PLACE, LANDSDOWNE ROAD',
               'APARTMENT 12 THE BURBRIDGE, BLOCK 5 LEVEL 3, LANDSDOWNE PLACE LANDSDOWNE ROAD',
               'APARTMENT 19, THE BURBRIDGE BLOCK 5, LANDSDOWNE PLACE LANDSDOWNE ROAD',
               'APT 13 THE BURBRIDGE, BLOCK 5 LEVEL 3 LANDSDOWNE PLACE, LANDSDOWNE ROAD', 'APT. 1, IVERAGH BLO',
               '2 THE BURBRIDGE, BLOCK 5, LANDSDOWNE PLACE LANDSDOWNE ROAD', 'APT 4 THE BURBRIDGE, BLOCK 5 LEVEL ]']

rows.remove(df[df['Address']==('UNIT 172 (BLOCK Z) LEARGAN, BALLYMONEEN WEST, GAWLAY')].index)
for apt in single_apts:
    index = list(df[(df['Price'] >= p_99) & (df['Address'].str.contains(apt))].index)
    for num in index:
        rows.remove(num)
# drop rows
df = df.drop(index=rows)
df = df.reset_index(drop=True)
df.shape
```

Out[37]: (592375, 8)

- Look for addresses with keyword 'RESIDENTIAL'.

In [38]: df[df['Address'].str.contains('RESIDENTIAL')]

Out[38]:		DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Prop
	35807	2011-11-15	RESIDENTIAL DWELLING, BALLYMACREDMOND, BUTLERS...	Cork	104000.00		No	No	Second-Hand Dwelling house /Apartment
	45689	2012-05-15	APARTMENT 2, EYRE SQUARE RESIDENTIAL CENTRE, E...	Galway	155000.00		No	No	Second-Hand Dwelling house /Apartment
	88288	2013-11-20	20 RESIDENTIAL PROPERTIES, THE COAST, BALDOYLE	Dublin	2129955.95		No	Yes	New Dwelling house /Apartment
	120796	2014-09-24	11 RESIDENTIAL DWELLINGS AT ARDMORE, BETTYSTOWN	Meath	704975.00		No	Yes	New Dwelling house /Apartment
	135720	2014-12-22	23 DUNARAS, RESIDENTIAL ACCOMMODATION, BISHOP ...	Galway	83000.00		No	No	Second-Hand Dwelling house /Apartment
	139169	2015-01-21	RESIDENTIAL PART OF THE HAZEL BAR, KILBEGGAN, ...	Westmeath	20000.00		No	No	Second-Hand Dwelling house /Apartment
	180706	2015-11-30	UNIT 3 BALLYLOGUE (RESIDENTIAL), THE ROWER, TH...	Kilkenny	210000.00		No	No	Second-Hand Dwelling house /Apartment
	191472	2016-02-19	51 DUNARAS, RESIDENTIAL ACCOMMODATION, BISHOP ...	Galway	70000.00		No	No	Second-Hand Dwelling house /Apartment
	198841	2016-04-25	36 MANOR RD, FOR RESIDENTIAL A C, PALMERSTOWN	Dublin	100000.00		No	No	Second-Hand Dwelling house /Apartment
	209254	2016-07-13	16 SKIBBEREEN RESIDENTIAL CARE CTR, BALTIMORE ...	Cork	102000.00		No	No	Second-Hand Dwelling house /Apartment
	228818	2016-11-25	19 EYRE SQ RESIDENTIAL CENTRE, EYRE SQ, GALWAY	Galway	160000.00		No	No	Second-Hand Dwelling house /Apartment
	237793	2017-01-20	RESIDENTIAL PROPERTY OVER AM-PM, MAIN STREET, ...	Leitrim	60000.00		No	No	Second-Hand Dwelling house /Apartment
	237944	2017-01-24	12 DUNARAS RESIDENTIAL ACCOMD, BISHOP O DONNEL...	Galway	122500.00		No	No	Second-Hand Dwelling house /Apartment
	238987	2017-01-31	RESIDENTIAL, 19/20 MEATH ST, DUBLIN 8	Dublin	609000.00		No	No	Second-Hand Dwelling house /Apartment
	257342	2017-06-19	32, DUNARAS RESIDENTIAL ACCOMODATION, BISHOP O...	Galway	136000.00		No	No	Second-Hand Dwelling house /Apartment
	261369	2017-07-13	RESIDENTIAL UNIT 1, ABBOTT CLOSE, ASKEATON	Limerick	50000.00		No	No	Second-Hand Dwelling house /Apartment
	271150	2017-09-14	247 SWORDS RD, FOR RESIDENTIAL A C, DUBLIN 9	Dublin	150000.00		No	No	Second-Hand Dwelling house /Apartment
	280947	2017-11-14	12 SKIBBEREEN RESIDENTIAL CARE CTR, BALTIMORE ...	Cork	110000.00		No	No	Second-Hand Dwelling house /Apartment
	297077	2018-02-22	RESIDENTIAL UNIT 7, NEIGHBOURHOOD VILLAGE CAST...	Galway	90000.00		No	No	Second-Hand Dwelling house /Apartment
	297926	2018-02-28	RESIDENTIAL OVER JAMESIE'S BAR, MAIN ST, BALLI...	Donegal	48000.00		No	No	Second-Hand Dwelling house /Apartment
	339176	2018-11-22	7 SKIBBEREEN RESIDENTIAL CARE CTR, BALTIMORE R...	Cork	135000.00		No	No	Second-Hand Dwelling house /Apartment
			RESIDENTIAL						

565706	2022-11-11	APT 2, EYRE SQ RESIDENTIAL CENTRE, EYRE SQ	Galway	307000.00	No	No	Second-Hand Dwelling house /Apartment
--------	------------	--	--------	-----------	----	----	--

```
In [39]: error_properties = ['20 RESIDENTIAL PROPERTIES, THE COAST, BALDOYLE', '11 RESIDENTIAL DWELLINGS AT ARDMORE, BET  
'102 RESIDENTIAL HOUSES, ST EDMUNDS, PALMERSTOWN', 'RESIDENTIAL PROPERTIES, STATION MANOR S'  
'RESIDENTIAL PROPERTIES, SCHOLARSTOWN WOOD SCHOLARSTOWN ROAD, RATHFARNHAM']

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)

# drop rows
df = df.reset_index(drop=True)
df.shape
```

Out[39]: (592370, 8)

- Look for addresses with keyword 'PROPERTIES'.

In [40]: df[df['Address'].str.contains('PROPERTIES')]

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Propert
116968	2014-08-22	PROPERTIES AT BREMORE PASTURES, BALBRIGGAN, CO...	Dublin	3315341.46		No	No	Second-Hand Dwelling house /Apartment
116969	2014-08-22	PROPERTIES AT BREMORE PASTURES, BALBRIGGAN, CO...	Dublin	2119644.54		No	No	Second-Hand Dwelling house /Apartment
131301	2014-12-09	12 PROPERTIES AT BAKER'S YARD, PORTLAND STREET...	Dublin	1229331.90		No	Yes	New Dwelling house /Apartment
198594	2016-04-22	2 CAROLANS COURT & OTHER PROPERTIES, KILRONAN ...	Roscommon	960000.00		No	No	Second-Hand Dwelling house /Apartment
199086	2016-04-26	VARIOUS PROPERTIES MILLHOUSE, WOODLAND GROVE ...	Wexford	409500.00		No	No	Second-Hand Dwelling house /Apartment
199087	2016-04-26	VARIOUS PROPERTIES OAKDALE PARK, TEMPLEMORE, T...	Tipperary	934426.00		No	No	Second-Hand Dwelling house /Apartment
202811	2016-05-25	VARIOUS PROPERTIES AT CLUAIN SHEE, AIKEN'S VIL...	Dublin	5594511.01		No	Yes	New Dwelling house /Apartment
289710	2017-12-21	5 DOOGART BEACH PROPERTIES, DOOGART, ACHILL	Mayo	130000.00		No	No	Second-Hand Dwelling house /Apartment
383987	2019-09-03	6 DOOGART BEACH PROPERTIES, DOOGART, ACHILL	Mayo	115000.00		No	No	Second-Hand Dwelling house /Apartment
409139	2020-01-29	APT 7 THE GARDENS, APPLEWOOD VILLAGE SWORDS, ...	Dublin	1920000.00		No	No	Second-Hand Dwelling house /Apartment
428561	2020-07-24	3 CRUISE PARK AVE CRUISE PARK, AND 29 OTHER P...	Dublin	8090730.00		No	No	Second-Hand Dwelling house /Apartment
438679	2020-10-09	VARIOUS PROPERTIES THE DRIVE, SCHOLARSTOWN WO...	Dublin	2144935.92		No	Yes	New Dwelling house /Apartment
		VARIOUS						

438680	2020-10-09	PROPERTIES, HOLLYWOODRATH GREEN, HOLLY...	Dublin	3108370.04	No	Yes	New Dwelling house /Apartment
462784	2021-02-26	VARIOUS PROPERTIES, HOLLYWOODRATH GREEN, HOLLY...	Dublin	5392070.48	No	Yes	New Dwelling house /Apartment
471598	2021-04-29	8 DOOGART BEACH PROPERTIES, DOOGART, ACHILL	Mayo	118000.00	No	No	Second-Hand Dwelling house /Apartment
528164	2022-04-06	OLD SCHOOL MASTERS PROPERTIES, BALLINDERREEN, ...	Galway	120000.00	No	No	Second-Hand Dwelling house /Apartment
544137	2022-07-14	6 PORTSIDE COURT, WEST RD EAST WALL, D3 & OT...	Dublin	17099496.00	No	No	Second-Hand Dwelling house /Apartment
544179	2022-07-14	APT 58 BELMONT HALL, GARDINER STREET MIDDLE, ...	Dublin	8533138.00	No	No	Second-Hand Dwelling house /Apartment

```
In [41]: multi_prop = df[(df['Address'].str.contains('PROPERTIES') & (~df['Address'].str.contains('DOOGART')))].index
df = df.drop(index=multi_prop)
df = df.reset_index(drop=True)
df.shape
```

Out[41]: (592355, 8)

- Look for addresses with keyword 'DEVELOPMENT'.

```
In [42]: pd.set_option('display.float_format', lambda x: '%.3f' % x)
df[(df['Address'].str.contains('DEVELOPMENT')) & (df['Price'] >= p_99)]
```

Out[42]:		DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Property:
	90227	2013-12-06	DEVELOPMENT, ISLAND KEY, EAST ROAD	Dublin	11603965.000		No	No	Second-Hand Dwelling house /Apartment
	90228	2013-12-06	DEVELOPMENT, ORCHARD GARDENS, DENNEHY'S CROSS	Cork	11789574.000		No	No	Second-Hand Dwelling house /Apartment
	90229	2013-12-06	DEVELOPMENT, TALLAGHT CROSS, TALLAGHT	Dublin	20804659.000		No	No	Second-Hand Dwelling house /Apartment
	157843	2015-06-17	HEUSTON SOUTH QUARTER DEVELOPMENT, ST. JOHN'S ...	Dublin	35045714.000		Yes	Yes	New Dwelling house /Apartment
	170922	2015-09-22	GORTANORA DEVELOPMENT, DINGLE	Kerry	2925000.000		No	Yes	New Dwelling house /Apartment
	426219	2020-07-02	12 UNIT DEVELOPMENT, AT CROMWELLSFORD COURT, C...	Dublin	5068387.380		No	Yes	New Dwelling house /Apartment
	428789	2020-07-27	34 UNIT DEVELOPMENT AT WESTON, KETTLE'S LANE ...	Dublin	12753320.000		No	Yes	New Dwelling house /Apartment
	440630	2020-10-22	9 UNIT DEVELOPMENT AT, RICHMOND WAY CLONDALKIN	Dublin	3052864.000		Yes	Yes	New Dwelling house /Apartment
	444055	2020-11-10	RESIDENTIAL DEVELOPMENT, THE OAKS, PORTERSTOWN...	Dublin	4396000.000		No	No	Second-Hand Dwelling house /Apartment
	459173	2021-02-04	MIXED USE DEVELOPMENT, AT SWISS COTTAGE, SANTRY	Dublin	38094573.370		No	Yes	New Dwelling house /Apartment
	486996	2021-08-09	40 UNIT DEVELOPMENT, AT FAIRGREEN MANOR, DUNLA...	Wicklow	10500000.000		No	Yes	New Dwelling house /Apartment
	530931	2022-04-26	SITE AT BARNOAKS DEVELOPMENT, CITYWEST ROAD, C...	Dublin	107488987.000		No	Yes	New Dwelling house /Apartment
	555463	2022-09-16	DAVITT PRS DEVELOPMENT, DAVITT ROAD, DRIMNAGH	Dublin	111280172.000		No	Yes	New Dwelling house /Apartment
	570925	2022-12-07	APARTMENT DEVELOPMENT, AT CASTLEVIEW, CARRIGLEA	Dublin	132675590.200		No	Yes	New Dwelling house /Apartment

```
In [43]: rows = df[(df['Address'].str.contains('DEVELOPMENT')) & (df['Price'] >= p_99)].index

# drop rows
df = df.drop(index=rows)
df = df.reset_index(drop=True)
df.shape
```

Out[43]: (592341, 8)

```
In [44]: df[(df['Address'].str.contains('DEVELOPMENT'))]
```

Out[44]:		DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Property:
	32571	2011-09-19	52 MARINER'S VIEW DEVELOPMENT, CASTLETOWNBERE	Cork	58562.000		No	No	Second-Hand Dwelling house /Apartment
			48 ROS ARD,						

```
In [45]: # drop rows
df = df.drop(index=df[df['Address'] == 'LINN CORR DEVELOPMENT, CASTLMARTYR'].index)
df = df.reset_index(drop=True)
df.shape
```

Out[45]: (592340, 8)

- Look for addresses with keyword 'HOUSES'.

```
In [46]: df[df['Address'].str.contains('HOUSES')]
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Prop
7891	2010-06-09	19 NEW HOUSES, TYHOLLAND, MONAGHAN	Monaghan	95000.000		No	No	Second-Hand Dwelling house /Apartment
19244	2010-12-06	NO. 1 THE COACH HOUSES, RICHMOND, TEMPLEMORE	Tipperary	120000.000		No	No	Second-Hand Dwelling house /Apartment
19248	2010-12-06	NO.4 THE COACH HOUSES, RICHMOND, TEMPLEMORE	Tipperary	120000.000		No	No	Second-Hand Dwelling house /Apartment
19348	2010-12-07	THE COACH HOUSES AND THE COURTYARD, GRANGEMOUN...	Dublin	15000.000		No	No	Second-Hand Dwelling house /Apartment
20601	2010-12-22	3 MINERS HOUSES, CLONMEL ROAD, TIPPERARY	Tipperary	130000.000		No	No	Second-Hand Dwelling house /Apartment
24778	2011-04-13	NEW HOUSES, KILL	Waterford	55000.000		Yes	No	Second-Hand Dwelling house /Apartment
33439	2011-10-03	GRAIGUE LOWER, SEVEN HOUSES	Kilkenny	140000.000		No	No	Second-Hand Dwelling house /Apartment
36782	2011-11-30	1 NEW HOUSES, KILL	Waterford	62500.000		No	No	Second-Hand Dwelling house /Apartment
39014	2011-12-22	CRONE HOUSES, SHILLELAGH	Wicklow	480000.000		No	No	Second-Hand Dwelling house /Apartment
39109	2011-12-23	8 NEW HOUSES, KINLOUGH	Leitrim	18000.000		No	No	Second-Hand Dwelling house /Apartment
49869	2012-07-18	'NEWHOUSES', KILRANE, ROSSLARE HARBOUR	Wexford	200000.000		No	No	Second-Hand Dwelling house /Apartment
66560	2013-02-13	21 THE COACH HOUSES, FOREST PARK, BALLINATRAY ...	Wexford	60000.000		No	No	Second-Hand Dwelling house /Apartment
85639	2013-10-23	13 HOUSES AT COOLINE DRIVE, COBH	Cork	1511011.000		No	Yes	New Dwelling house /Apartment
94376	2014-01-06	NEW HOUSES, KILRANE, CO. WEXFORD	Wexford	110000.000		No	No	Second-Hand Dwelling house /Apartment
99390	2014-03-12	3 CASTLE HOUSES, NEWTOWNFORBES, CO LONGFORD	Longford	50250.000		No	No	Second-Hand Dwelling house /Apartment
102612	2014-04-16	9 THE COACHOUSES, COURTOWN, WEXFORD	Wexford	58500.000		No	No	Second-Hand Dwelling house /Apartment
120077	2014-09-18	1 HARTYS HOUSES, THE SQUARE, ARDFERT	Kerry	132500.000		No	No	Second-Hand Dwelling house /Apartment
121424	2014-09-29	24 HOUSES AT THE TANNERY COOLFADDA, DUNMANWAY...	Cork	2303695.000		No	Yes	New Dwelling house /Apartment
121425	2014-09-29	24 HOUSES AT THE TANNERY COOLFADDA, DUNMANWAY...	Cork	2303965.000		No	Yes	New Dwelling house /Apartment

510829	2021-12-15	CASTLEGREGORY	Kerry	280000.000	No	No	house /Apartment
514142	2021-12-23	9 BREWERY HOUSES, KILSARAN, CASTLEBELLINGHAM	Louth	186000.000	No	No	Second-Hand Dwelling house /Apartment
515670	2022-01-18	2 APT ROOF HOUSES, NEW LINE RD, LETTERKENNY	Donegal	128000.000	No	No	Second-Hand Dwelling house /Apartment
517092	2022-01-27	10 TOWN HOUSES, SHANNON GROVE, CARRICK ON SHANNON	Leitrim	165000.000	No	No	Second-Hand Dwelling house /Apartment
529087	2022-04-12	4 TOWN HOUSES, SHANNON GROVE, CARRICK ON SHANNON	Leitrim	147137.000	No	Yes	New Dwelling house /Apartment
530739	2022-04-25	NEW HOUSES, SLEEVEN, KILL	Waterford	228000.000	No	No	Second-Hand Dwelling house /Apartment
530754	2022-04-26	1 SHANNON GROVE TOWNHOUSES, LISNAGOT, CARRICK	Leitrim	122000.000	No	No	Second-Hand Dwelling house /Apartment
532621	2022-05-09	...	Limerick	31438.000	Yes	No	Second-Hand Dwelling house /Apartment
540494	2022-06-23	1 NEW HOUSES, FAIR HILL, RATHKEALE	Leitrim	168000.000	No	No	Second-Hand Dwelling house /Apartment
546216	2022-07-26	9 TOWN HOUSES, SHANNON GROVE, CARRICK ON SHANNON	Wexford	215000.000	No	No	Second-Hand Dwelling house /Apartment
550074	2022-08-18	4 THE COACH HOUSES, FOREST PARK, COURTOWN HARBOUR	Mayo	250000.000	No	No	Second-Hand Dwelling house /Apartment
550360	2022-08-19	3 HOUSES AT CORNMARKET, BALLINROBE, COUNTY MAYO	Leitrim	149779.000	No	Yes	New Dwelling house /Apartment
551368	2022-08-25	5 TOWN HOUSES, SHANNON GROVE, CARRICK ON SHANNON	Limerick	185000.000	No	No	Second-Hand Dwelling house /Apartment
559395	2022-10-10	LONG HOUSES, KNOCKLONG, CO LIMERICK	Leitrim	120000.000	No	No	Second-Hand Dwelling house /Apartment
566540	2022-11-17	3 TOWN HOUSES, SHANNON GROVE, CARRICK ON SHANNON	Carlow	175000.000	No	No	Second-Hand Dwelling house /Apartment
584975	2023-03-07	14 CASTLEVIEW TOWNHOUSES, GRAIGUECULLEN, CARLOW	Monaghan	113000.000	No	No	Second-Hand Dwelling house /Apartment
585066	2023-03-07	13 NEW HOUSES, TYHOLLAND, CO MONAGHAN	Kerry	315000.000	No	No	Second-Hand Dwelling house /Apartment

- Many addresses begin with a number followed by 'NEW HOUSES'. This could be interpreted as the sale of multiple houses but looking at the prices, I believe that 'NEW HOUSES' is actually part of the address. I will drop the rows where it is clear multiple houses were included in the sale.

```
In [47]: # remove multiple houses in top 1%
rows = df[(df['Address'].str.contains('HOUSES')) & (df['Price'] >= p_99)].index
# drop rows
df = df.drop(index=rows)
```

```
In [48]: df[df['Address'].str.contains('HOUSES')]
```

DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Propo
7891 2010-06-09	19 NEW HOUSES, TYHOLLAND, MONAGHAN	Monaghan	95000.000		No	Second-Hand Dwelling house /Apartment	

551368	2022-08-25	KNOCKLONG, CO LIMERICK	Limerick	185000.000	No	No	house /Apartment
559395	2022-10-10	3 TOWN HOUSES, SHANNON GROVE, CARRICK ON SHANNON	Leitrim	120000.000	No	No	Second-Hand Dwelling house /Apartment
566540	2022-11-17	14 CASTLEVIEW TOWNHOUSES, GRAIGUECULLEN, CARLOW	Carlow	175000.000	No	No	Second-Hand Dwelling house /Apartment
584975	2023-03-07	13 NEW HOUSES, TYHOLLAND, CO MONAGHAN	Monaghan	113000.000	No	No	Second-Hand Dwelling house /Apartment
585066	2023-03-07	8 STATION HOUSES, CASTLEGROGERY, KERRY	Kerry	315000.000	No	No	Second-Hand Dwelling house /Apartment

```
In [49]: # remove other multiple houses, finished or otherwise
error_properties = ['2 UNFINISHED HOUSES KILGLASS, KILGLASS, COUNTY ROSCOMMON',
                    '10 PART BUILT HOUSES, AT NEWCASTLE LYONS, LUCAN', 'PART CONSTRUCTED DWELLINGHOUSE @, NEW HOI
                    '3 HOUSES AT CORNMARKET, BALLINROBE, COUNTY MAYO']

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)

# drop rows
df = df.reset_index(drop=True)
df.shape
```

Out[49]: (592329, 8)

- Look for addresses with keyword 'HOMES'.

In [50]: df[df['Address'].str.contains('HOMES')]

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Prop
2347	2010-02-25	22 LISCANNOR HOLIDAY HOMES, LISCANNOR, CO. CLARE	Clare	450000.000		No	No	Second-Hand Dwelling house /Apartment
9588	2010-07-05	APT F3 RAHEEN HOLIDAY HOMES, CASTLEGRANGE, RAHEEN	Limerick	120000.000		No	No	Second-Hand Dwelling house /Apartment
9703	2010-07-07	2 FENIT ATLANTIC HOLIDAY HOMES, FENIT	Kerry	155000.000		No	No	Second-Hand Dwelling house /Apartment
25749	2011-05-09	9 AN GLEANTHAIN HOLIDAY HOMES, KENMORE	Kerry	200000.000		No	No	Second-Hand Dwelling house /Apartment
26504	2011-05-25	9 BALLYBUNION SEASIDE HOLIDAY HOMES, AHERN ROA...	Kerry	155000.000		No	No	Second-Hand Dwelling house /Apartment
28040	2011-06-24	3 GORTAMULLIN HOLIDAY HOMES, KENMORE	Kerry	120000.000		No	No	Second-Hand Dwelling house /Apartment
30264	2011-08-05	1 TRIDENT HOMES, CASTLE STREET, LISCANNOR	Clare	150000.000		No	No	Second-Hand Dwelling house /Apartment
40783	2012-02-09	11 GORTAMULLEN HOLIDAY HOMES, KENMORE	Kerry	116000.000		No	No	Second-Hand Dwelling house /Apartment
41899	2012-03-01	HOMESTEAD, 19 RATHDOWN PARK, TERENURE	Dublin	1285000.000		No	No	Second-Hand Dwelling house /Apartment
43501	2012-04-02	NO. 1 BUNBEG HOLIDAY HOMES, BUNBEG	Donegal	100000.000		No	No	Second-Hand Dwelling house /Apartment
43759	2012-04-05	8 CREEESLOUGH HOLIDAY HOMES, CREEESLOUGH	Donegal	88500.000		No	No	Second-Hand Dwelling house /Apartment
49902	2012-07-18	5 BALLINGLANTA HOLIDAY HOMES, BALLINGLANTA, CL...	Cork	135000.000		No	No	Second-Hand Dwelling house /Apartment

DONEGAL							
561020	2022-10-19	11 BALLYHEIGUE, HOLIDAY HOMES, KERRY	Kerry	230000.000	No	No	Second-Hand Dwelling house /Apartment
561609	2022-10-21	25 LISHEEN HOMES, CAHIRDAVIN, LIMERICK	Limerick	205000.000	No	No	Second-Hand Dwelling house /Apartment
563322	2022-10-28	9 MUCKROSS HOLIDAY HOMES, MUCKROSS RD, KILLARNEY	Kerry	300000.000	No	No	Second-Hand Dwelling house /Apartment
567750	2022-11-23	3 BALLYVAUGHAN HOLIDAY HOMES, BALLYVAUGHAN, CLARE	Clare	194000.000	No	No	Second-Hand Dwelling house /Apartment
571366	2022-12-09	1 DOOLIN HOLIDAY HOMES, DOOLIN CROSS, CLARE	Clare	262500.000	No	No	Second-Hand Dwelling house /Apartment
573288	2022-12-15	8 SHANAGARRY HOLIDAY HOMES, SHANAGARRY, CORK	Cork	235000.000	No	No	Second-Hand Dwelling house /Apartment
573302	2022-12-15	9 SHANAGARRY HOLIDAY HOMES, SHANAGARRY, CORK	Cork	225000.000	No	No	Second-Hand Dwelling house /Apartment
574261	2022-12-19	58 CHARLESTON TOWN HOMES, CHARLESTON ROAD, RAN...	Dublin	2202643.170	No	Yes	New Dwelling house /Apartment
574458	2022-12-20	10 PARKLANDS HOLIDAY HOMES, PORT RD, KILLARNEY	Kerry	295000.000	No	No	Second-Hand Dwelling house /Apartment
574784	2022-12-20	67 BANNA HOLIDAY HOMES, BANNA, KERRY	Kerry	85000.000	No	No	Second-Hand Dwelling house /Apartment
578756	2023-01-25	7 TRAMORE HOLIDAY HOMES, KEEL, ACHILL	Mayo	117500.000	Yes	No	Second-Hand Dwelling house /Apartment
586553	2023-03-16	51 BANNA HOLIDAY HOMES, BANNA, TRALEE	Kerry	106250.000	No	No	Second-Hand Dwelling house /Apartment
590569	2023-04-17	APT B2 RAHEEN HOLIDAY HOMES, RAHEEN, LIMERICK	Limerick	190000.000	No	No	Second-Hand Dwelling house /Apartment
591123	2023-04-21	23 SHANAGARRY HOLIDAY HOMES, SHANAGARRY, CORK	Cork	220000.000	No	No	Second-Hand Dwelling house /Apartment

```
In [51]: error_properties = ['HOLIDAY HOMES, BALLYVAUGHAN, CO.CLAIRE', '74 HOOKLESS HOLIDAY HOMES, FETHARD ON SEA', 'GLENDALOUGH HOLIDAY HOMES, LARAGH']

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)

# drop rows
df = df.reset_index(drop=True)
df.shape
```

Out[51]: (592326, 8)

- Look for addresses with keyword 'OTHERS'.

```
In [52]: df[df['Address'].str.contains(' OTHERS ')]
```

Out [52] :	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Pro
268374	2017-08-28	APT 32 RIVERBANK HALL & OTHERS, ADDISON PARK, ...	Dublin	915000.000	Yes	No	Second-Hand Dwelling house /Apartment	
282670	2017-11-23	30 AND OTHERS BLOCK B HAZELWOOD, NORTHWOOD, SA...	Dublin	1092857.140	No	No	Second-Hand Dwelling house /Apartment	
291821	2018-01-16	3 THE MEWS & OTHERS, LOWER GATE, CASHEL	Tipperary	425000.000	No	No	Second-Hand Dwelling house /Apartment	
322558	2018-08-15	APT 1 (AND OTHERS), 62 LOWER DORSET ST, DUBLIN 1	Dublin	500000.000	No	No	Second-Hand Dwelling house /Apartment	
344797	2018-12-17	APT1 AND OTHERS, NEW BANCROFT SQUARE, MAIN STREET	Dublin	33920704.000	No	No	Second-Hand Dwelling house /Apartment	
346116	2018-12-20	1 THE GROODY STUDENT PARK & OTHERS, CASTLETROY...	Limerick	6050000.000	No	No	Second-Hand Dwelling house /Apartment	
376176	2019-07-18	31 AND OTHERS FENTON GREEN, CHURCH STREET, KIL...	Kildare	1607048.460	No	Yes	New Dwelling house /Apartment	
397338	2019-11-20	53 CARTON GROVE AND OTHERS, DUBLIN ROAD, MAYN...	Kildare	4833945.730	No	Yes	New Dwelling house /Apartment	
401489	2019-12-10	APT 1 (AND OTHERS), SHELBOURNE PLAZA, CHARLOTT...	Dublin	22092260.470	No	No	Second-Hand Dwelling house /Apartment	
402210	2019-12-12	9 THE AVENUE (AND OTHERS), SEMPLE WOODS, DONABATE	Dublin	8757890.750	No	Yes	New Dwelling house /Apartment	
404805	2019-12-19	47 MARIAVILLA WOODS (AND OTHERS), MOYGLARE ROA...	Kildare	21155144.720	No	Yes	New Dwelling house /Apartment	
428114	2020-07-21	APT 1 & OTHERS, 53 NORTHUMBERLAND ROAD, BALLSB...	Dublin	1800000.000	No	No	Second-Hand Dwelling house /Apartment	
450678	2020-12-11	6 WILLOUISE & OTHERS, SALLINS, COUNTY KILDARE	Kildare	7753303.960	No	Yes	New Dwelling house /Apartment	
464126	2021-03-09	20 BARONS WAY (AND OTHERS), DUNDALK, LISMULLEN	Louth	1550661.600	No	Yes	New Dwelling house /Apartment	
474966	2021-05-21	UNIT 4 THE AVENUE AND OTHERS, BELLINGSMORE, CH...	Dublin	3118942.720	No	Yes	New Dwelling house /Apartment	
478006	2021-06-11	UNIT 2 THE CRESCENT AND OTHERS, BELLINGSMORE, ...	Dublin	3858059.000	No	Yes	New Dwelling house /Apartment	
516690	2022-01-25	11 RIVER OAKS & OTHERS, BALLAGHADERREEN, CO. R...	Roscommon	750000.000	No	Yes	New Dwelling house /Apartment	
524209	2022-03-10	BLOCK 7 NO. 34 THE DARGLE & OTHERS, PROSPECT H...	Dublin	970000.000	No	No	Second-Hand Dwelling house /Apartment	
567801	2022-11-23	50 A AND OTHERS, FERRYBANK, ARKLOW	Wicklow	266667.000	Yes	No	Second-Hand Dwelling house /Apartment	

In [53]: rows = df[df['Address'].str.contains(' OTHERS')].index

drop rows

```
df = df.drop(index=rows)
df = df.reset_index(drop=True)
df.shape
```

Out[53]: (592307, 8)

- Look for addresses with keywords '&/'AND' and 'APT'/'APARTMENT'.

In [54]: df[(df['Address'].str.contains('&') | df['Address'].str.contains(' AND')) & (df['Address'].str.contains('APT') | df['Address'].str.contains('APARTMENT'))]

		DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Pro
21539	2011-01-24	SUPERMARKET CARPARK &APT, THE FAYTHE	Wexford	85000.000		No	No	Second-Hand Dwelling house /Apartment	
23368	2011-03-09	APT B & D, 14 DOUGHISKA ROAD, DOUGHISKA	Galway	220000.000		No	Yes	New Dwelling house /Apartment	g 38
68299	2013-03-14	APT NO. 15 5TH FLOOR BLOCK A, STONETOWN ENN...	Limerick	358943.000		No	Yes	New Dwelling house /Apartment	g
76976	2013-07-17	THE RED COTTAGE & APTS, LISNALURG, BUNDORAN ROAD	Sligo	150000.000		No	No	Second-Hand Dwelling house /Apartment	
84420	2013-10-09	THE TEMPLE BAR PUB & 9 APARTMENTS, TEMPLE BAR	Dublin	1195000.000		Yes	No	Second-Hand Dwelling house /Apartment	
94807	2014-01-15	123A CORK ST, APT1 & 2, DUBLIN 8	Dublin	180000.000		No	No	Second-Hand Dwelling house /Apartment	
100038	2014-03-20	APARTMENTS 18 AND 18A, HARLEY WOOD, TOGHER	Cork	108000.000		No	No	Second-Hand Dwelling house /Apartment	
106200	2014-05-27	GROUND FLOOR APT1 APT2 AND APT3, 50 UPPER BA...	Dublin	761400.000		No	No	Second-Hand Dwelling house /Apartment	
116729	2014-08-21	OFFICES APARTMENT AND CAR PARK, 35-45 SOUTH KI...	Dublin	492000.000		No	No	Second-Hand Dwelling house /Apartment	
140734	2015-01-30	APARTMENTS AND SMYTHS PUBLIC HOUSE, 75/77 RANE...	Dublin	320000.000		No	No	Second-Hand Dwelling house /Apartment	
145190	2015-03-06	PH APT 4TH FLOOR, 73 & 74 SOUTH MAIN STREET, B...	Cork	50000.000		No	No	Second-Hand Dwelling house /Apartment	
152450	2015-05-06	APT1 AND 2, 19 TRINITY ST, WEXFORD	Wexford	62100.000		No	No	Second-Hand Dwelling house /Apartment	
160862	2015-07-09	APT 16D AND MOORING BERTH, 10 HARBOUR VILLAGE ...	Clare	140000.000		No	No	Second-Hand Dwelling house /Apartment	
161792	2015-07-16	APT. 51 ST. ANN'S AILESBUERY, ROAD DUBLIN 4 A...	Dublin	1740000.000		No	No	Second-Hand Dwelling house /Apartment	
164625	2015-08-05	APARTMENT 13 DUN NA SI, RATHCORE GOLF & COUNT...	Meath	63000.000		No	Yes	New Dwelling house /Apartment	g 38
183948	2015-12-17	APT.6 FOX AND MARINE COTTAGE, KILMORE QUAY, CO...	Wexford	145000.000		No	No	Second-Hand Dwelling house /Apartment	
190116	2016-02-10	APT & CAR SPACE NO. 10, HOBBLE COURT, LEAMY ST...	Waterford	40000.000		No	No	Second-Hand Dwelling house /Apartment	
190117	2016-02-10	APT & CAR SPACE NO. 1, HOBBLE COURT, LEAMY STREET	Waterford	40000.000		No	No	Second-Hand Dwelling house /Apartment	
190118	2016-02-10	APT & CAR SPACE NO. 2, HOBBLE COURT, LEAMY STREET	Waterford	40000.000		No	No	Second-Hand Dwelling house /Apartment	
190119	2016-02-10	APT & CAR SPACE NO. 3, HOBBLE COURT, LEAMY STREET	Waterford	40000.000		No	No	Second-Hand Dwelling house /Apartment	
190120	2016-02-10	APT & CAR SPACE NO. 4, HOBBLE COURT, LEAMY STREET	Waterford	40000.000		No	No	Second-Hand Dwelling house /Apartment	

426330	2020-07-03	APT 2 3 4 5 6 BARROW COURT...	Laois	141890.000	No	No	Second-Hand Dwelling house /Apartment
450795	2020-12-11	APT AT JJ BRENNAN & SON, MAIN STREET, CASTLEB...	Monaghan	40000.000	No	No	Second-Hand Dwelling house /Apartment
453262	2020-12-18	APT NOS. 49A 49B 50A 50D & 50E, FERRYBANK, ...	Wicklow	245000.000	No	No	New Dwelling house /Apartment
475444	2021-05-26	APT.11 - BURGUNDY COURT, AND BURGUNDY HOUSE, ...	Dublin	250000.000	No	No	Second-Hand Dwelling house /Apartment
495996	2021-09-30	APARTMENT 3, GROUND & FIRST FLOOR, KILMACUD HO...	Dublin	495392.000	No	No	New Dwelling house /Apartment
513780	2021-12-22	APT 1 APT2 APT3 & APT4, 1 NEW RD, TRALEE	Kerry	70000.000	No	No	Second-Hand Dwelling house /Apartment
521928	2022-02-25	APARTMENT 1, BRANNIGAN'S SPORTS BAR, 17/1...	Cork	100000.000	No	No	Second-Hand Dwelling house /Apartment
525479	2022-03-21	86A CAPTAIN'S AVENUE, CRUMLIN	Dublin	460000.000	No	Yes	New Dwelling house /Apartment
532558	2022-05-06	GROUND & FIRST FLOOR APT, KELLY'S LANE LOWER ...	Donegal	85000.000	No	No	Second-Hand Dwelling house /Apartment
560578	2022-10-14	T/A MORIARTY AND MURPHY, APT 2 FLEMING'S LAN...	Kerry	50000.000	No	No	Second-Hand Dwelling house /Apartment
572476	2022-12-13	APARTMENT (FISH & CHIP SHOP), 291 RICHMOND ROA...	Dublin	175000.000	No	No	Second-Hand Dwelling house /Apartment

```
In [55]: error_properties = ['SUPERMARKET CARPARK &APT, THE FAYTHE', 'APT B & D, 14 DOUGHISKA ROAD, DOUGHISKA', 'THE TEMPLE BAR PUB & 9 APARTMENTS, TEMPLE BAR', 'THE RED COTTAGE & APTS, LISNALURG, BUNDORAI', '23A CORK ST, APT1 & 2', 'APARTMENTS 18 AND 18A, HARLEY WOOD, TOGHER', 'GROUND FLOOR APT1 APT2 AND APT3, 50 UPPER BAGGOT ST, DUBLIN 4', 'OFFICES APARTMENT AND CAI', 'APARTMENTS AND SMYTHS PUBLIC HOUSE, 75/77', 'APT1 AND 2, 19 TRINITY ST, WEXFORD', 'APT 16D AND MOORING BERTH, 10 HARBOUR VILLAGE', 'BLOCK AND 7 APTS AT LA TOUCHE PLACE,', 'APT 1 2 3 & 4 BLOCK 3, ROSSMORE NEIGHBOURHOOD CENTRE, TIPPERARY', 'APT 1 2 & 3 BLOCK 5, ROSSMORE NEIGHBOURHOOD CENTRE, TIPPERARY', 'APT 5 6 7 & 8 BLOCK 3, ROSSMORE NEIGHBOURHOOD CENTRE, TIPPERARY', '3 APARTMENTS AND FIRST FLOOR OF, GRATTAN COURT GRATTAN PLACE, GREYSTONES', 'KCR HOUSE &, 324 KIMMAGE RD LOWER, APT 1 KIMMAGE', 'CROUGH\&S PUBLIC HOUSE AND APT 3', 'APARTMENTS ABOVE, 9 & 10 FRIARS WALK, CORK', 'APARTMENT 49A 49B & 49C, ST BRENDANS AVENUE', 'APARTMENTS, 11A 11B AND 13A BARRACK STREET, CITY CENTRE SOUTH', 'APT 418 & 327 BLOCK A, APT 218 216 316 125 221 223, 410 & 320 BLOCK B', 'APT 1 AND NEWCASTLE SERVICE STATION, NEWCASTLE SERVICE STATION, NEWCASTLE', 'APARTMENT BLOCK, AT ARD ALAINN & CAR PARK', 'GARDEN APT. 4GROSVENORTERRACE', 'GARDEN APT HALL APT AND TOP APT, 38 LOWER MOUNT PLEASANT, RATHMINES DUBLIN 6', 'APT NOS. 49A 49B 50A 50D & 50E, FERRYBANK, ARKLOW', 'APT 1 APT2 APT3 & APT4, 1 NEW RD 1 BARROW COURT']

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)

# drop rows
df = df.reset_index(drop=True)
df.shape
```

Out[55]: (592275, 8)

- Look for addresses with keyword 'FOLIOS'.

```
In [56]: df[df['Address'].str.contains('FOLIOS')]
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescrip
33754	2011-10-07	FOLIOS 8284F AND 5054F COUNTY CAVAN, COUNTY CAVAN	Cavan	62500.000	Yes	No	Second-Hand Dwelling house /Apartment	
402295	2019-12-12	LANDS COMPRISED IN, FOLIOS KY24798, AND KY71789F	Kerry	880000.000	No	No	Second-Hand Dwelling house /Apartment	

```
In [57]: rows = df[df['Address'].str.contains('FOLIOS')].index

# drop rows
df = df.drop(index=rows)
df = df.reset_index(drop=True)
df.shape
```

Out[57]: (592273, 8)

- Look for addresses with keyword 'STUDENT'.

```
In [58]: df[(df['Address'].str.contains('STUDENT')) & (df['Price'] >= p_99)]
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySi
126891	2014-11-07	286 STUDENT RESIDENCES, NATIONAL COLLEGE OF IR...	Mayo	7762438.000	Yes	No	Second-Hand Dwelling house /Apartment	
138022	2015-01-09	INTEREST IN 286 STUDENT RESIDENCES, MAYOR STRE...	Mayo	6986194.000	Yes	No	Second-Hand Dwelling house /Apartment	
245817	2017-03-24	MONTROSE STUDENT RESIDENCE, STILLORGAN ROAD, D...	Dublin	33538000.000	No	No	Second-Hand Dwelling house /Apartment	
274720	2017-10-06	1 DUNDALK STUDENT VILLAGE, DUNDALK INSTITUTE O...	Louth	2300000.000	No	No	Second-Hand Dwelling house /Apartment	
287867	2017-12-18	1 MARINO STUDENT RESIDENCE, MARION INSTITUTE O...	Dublin	11225721.000	No	No	Second-Hand Dwelling house /Apartment	
305172	2018-04-24	STUDENT ACCOMODATION KNOWN AS, CUIRT NA COIRIB...	Galway	34781000.000	No	No	Second-Hand Dwelling house /Apartment	
460535	2021-02-15	41 GROVE ISLAND STUDENT VILLAGE, GROVE ISLAND,...	Limerick	1890000.000	No	No	Second-Hand Dwelling house /Apartment	
528689	2022-04-08	STUDENT ACCOMMODATION BUILDING, 35-38 CORK STREET	Dublin	80685721.460	No	No	Second-Hand Dwelling house /Apartment	
528690	2022-04-08	STUDENT ACCOMMODATION BUILDING, 9 11 & 13 BLA...	Dublin	18208921.020	No	No	Second-Hand Dwelling house /Apartment	

```
In [59]: rows = df[(df['Address'].str.contains('STUDENT')) & (df['Price'] >= p_99)].index

# drop rows
df = df.drop(index=rows)
df = df.reset_index(drop=True)
df.shape
```

Out[59]: (592264, 8)

In [60]: df[(df['Address'].str.contains('STUDENT'))]

Out[60]:		DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Property
	48744	2012-06-29	APARTMENT 101 STUDENT VILLAGE, CRATLOE WOOD	Limerick	26500.000		No	No	Second-Hand Dwelling house /Apartment
	48746	2012-06-29	APARTMENT 57 STUDENT VILLAGE, CRATLOE WOOD	Limerick	26500.000		No	No	Second-Hand Dwelling house /Apartment
	48747	2012-06-29	APARTMENT 90 STUDENT VILLAGE, CRATLOE WOOD	Limerick	26500.000		No	No	Second-Hand Dwelling house /Apartment
	48748	2012-06-29	APARTMENT 93 STUDENT VILLAGE, CRATLOE WOOD	Limerick	26500.000		No	No	Second-Hand Dwelling house /Apartment
	48749	2012-06-29	APARTMENT 97 STUDENT VILLAGE, CRATLOE WOOD	Limerick	26500.000		No	No	Second-Hand Dwelling house /Apartment
	64201	2012-12-21	STUDENT RESIDENCES, DUBLIN CITY UNIVERSITY, GL...	Dublin	228976.900		No	No	Second-Hand Dwelling house /Apartment
	64334	2012-12-28	STUDENT RESIDENCES, WATERFORD INSTITUTE OF TEC...	Waterford	699563.900		No	No	Second-Hand Dwelling house /Apartment
	72363	2013-05-21	APT 50, GROODY STUDENT VILLAGE, CASTLETROY	Limerick	80000.000		No	No	Second-Hand Dwelling house /Apartment
	74784	2013-06-21	APT. 40 WELLMOUNT STUDENT VILLAGE, DUBLIN ROAD...	Westmeath	45000.000		No	No	Second-Hand Dwelling house /Apartment
	78718	2013-08-06	32 GORT NA COIRIBE STUDENT VILLAGE, HEADFORD R...	Galway	110000.000		No	No	Second-Hand Dwelling house /Apartment
	87089	2013-11-07	65A THOMOND STUDENT VILLAGE, LIMERICK, LIMERICK	Limerick	25000.000		No	No	Second-Hand Dwelling house /Apartment
	89037	2013-11-27	UNIT 102 GLASAN STUDENT VILLAGE, BALLYBANE RD,...	Galway	78000.000		No	No	Second-Hand Dwelling house /Apartment
	92098	2013-12-18	APT 40A THOMOND STUDENT VILLAGE, BALLYGRENNAN,...	Limerick	25000.000		No	No	Second-Hand Dwelling house /Apartment
	93631	2013-12-20	APT. B3 WELLMOUNT STUDENT VILLAGE, DUBLIN ROAD...	Westmeath	40000.000		No	No	Second-Hand Dwelling house /Apartment
	94051	2013-12-23	APT B7, ST ANGELAS STUDENT RESIDENCES, CLOGHER...	Sligo	65000.000		No	No	Second-Hand Dwelling house /Apartment
	99419	2014-03-12	61 GORT NA COIRIBE, STUDENT VILLAGE, HEADFORD RD	Galway	155000.000		No	No	Second-Hand Dwelling house /Apartment
	99502	2014-03-13	22 A GLASSAN GNIT STUDENT VILLAGE, BALLYBANE ...	Galway	258391.000		No	Yes	New Dwelling house /Apartment
			APT C13, BLOCK C,						greater than 38 sq m

589468	2023-04-05	VILLAGE, BALLYMUN, DUBL...	Dublin	200000.000	No	No	house /Apartment
590952	2023-04-20	88A THOMOND, STUDENT VILLAGE, OLD CRATLOE RD	Limerick	115000.000	No	No	Second-Hand Dwelling house /Apartment
591952	2023-04-28	APARTMENT 52, GATEWAY STUDENT VILLAGE, BALLYMU...	Dublin	215000.000	No	No	Second-Hand Dwelling house /Apartment

```
In [61]: error_properties = ['88 GROODY STUDENT PARK, CASTLETROY', 'STUDENT RESIDENCES, DUBLIN CITY UNIVERSITY',  
'STUDENT RESIDENCES, WATERFORD INSTITUTE', 'APTS 7-16 BLOCK4, BRIAN FRIEL HALL, MCCAFFERY',  
'BLOCK A APT 10, ST ANGELAS COLLEGE']  
  
for address in error_properties:  
    df = df.drop(index=df[df['Address'].str.contains(address)].index)  
  
# drop rows  
df = df.reset_index(drop=True)  
df.shape
```

Out[61]: (592259, 8)

- Recheck New Highest Prices

```
In [62]: # check new highest price  
df[df['Price'] == df['Price'].max()]
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySize
469516	2021-04-15	8TH LOCK, RATOATH ROAD, PELLETSTOWN	Dublin	170142820.000	No	Yes	New Dwelling house /Apartment	

- After removing multiple apartments, we can see that the new highest priced property is not a single residential property. This means there are still non-residential single properties in the data. We'll examine the top 50 properties with regards to price below.

```
In [63]: # show top 50 properties with highest price  
df.nlargest(50, 'Price')
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySize
469516	2021-04-15	8TH LOCK, RATOATH ROAD, PELLETSTOWN	Dublin	170142820.000	No	Yes	New Dwelling house /Apartment	
513889	2021-12-22	ST CLARES PARK, HAROLDS CROSS ROAD	Dublin	107488987.000	No	No	Second-Hand Dwelling house /Apartment	
588125	2023-03-28	OPUS 6 HANOVER QUAY, DUBLIN 2	Dublin	93813416.400	No	No	Second-Hand Dwelling house /Apartment	
372884	2019-06-28	6 HANOVER QUAY, DUBLIN 2	Dublin	84420800.000	No	Yes	New Dwelling house /Apartment	
396135	2019-11-13	FERNBANK, CHURCHTOWN ROAD, DUBLIN 14	Dublin	79306927.000	No	Yes	New Dwelling house /Apartment	
319801	2018-07-27	182 THE ELYSIAN, EGLINTON STREET, CORK	Cork	69873482.000	Yes	No	Second-Hand Dwelling house /Apartment	
233835	2016-12-19	BINARY HUB ROE'S LANE, BONHAM STREET	Dublin	69208163.000	No	No	New Dwelling house /Apartment	greater than 38 sq m
305529	2018-04-27	1 THE ELYSIAN, EGLINTON STREET, CORK	Cork	67700000.000	No	No	Second-Hand Dwelling house /Apartment	
187372	2016-01-15	TALLAGHT CROSS WEST, TALLAGHT	Dublin	64085057.000	No	Yes	New Dwelling house /Apartment	greater than 38 sq m
81491	2013-09-06	CLANCY BARRACKS,	Dublin	61828634.000	No	No	Second-Hand Dwelling	

- After further research, it appears there are no in the highest 50 properties which mainly consist of property developments and apartment complexes. Addresses that appear to be single apartments are actually full buildings of apartments. I'll drop these rows since our focus is single residential properties.

```
In [65]: rows = list(df.nlargest(50, 'Price').index)
```

```
# drop rows
df = df.drop(index=rows)
df = df.reset_index(drop=True)
df.shape
```

```
Out[65]: (592209, 8)
```

```
In [66]: # show top 50 properties with highest price
df.nlargest(50, 'Price')
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Prope
215856	2016-08-31	1 COLDCUT PARK, COLDCUT RD, CLONDALKIN DUBLIN 22	Dublin	18150000.000		No	No	Second-Hand Dwelling house /Apartment
487417	2021-08-12	1 THE CASINO, DUBLIN RD, MALAHIDE	Dublin	17200000.000		No	No	Second-Hand Dwelling house /Apartment
61541	2012-12-12	COLLEGE STREET, BALDOYLE	Dublin	15785185.000		No	No	Second-Hand Dwelling house /Apartment
401818	2019-12-12	1 THE CHESTNUT, PARK AVE, DOUGLAS RD SOUTH	Cork	15507048.000		No	No	Second-Hand Dwelling house /Apartment
196866	2016-04-07	KYLEMORE CLINIC, 49 CHURCH ROAD, BALLYBRACK	Dublin	15450000.000		No	No	Second-Hand Dwelling house /Apartment
138042	2015-01-12	1 VICTORIA MILLS, VICTORIA CROSS, CORK	Cork	15176000.000		No	No	Second-Hand Dwelling house /Apartment
378909	2019-08-01	XAVIER COURT, 21- 25 SHERRARD STREET UPPER, 21 ...	Dublin	15002000.000		No	No	Second-Hand Dwelling house /Apartment
572765	2022-12-14	APT 1, 59 PATRICK STREET, DUN LAOGHAIRE	Dublin	14820000.000		No	No	Second-Hand Dwelling house /Apartment
340208	2018-11-28	ABBEY GLEN, OFF POTTERY RD, CABINTEELY	Dublin	14800000.000		No	No	Second-Hand Dwelling house /Apartment
232827	2016-12-15	WALFORD, SHREWSBURY RD, DUBLIN 4	Dublin	14250000.000		No	No	Second-Hand Dwelling house /Apartment
408290	2020-01-23	HAZELBROOK RISE, RATHNEW	Wicklow	14176211.450		No	Yes	New Dwelling house /Apartment
69260	2013-03-29	WALFORD, SHREWSBURY ROAD, DUBLIN 4	Dublin	14000000.000		No	No	Second-Hand Dwelling house /Apartment
172367	2015-10-02	2 PARKGROVE, CITYQUARTER, ATHLONE TOWN CENTRE	Westmeath	14000000.000		No	No	Second-Hand Dwelling house /Apartment
73467	2013-06-05	ROCHDALE, HONEY PARK, DUN LAOGHAIRE	Dublin	13389846.000		Yes	No	Second-Hand Dwelling house /Apartment
474359	2021-05-19	UIMHI A NAOI, BÓTHA SRIÚSBAIRE	Dublin	13250000.000		No	No	Second-Hand Dwelling house /Apartment
109520	2014-06-25	APT 201 LIFFEY TRUST BUILDING, 117 TO 126 UPPE...	Dublin	12858202.000		Yes	No	Second-Hand Dwelling house /Apartment
397355	2019-11-20	PROPERTY AT CASTLETOWN DEMESNE, CARRICK-ON-SUI...	Kilkenny	12600000.000		No	No	Second-Hand Dwelling house /Apartment
554333	2022-09-12	ST KILDAS SANDYCOVE AVE EAST, SANDYCOVE CO DUBLIN	Dublin	12500000.000		No	No	Second-Hand Dwelling house /Apartment

ROAD

67037	2013-02-21	PARK LODGE, 57 AND 9 NORTH CIRCULAR ROAD, DUBLIN	Dublin	9601711.000	No	No	Second-Hand Dwelling house /Apartment
453546	2020-12-21	2 GREEG COURT, PARNELL STREET, DUBLIN 1	Dublin	9600000.000	No	No	Second-Hand Dwelling house /Apartment
551270	2022-08-26	1 ARDCLOUGH ROAD, CELBRIDGE	Kildare	9552782.000	Yes	Yes	New Dwelling house /Apartment
98223	2014-02-28	1 ALTO VETRO, GRAND CANAL QUAY, DUBLIN 2	Dublin	9515419.000	No	No	Second-Hand Dwelling house /Apartment
260779	2017-07-10	GORSE HILL, VICO RD, KILLINEY	Dublin	9500000.000	No	No	Second-Hand Dwelling house /Apartment
546440	2022-07-27	CEDAR LODGE, 98 MERRION RD, BALLSBRIDGE DUBLIN 4	Dublin	9400000.000	No	No	Second-Hand Dwelling house /Apartment
460116	2021-02-12	189 KILLINARDEN EST, TALLAGHT, DUBLIN	Dublin	9249397.000	No	No	Second-Hand Dwelling house /Apartment
190865	2016-02-17	11 THE LINKS, ELM PARK, MERRION ROAD	Dublin	9173500.000	No	No	Second-Hand Dwelling house /Apartment
298873	2018-03-09	APT 23 BLOCK L TEMPLEGROVE, CASTLETREASURE, DO...	Cork	9115000.000	No	No	Second-Hand Dwelling house /Apartment

In [67]:

```

rows = list(df.nlargest(50, 'Price').index)

# WALFORD, SHREWSBURY RD, DUBLIN 4
rows.remove(df[df['Address'].str.contains('WALFORD, SHREWSBURY RD, DUBLIN 4')].index)
# WALFORD, SHREWSBURY RD, DUBLIN 4
rows.remove(df[df['Address'].str.contains('WALFORD, SHREWSBURY ROAD, DUBLIN 4')].index)
# UIMHI A NAOI, BÓTHA SRIÚSBAIRE
rows.remove(df[df['Address'].str.contains('UIMHI A NAOI, BÓTHA SRIÚSBAIRE')].index)
# APT 201 LIFFEY TRUST BUILDING,
rows.remove(df[df['Address'].str.contains('APT 201 LIFFEY TRUST BUILDING, ')].index)
# ST KILDAS SANDYCOVE AVE EAST, SANDYCOVE CO DUBLIN
rows.remove(df[df['Address'].str.contains('ST KILDAS SANDYCOVE AVE EAST, SANDYCOVE CO DUBLIN')].index)

# 73 AILESBUARY RD, BALLSBRIDGE, DUBLIN 4
address = list(df[df['Address'].str.contains('73 AILESBUARY RD, BALLSBRIDGE, DUBLIN 4')].index)
for a in address:
    try:
        rows.remove(a)
    except:
        pass

# CARRICKBRACK HOUSE, OLD CARRICKBRACK RD, HOWTH
rows.remove(df[df['Address'].str.contains('CARRICKBRACK HOUSE, OLD CARRICKBRACK RD, HOWTH')].index)

# SORRENTO HOUSE, 1 SORRENTO TERRACE, SORRENTO ROAD
address = list(df[df['Address'].str.contains('SORRENTO HOUSE, 1 SORRENTO TERRACE, SORRENTO ROAD')].index)
for a in address:
    try:
        rows.remove(a)
    except:
        pass

# 40 THE CASINO, MALAHIDE, DUBLIN
rows.remove(df[df['Address'].str.contains('40 THE CASINO, MALAHIDE, DUBLIN')].index)

# THE GATE LODGE, SOMERTON, CASTLEKNOCK
rows.remove(569369)

# GORSE HILL, VICO RD, KILLINEY
rows.remove(df[df['Address'].str.contains('GORSE HILL, VICO RD, KILLINEY')].index)

# drop rows
df = df.drop(index=rows)
df = df.reset_index(drop=True)
df.shape

```

Out[67]: (592171, 8)

```
In [68]: # show top 50 properties with highest price
df.nlargest(50, 'Price')
```

		DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertyS
232813	2016-12-15	WALFORD, SHREWSBURY RD, DUBLIN 4	Dublin	14250000.000		No	No	Second-Hand Dwelling house /Apartment	
69257	2013-03-29	WALFORD, SHREWSBURY ROAD, DUBLIN 4	Dublin	14000000.000		No	No	Second-Hand Dwelling house /Apartment	
474328	2021-05-19	UIMHI A NAOI, BÓTHA SRIÚSBAIRE	Dublin	13250000.000		No	No	Second-Hand Dwelling house /Apartment	
109515	2014-06-25	APT 201 LIFFEY TRUST BUILDING, 117 TO 126 UPPE...	Dublin	12858202.000		Yes	No	Second-Hand Dwelling house /Apartment	
554296	2022-09-12	ST KILDAS SANDYCOVE AVE EAST, SANDYCOVE CO DUBLIN	Dublin	12500000.000		No	No	Second-Hand Dwelling house /Apartment	
520115	2022-02-16	73 AILESBUARY RD, BALLSBIDGE, DUBLIN 4	Dublin	11700000.000		No	No	Second-Hand Dwelling house /Apartment	
422280	2020-05-25	CARRICKBRACK HOUSE, OLD CARRICKBRACK RD, HOWTH	Dublin	11500000.000		No	No	Second-Hand Dwelling house /Apartment	
481348	2021-07-02	SORRENTO HOUSE, 1 SORRENTO TERRACE, SORRENTO ROAD	Dublin	10650000.000		No	No	Second-Hand Dwelling house /Apartment	
167910	2015-08-31	SORRENTO HOUSE, 1 SORRENTO TERRACE, SORRENTO ROAD	Dublin	10000000.000		No	No	Second-Hand Dwelling house /Apartment	
487479	2021-08-12	40 THE CASINO, MALAHIDE, DUBLIN	Dublin	9910000.000		No	No	Second-Hand Dwelling house /Apartment	
569332	2022-11-30	THE GATE LODGE, SOMERTON, CASTLEKNOCK	Dublin	9814000.000		No	No	Second-Hand Dwelling house /Apartment	
260764	2017-07-10	GORSE HILL, VICO RD, KILLINEY	Dublin	9500000.000		No	No	Second-Hand Dwelling house /Apartment	
588237	2023-03-29	PARKLANDS, CITY WEST, DUBLIN 24	Dublin	9101322.000		No	Yes	New Dwelling house /Apartment	
314106	2018-06-22	THE GALLERY CITY SQUARE, WATERCOURSE ROAD, BL...	Cork	9096916.300		No	No	Second-Hand Dwelling house /Apartment	
419973	2020-04-27	APARTMENT NO 1, HANBURY MEWS, HANBURY LANE	Dublin	9050000.000		No	No	Second-Hand Dwelling house /Apartment	
133581	2014-12-17	CONTRACT-S31A SDCA 1999, 152/154 PEMBROKE ROAD...	Dublin	9000000.000		No	No	Second-Hand Dwelling house /Apartment	
180414	2015-11-27	KENAH HILL, ST GEORGES AVENUE, KILLINEY	Dublin	9000000.000		No	No	Second-Hand Dwelling house /Apartment	
233816	2016-12-19	CLANCOOL, 15 SHREWSBURY RD, BALLSBIDGE DUBLIN 4	Dublin	8761192.000		No	No	Second-Hand Dwelling house /Apartment	
320087	2018-07-30	1A HAROLD S COURT 23 UNIT, PARNELL ROAD, DUBLI...	Dublin	8750750.000		No	No	Second-Hand Dwelling house /Apartment	
81483	2013-09-06	CAMAC CRESCENT, TURVEY AVENUE, INCHICORE	Dublin	8734134.800		No	No	Second-Hand Dwelling house /Apartment	
201732	2016-05-18	19 TEMPLE RD, DARTRY, DUBLIN 6	Dublin	8725000.000		No	No	Second-Hand Dwelling house /Apartment	
		1 RATHASKER							

BALBUTCHER LANE, DUBLIN 11							house /Apartment
261809	2017-07-17	THE HERMITAGE, CROSS AVE, BLACKROCK	Dublin	7600000.000	No	No	Second-Hand Dwelling house /Apartment
230596	2016-12-06	LORETO HALL, ST STEPHENS GREEN, DUBLIN 2	Dublin	7550000.000	No	No	Second-Hand Dwelling house /Apartment
141231	2015-02-04	STRATHMORE HOUSE AND GATE LODGE, STRATHMORE RO...	Dublin	7500000.000	No	No	Second-Hand Dwelling house /Apartment
324258	2018-08-27	26 COLIEMORE RD, DALKEY, DUBLIN	Dublin	7500000.000	No	No	Second-Hand Dwelling house /Apartment
6923	2010-05-21	BEAUPARK, CLARE HALL, CLONGRIFFIN	Dublin	7333121.000	No	No	Second-Hand Dwelling house /Apartment
518251	2022-02-03	ABINGTON WOOD, SWORDS ROAD, MALAHIDE	Dublin	7250000.000	No	No	Second-Hand Dwelling house /Apartment

- At first glance, there are still multiple properties listed as single sales. I will remove these.

```
In [69]: error_properties = ['PARKLANDS, CITY WEST, DUBLIN 24', 'THE GALLERY CITY SQUARE, WATERCOURSE ROAD, BLACKPOOL', 'CONTRACT-S31A SDCA 1999, 152/154 PEMBROKE ROAD, BALLSBIDGE', 'CAMAC CRESCENT, TURVEY AVENUE']

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)

# drop rows
df = df.reset_index(drop=True)
df.shape
```

Out[69]: (592167, 8)

```
In [70]: # check new most expensive property
df[df['Price'] == df['Price'].max()]
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDe
232811	2016-12-15	WALFORD, SHREWSBURY RD, DUBLIN 4	Dublin	14250000.000	No	No	Second-Hand Dwelling house /Apartment	

```
In [71]: # show min price for top 50 properties
df.nlargest(100, 'Price')['Price'].min()
```

Out[71]: 5900000.0

- After removing most apartment complexes and developments, the 100 most expensive property sales are between €5,854,801 and €14,250,000. The most expensive property appears to be a single apartment. It doesn't seem credible that a single apartment can cost that much, I will further investigate the top 100 most expensive properties.

```
In [72]: # check top 100 most expensive sales
df.nlargest(100, 'Price')
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Property
232811	2016-12-15	WALFORD, SHREWSBURY RD, DUBLIN 4	Dublin	14250000.000	No	No	Second-Hand Dwelling house /Apartment	
69257	2013-03-29	WALFORD, SHREWSBURY ROAD, DUBLIN 4	Dublin	14000000.000	No	No	Second-Hand Dwelling house /Apartment	
474325	2021-05-19	UIMHI A NAOI, BÓTHA SRIÚSBAIRE	Dublin	13250000.000	No	No	Second-Hand Dwelling house /Apartment	
109514	2014-06-25	APT 201 LIFFEY TRUST BUILDING, 117 TO 126 UPPE...	Dublin	12858202.000	Yes	No	Second-Hand Dwelling house /Apartment	

386201	2019-09-16	AUBURN HOUSE, MALAHIDE, RIGHT OF WAY AT DN226834F	Dublin	6000000.000	No	No	Second-Hand Dwelling house /Apartment
406803	2020-01-10	APARTMENT 3, GROVE COURT APARTMENTS, BLANCHARD...	Dublin	5961000.000	No	Yes	New Dwelling house /Apartment
207605	2016-07-01	1 SHEARES GATE, GLASHEEN RD, CORK	Cork	5950000.000	No	No	Second-Hand Dwelling house /Apartment
224755	2016-10-28	51 AILESBUY RD, DUBLIN 4, DUBLIN	Dublin	5900000.000	No	No	Second-Hand Dwelling house /Apartment
556707	2022-09-26	CARMELITE MONASTERY, FIRHOUSE RD, DUBLIN 24	Dublin	5900000.000	No	No	Second-Hand Dwelling house /Apartment

I used Google to find information on each of these properties. After further research, I have concluded that there are some errors in the 100 most expensive property listings:

- WALFORD, SHREWSBURY RD, DUBLIN 4: The price for this property appears to be correct.
- WALFORD, SHREWSBURY ROAD, DUBLIN 4: The price for this property appears to be correct.
- UIMHI A NAOI, BÓTHA SRIÚSBAIRE: The price for this property appears to be correct.
- APT 201 LIFFEY TRUST BUILDING, 117 TO 126 UPPER SHERIFF STREET: The price for this property appears to be an error. It does not seem in line with the price of other apartments in the building.
- ST KILDAS SANDYCOVE AVE EAST, SANDYCOVE CO DUBLIN: The price for this property appears to be correct.
- 73 AILESBUY RD, BALLSBIDGE, DUBLIN 4: The price for this property appears to be correct.
- CARRICKBRACK HOUSE, OLD CARRICKBRACK RD, HOWTH: The price for this property appears to be correct.
- SORRENTO HOUSE, 1 SORRENTO TERRACE, SORRENTO ROAD: The price for this property appears to be correct.
- 40 THE CASINO, MALAHIDE, DUBLIN: The price for this property appears to be an error. It is more than 10 times the price of other apartments in the same building. I will drop this row.
- THE GATE LODGE, SOMERTON, CASTLEKNOCK: Unclear if this is an error. It was sold for €450,000 in 2016. I doubt it has gained that much value in 6 years and so I will drop it just in case.
- GORSE HILL, VICO RD, KILLINEY: The price appears to be correct.
- APARTMENT NO 1, HANBURY MEWS, HANBURY LANE: I believe the full building was sold and not just a single apartment. I will drop this row.
- KENAH HILL, ST GEORGES AVENUE, KILLINEY: The price appears to be correct.
- CLANCOOL, 15 SHREWSBURY RD, BALLSBIDGE DUBLIN 4: The price appears to be correct.
- 1A HAROLD S COURT 23 UNIT, PARNELL ROAD: I believe the full building was sold and not just a single apartment. I will drop this row.
- 19 TEMPLE RD, DARTRY, DUBLIN 6: The price appears to be correct.
- 1 RATHASKER COURT, WHITEWATER PLAZA, NEWBRIDGE: Apartment complex. I will drop this row.
- APT. 1 BLOCK A, BLACKHALL PLACE, ARBOUR HILL: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- SUMMERHILL, MARINO AVE WEST, KILLINEY: The price appears to be correct.
- APT 23 BLOCK L TEMPLEGROVE, CASTLETREASURE: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- FINTRAGH, 11 SHREWSBURY ROAD, DUBLIN 4: The price appears to be correct.
- APT 1, LEESIDE APTS, GRATTAN ST: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- 1 RATHASKER COURT, WHITEWATER PLAZA, NEWBRIDGE: Apartment complex. I will drop this row.
- GLENHEATHER, AVOCAS AVE, BLACKROCK: The price appears to be correct.

- DANES HOLLOW, THORMANBY RD, HOWTH: The price appears to be correct.
- CLONMORE, 15A SHREWSBURY RD, BALLSBIDGE DUBLIN 4: The price appears to be correct.
- MILNER'S SQUARE, SHANOWEN ROAD, SANTRY: Apartment complex. I will drop this row.
- ABILENE HOUSE, NEWTOWNPARK AVENUE, BLACKROCK: The price appears to be correct.
- APT 1, MONTPELIER SQ, 57 63 MONTPELIER HILL: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- DEEPWELL, ROCK HILL, BLACKROCK: The price appears to be correct.
- 10 THE CRESCENT, OLD FORT RD, BALLINCOLLIG: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- OLD QUARTER, BALLINCOLLIG: Residential development. I will drop this row.
- INNISCORRIG, COLIMORE ROAD, DALKEY: The price appears to be correct.
- APARTMENT NO 26, CASTLEFORBES SQUARE, DUBLIN 1: I believe the price for this property is an error. This apartment in was sold in 2017 for €208,000. I will drop this row.
- THE PLAZA CITY SQUARE, WATERCOURSE ROAD: I believe the price for this property is an error. I will drop this row.
- 1 ALEXANDER COURT, 25 UPPER PEMBROKE ST, DUBLIN 2: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- ANNAVILLE RESIDENCE, DUNDRUM ROAD: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- OLD QUARTER, BALLINCOLLIG: Residential development. I will drop this row.
- APT 35 PARKVIEW CLOSE, BALBUTCHER LANE, DUBLIN 11: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- THE HERMITAGE, CROSS AVE, BLACKROCK: The price appears to be correct.
- LORETO HALL, ST STEPHENS GREEN, DUBLIN 2: The price appears to be correct.
- STRATHMORE HOUSE AND GATE LODGE: The price appears to be correct.
- 26 COLIEMORE RD, DALKEY, DUBLIN: The price appears to be correct.
- BEAUPARK, CLARE HALL, CLONGRIGGIN: Residential development. I will drop this row.
- ABINGTON WOOD, SWORDS ROAD, MALAHIDE: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- HUMEWOOD CASTLE, KILTEGAN: The price appears to be correct.
- 1 BUTTERGATE, DONORE RD, DROGHEDA: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- MORNINGTON HOUSE, SANDYCOVE POINT, SANDYCOVE: The price appears to be correct.
- 86 UPPER LEESON ST, DUBLIN 4, DUBLIN: The price appears to be correct.
- 39 GORT NA GLAISE, KILMONEY ROAD, CARRIGALINE: Price appears to be an error. I will drop this row.
- 14 SLI NA CHAIL, LEIXLIP, KILDARE: Price appears to be an error. I will drop this row.
- GAS YARD LANE, MALAHIDE: Not residential property. I will drop this row.
- DERRYMORE, 6 SHREWSBURY ROAD, BALLSBIDGE: The price appears to be correct.
- TOWER GREEN AND CLAREVILLE, CROSS AVENUE: Multiple properties. I will drop this row.
- SEAFIELD HOUSE, KILCREA, DONABATE: The price appears to be correct.
- APARTMENT 1A, FLORAVILLE LODGE, MONASTERY ROAD: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- PART OF THE PROPERTY COMPRISED IN, FOLIO DN184: No address I will drop this row.
- 8 HALCYON PLACE, THE PARK RESIDENTIAL VILLAGE: I believe multiple apartments were sold and not just a single apartment. I

will drop this row.

- 22 APARTMENT AT CLORAGH MILLS, EDMONDSTOWN RD: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- 21 PENNEFEATHER GREEN, HEBRON ROAD: I believe multiple properties were sold and not just a single property. I will drop this row.
- APARTMENT 115, LONGBOAT QUAY NORTH, DUBLIN2: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- APT 8, THE GARDEN HOUSE, WATERFALL AVE: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- 49 AILESURY RD, BALLSBIDGE, DUBLIN 4: The price appears to be correct.
- DUNLIN HOUSE, REDCOURT OAKS, SEAFIELD RD EAST: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- 1 THE HALL, YEW TREE SQ, PROSPEROUS RD: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- SUMMERHILL, MARINO AVE WEST, KILLINEY: The price appears to be correct.
- NERANO RD, NERANO HOUSE, DALKEY: The price appears to be correct.
- MOUNT ALVERNO, SORRENTO RD, DALKEY: The price appears to be correct.
- 2 DUNLIN HOUSE, REDCOURT OAKS, SEAFIELD RD EAST: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- 2 SHREWSBURY ROAD, BALLSBIDGE: The price appears to be correct.
- 5 TEMPLE GARDENS, RATHMINES, DUBLIN 6: The price appears to be correct.
- RATHMORE, ADELAIDE RD, GLENAGEARY: The price appears to be correct.
- HATHERTON, RICHMOND AVENUE SOUTH, DARTRY: The price appears to be correct.
- 8 MARLBOROUGH RD, GLENAGEARY, DUBLIN: The price appears to be correct.
- BELFIELD ABBEY, BOREENMANNA ROAD, CORK: I believe multiple properties were sold and not just a single property. I will drop this row.
- 1A HAZELWOOD APTS, NORTHWOOD, SANTRY: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- PAIRC NA GREINE, COOKSTOWN, TALLAGHT: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- PRIORSGATE, GREENHILLS ROAD, TALLAGHT: I believe multiple apartments were sold and not just a single apartment. I will drop this row.
- BOYSTOWN HOUSE AND ESTATE, BLESSINGTON: The price appears to be correct.
- MINORE, 23 ST KEVINS PARK, DARTRY: The price appears to be correct.
- THE LAURELS, 54 INCHICORE ROAD, KILMAINHAM: Seems to be a development site of multiple properties. I will drop this row.
- ALEXANDER COURT, 25 UPPER PEMBROKE ST, DUBLIN 2: 5 storey apartment block. I will drop this row.
- COURTTOWN HOUSE, COURTTOWN STUD, KILCOCK: The price appears to be correct.
- ELLERSLIE, 46 TEMPLE RD, DARTRY DUBLIN 6: The price appears to be correct.
- LANCASTER GATE, LANCASTER QUAY: Residential apartment development. I will drop this row.
- WILTON, 16 SHREWSBURY RD, BALLSBIDGE DUBLIN 4: The price appears to be correct.
- SORRENTO, NERANO RD, DALKEY: The price appears to be correct.
- 28 AN DUICHE, SCALLAGHEEN, TIPPERARY TOWN: Price appears to be an error. I will drop this row.
- 2 CASTLE PLACE, BREFFNI RD, SANDYCOVE: The price appears to be correct.
- SPICERS MILL, BALBRIGGAN, DUBLIN: Full apartment block. I will drop this row.

- SEAFIELD ROAD, BLACKROCK: Not a single property. I will drop this row.
- 'LISSADELL', 9 SHREWSBURY ROAD, BALLSBIDGE: The price appears to be correct.
- 73 AILESBUY RD, BALLSBIDGE, DUBLIN 4: The price appears to be correct.
- BEULAH, HARBOUR RD, DALKEY: The price appears to be correct.
- AUBURN HOUSE, MALAHIDE, RIGHT OF WAY AT DN226834F: The price appears to be correct.
- APARTMENT 3, GROVE COURT APARTMENTS, BLANCHARDSTOWN: Seems to be multiple apartments. I will drop this row.
- 1 SHEARES GATE, GLASHEEN RD, CORK: Student accomodation. I will drop this row.
- 51 AILESBUY RD, DUBLIN 4, DUBLIN: The price appears to be correct.
- CARMELITE MONASTERY, FIRHOUSE RD, DUBLIN 24: Not a residential property. I will drop this row.
- 16 BLOCK B, BRUACH NA HABHAINN, ALTAMOUNT ST: Seems to be multiple properties. I will drop this row.

In [73]:

```
error_properties = [
    'APT 201 LIFFEY TRUST BUILDING, 117 TO 126 UPPER SHERIFF STREET',
    '40 THE CASINO, MALAHIDE, DUBLIN', 'APARTMENT NO 1, HANBURY MEWS, HANBURY LANE',
    '1A HAROLD S COURT 23 UNIT, PARNELL ROAD', '1 RATHASKER COURT, WHITEWATER PLAZA, NEWBRIDGE',
    'APT. 1 BLOCK A, BLACKHALL PLACE, ARBOUR HILL', 'APT 23 BLOCK L TEMPLEGROVE, CASTLETREASURE',
    'APT 1, LEESIDE APTS, GRATTAN ST', 'MILNER\'S SQUARE, SHANOWEN ROAD, SANTRY',
    'APT 1, MONTPELIER SQ, 57 63 MONTPELIER HILL', '10 THE CRESCENT, OLD FORT RD, BALLINCOLLIG',
    'APT 35 PARKVIEW CLOSE, BALBUTCHER LANE, DUBLIN 11', 'APARTMENT NO 26, CASTLEFORBES SQUARE',
    '1 ALEXANDER COURT, 25 UPPER PEMBROKE ST, DUBLIN 2', 'ANNAVILLE RESIDENCE, DUNDRUM ROAD',
    'BEAUPARK, CLARE HALL, CLONGRIFFIN', 'ABINGTON WOOD, SWORDS ROAD, MALAHIDE', '1 BUTTERGATE',
    '39 GORT NA GLAISE, KILMONEY ROAD, CARRIGALINE', '14 SLI NA CHAIL, LEIXLIP, KILDARE',
    'TOWER GREEN AND CLAREVILLE, CROSS AVENUE', 'APARTMENT 1A, FLORAVILLE LODGE, MONASTERY ROAD',
    'PART OF THE PROPERTY COMPRISED IN, FOLIO DN184', '8 HALCYON PLACE, THE PARK RESIDENTIAL VILLAGE',
    '22 APARTMENT AT CLORAGH MILLS, EDMONDSTOWN RD', '21 PENNEFEATHER GREEN, HEBRON ROAD',
    'APARTMENT 115, LONGBOAT QUAY NORTH, DUBLIN2', 'APT 8, THE GARDEN HOUSE, WATERFALL AVE',
    'DUNLIN HOUSE, REDCOURT OAKS, SEAFIELD RD EAST', 'PAIRC NA GREINE, COOKSTOWN, TALLAGHT',
    'PRIORS GATE, GREENHILLS ROAD, TALLAGHT', 'THE LAURELS, 54 INCHICORE ROAD, KILMAINHAM',
    'ALEXANDER COURT, 25 UPPER PEMBROKE ST, DUBLIN 2', '28 AN DUICHE, SCALLAGHEEN, TIPPERARY TOWNS',
    'APARTMENT 3, GROVE COURT APARTMENTS, BLANCHARDSTOWN', '1 SHEARES GATE, GLASHEEN RD, CORK',
    'CARMELITE MONASTERY, FIRHOUSE RD, DUBLIN 24', '16 BLOCK B, BRUACH NA HABHAINN, ALTAMOUNT ST'

    # '1 ARDCLOUGH ROAD, CELBRIDGE', '1 ALTO VETRO, GRAND CANAL QUAY, DUBLIN 2',
    # 'CEDAR LODGE, 98 MERRION RD, BALLSBIDGE DUBLIN 4', '11 THE LINKS, ELM PARK, MERRION ROAD',
    # 'THE GALLERY CITY SQUARE, WATERCOURSE ROAD, BLACKPOOL', '1A HAROLD S COURT 23 UNIT, PARNELL ROAD',
    # '1 RATHASKER COURT, WHITEWATER PLAZA, NEWBRIDGE', 'MILNER\'S SQUARE, SHANOWEN ROAD, SANTRY',
    # '10 THE CRESCENT, OLD FORT RD, BALLINCOLLIG', 'BEAUPARK, CLARE HALL, CLONGRIFFIN',
    # 'THE PLAZA CITY SQUARE, WATERCOURSE ROAD, ', '1 ALEXANDER COURT, 25 UPPER PEMBROKE ST, DUBLIN 2',
    # 'ANNAVILLE RESIDENCE, DUNDRUM ROAD', 'ABINGTON WOOD, SWORDS ROAD, MALAHIDE', '1 BUTTERGATE',
    # '39 GORT NA GLAISE, KILMONEY ROAD, CARRIGALINE', '14 SLI NA CHAIL, LEIXLIP, KILDARE',
    # 'TOWER GREEN AND CLAREVILLE, CROSS AVENUE', 'PART OF THE PROPERTY COMPRISED IN, FOLIO DN184',
    # '8 HALCYON PLACE, THE PARK RESIDENTIAL VILLAGE', '21 PENNEFEATHER GREEN, HEBRON ROAD',
    # 'DUNLIN HOUSE, REDCOURT OAKS, SEAFIELD RD EAST', 'PAIRC NA GREINE, COOKSTOWN, TALLAGHT',
    # 'PRIORS GATE, GREENHILLS ROAD, TALLAGHT', '1A HAROLD S COURT 23 UNIT, PARNELL ROAD,',
    # 'PART OF THE PROPERTY COMPRISED IN, FOLIO DN184', 'BELFIELD ABBEY, BOREENMANNA ROAD, CORK',
    # 'APT 1 RICHMOND CHESHIRE HOUSE, RICHMOND PARK']

# 'THE GATE LODGE, SOMERTON, CASTLEKNOCK' & '1A HAZELWOOD APTS, NORTHWOOD, SANTRY'
# SEAFIELD ROAD, BLACKROCK
rows = [569329, 263036, 45575]
df = df.drop(index=rows)
df = df.drop(index=df[df['Address'] == ('OLD QUARTER, BALLINCOLLIG')].index)
df = df.drop(index=df[df['Address'] == ('GAS YARD LANE, MALAHIDE')].index)
df = df.drop(index=df[(df['Address'].str.contains('1 THE HALL, YEW TREE SQ, PROSPEROUS RD')) & (df['Price'] >= p_99)].index)
df = df.drop(index=df[df['Address'] == ('BELFIELD ABBEY, BOREENMANNA ROAD, CORK')].index)
df = df.drop(index=df[df['Address'] == ('LANCASTER GATE, LANCASTER QUAY')].index)
df = df.drop(index=df[df['Address'] == ('SPICERS MILL, BALBRIGGAN, DUBLIN')].index)

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)

# # drop rows
df = df.reset_index(drop=True)
df.shape
```

Out[73]: (592111, 8)

- There seems to be a common theme with single apartments listed as being sold for millions, many are actually multiple apartments sold under the one listing. After investigated the priciest properties, there are some legitimately priced properties and so it makes sense to keep these. However, I will remove any apartment listings within the top percent (any apartment with a price over

€1,275,000).

```
In [74]: # find apartment sales in the top 1%
df[((df['Address'].str.contains('APT ') & (~df['Address'].str.contains(' APT')))) | df['Address'].str.contains(
```

Out[74]:	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertyType
	26038	APARTMENT 3 MERRION HOTEL, 21/24 UPPER MERRION...	Dublin	2650000.000		No	No	Second-Hand Dwelling house /Apartment
	105275	APT 504 STEPHENS HALL, EARLSFORT CENTRE, LR LE...	Dublin	3200000.000		No	No	Second-Hand Dwelling house /Apartment
	112468	APTMENT 61 BALLSBRIDGE WOOD, SHELBOURNE ROAD...	Dublin	1520000.000		No	No	Second-Hand Dwelling house /Apartment
	133513	APT 1 DURROW MEWS, FLEMMING PLACE, DUBLIN 4	Dublin	1758000.000		No	No	Second-Hand Dwelling house /Apartment
	133529	APT 26, RYE ABBEY, KILCOCK	Kildare	1561000.000		No	No	Second-Hand Dwelling house /Apartment
	138964	APT 12 REXDON COURT, SANDYMOUNT AVENUE, BALLSB...	Dublin	1275000.000		No	No	Second-Hand Dwelling house /Apartment
	139506	APT 54, GREENHILLS COURT, GREENHILLS RD DUBLIN 24	Dublin	1900000.000		No	No	Second-Hand Dwelling house /Apartment
	140701	APT 12, 28 NORTH FREDERICK ST, DUBLIN 1	Dublin	1535500.000		No	No	Second-Hand Dwelling house /Apartment
	158787	APT 11 FOUR SEASONS HOTEL, BALLSBRIDGE, DUBLIN 4	Dublin	1400000.000		No	No	Second-Hand Dwelling house /Apartment
	177333	APARTMENT 1, 130 PEARSE STREET, DUBLIN 2	Dublin	1635000.000		No	No	Second-Hand Dwelling house /Apartment
	182309	APT 4, THORNFIELD, STILLORGAN RD DUBLIN 4	Dublin	2125000.000		No	No	Second-Hand Dwelling house /Apartment
	194029	APT 10, 16 GARDINER PLACE, DUBLIN 11	Dublin	1300000.000		No	No	Second-Hand Dwelling house /Apartment
	194031	APT 1, 16 GARDINER PLACE, DUBLIN 1	Dublin	1300000.000		No	No	Second-Hand Dwelling house /Apartment
	194033	APT 2, 16 GARDINER PLACE, DUBLIN 11	Dublin	1300000.000		No	No	Second-Hand Dwelling house /Apartment
	194036	APT 3, 16 GARDINER PLACE, DUBLIN 11	Dublin	1300000.000		No	No	Second-Hand Dwelling house /Apartment
	194038	APT 4, 16 GARDINER PLACE, DUBLIN 11	Dublin	1300000.000		No	No	Second-Hand Dwelling house /Apartment
	194039	APT 5, 16 GARDINER PLACE, DUBLIN 11	Dublin	1300000.000		No	No	Second-Hand Dwelling house /Apartment

581665	2023-02-16	CABRA ROAD, DUBLIN 7	Dublin	2800000.000	No	No	house /Apartment
586622	2023-03-20	HALL FLOOR APARTMENT, 40 LEESON PARK, RANELAGH	Dublin	2020000.000	No	No	Second-Hand Dwelling house /Apartment
591801	2023-04-28	APT 1 - 59A INCHICORE ROAD, DUBLIN 8	Dublin	1800000.000	No	No	Second-Hand Dwelling house /Apartment

- Some apartments in the same building were bought at the same price, on the same date. I would assume that these were all in the same transaction and the price reflects the multiple apartments e.g THE PINES, HERBERT PARK LANE DUBLIN 4. Other apartments are multi-purpose units and so I will drop those.

- After researching these properties, it appears that the only single apartments in this list are penthouse apartments. They appear to be exclusive to areas around Dublin. I will filter these out and drop the rest.

```
In [75]: areas = ['MONKSTOWN', 'BALLSBIDGE', 'AILESBURY', 'KNAPTON', 'DUBLIN 4', 'DALKEY', 'LEESON', 'BAGGOT', 'HOTEL', 'DUBLIN 2', 'KILLINEY', 'HANOVER QUAY', 'TRIMBLESTON', 'GRAFTON ST', 'DUBLIN2', 'BLACKROCK', 'ANGLESEA', 'MERRION', 'DONNYBROOK', 'HOWTH', 'MARIANELLA', 'DUN LAOGHAIRE', 'MARIANELLA', 'RATHMINES', 'NORTH CIRCULAR ROAD', 'SOUTH CIRCULAR ROAD', 'DELGANY', 'RANELAGH', 'RATHGAR', 'THE WATER FRONT, HANOV 'DUBLIN 6']

apts = df[(df['Address'].str.contains('APT') | df['Address'].str.contains('APARTMENT')) & (df['Price'] >= p_99)]
error_properties = []

# Manually keep correct properties
# THE ASGARD, APARTMENT NO. 13 ANGLESEA,
apts = apts.drop(index=df[df['Address'].str.contains('THE ASGARD, APARTMENT NO. 13 ANGLESEA,')].index)
# APARTMENT 26, DUNLUCE, 21 ANGLESEA ROAD
apts = apts.drop(index=df[df['Address'].str.contains('APARTMENT 26, DUNLUCE, 21 ANGLESEA ROAD')].index)
# 'BASEMENT APT, JASONVILLE HOUSE, 3 BRIGHTON TERRACE'
apts = apts.drop(index=df[df['Address'].str.contains('BASEMENT APT, JASONVILLE HOUSE, 3 BRIGHTON TERRACE')].index)
# APARTMENT 14, CASTLE COURT, 5-6 CASTLE COURT
apts = apts.drop(index=df[df['Address'].str.contains('APARTMENT 14, CASTLE COURT, 5-6 CASTLE COURT')].index)
# APARTMENT 4 WESTFIELD, BLOOMFIELD HOUSE, BLOOMFIELD AVENUE
apts = apts.drop(index=df[df['Address'].str.contains('APARTMENT 4 WESTFIELD, BLOOMFIELD HOUSE, BLOOMFIELD AVENUE')].index)
# '6 VICTORIA ST, SOUTH CIRCULAR RD, APARTMENT 7'
apts = apts.drop(index=468661)
# APARTMENT 9 THE IVORY BUILDING, HANOVER STREET EAST, SIR JOHN ROGERSONS QUAY
apts = apts.drop(index=df[df['Address'].str.contains('APARTMENT 9 THE IVORY BUILDING, HANOVER STREET EAST, SIR . # APT 1, 9 HATCH STREET LOWER, DUBLIN 2
apts = apts.drop(index=df[df['Address'].str.contains('APT 1, 9 HATCH STREET LOWER, DUBLIN 2')].index)
# APT 25, BLOCK 1, GALLERY QUAY
apts = apts.drop(index=df[df['Address'].str.contains('APT 25, BLOCK 1, GALLERY QUAY')].index)
# APT 1, 81 LOWER GARDINER ST, DUBLIN 1
apts = apts.drop(index=df[df['Address'].str.contains('APT 1, 81 LOWER GARDINER ST, DUBLIN 1')].index)
# APT 1, 138A SANDFORD RD, DUBLIN
apts = apts.drop(index=df[df['Address'].str.contains('APT 1, 138A SANDFORD RD, DUBLIN')].index)
# APT 1 COPPER BEACH, 16 HOLLYBROOK PARK, CLONTARF
apts = apts.drop(index=df[df['Address'].str.contains('APT 1 COPPER BEACH, 16 HOLLYBROOK PARK, CLONTARF')].index)
# APT 1, 46A NEW CABRA RD, DUBLIN 7
apts = apts.drop(index=df[df['Address'].str.contains('APT 1, 46A NEW CABRA RD, DUBLIN 7')].index)
# APT 1, 109 CORK ST, DUBLIN 8
apts = apts.drop(index=df[df['Address'].str.contains('APT 1, 109 CORK ST, DUBLIN 8')].index)
# APT D, 116 GRAFTON ST, DUBLIN
apts = apts.drop(index=df[df['Address'].str.contains('APT D, 116 GRAFTON ST, DUBLIN')].index)
# APT 1, 19 NOTTINGHAM ST, DUBLIN 3
apts = apts.drop(index=547072)
# APT 1 - 59A INCHICORE ROAD, DUBLIN 8
apts = apts.drop(index=df[df['Address'].str.contains('APT 1 - 59A INCHICORE ROAD, DUBLIN 8')].index)

for col, row in apts.iterrows():
    for area in areas:
        if area in row.Address:
            if row.Address not in error_properties:
                error_properties += [row.Address]

# Manually drop error properties
# APT 504 STEPHENS HALL, EARLSFORT
df = df.drop(index=105275)
# APARTMENT 11 THE BROKERAGE, TOWNSEND STREET
df = df.drop(index=df[df['Address'].str.contains('APARTMENT 11 THE BROKERAGE, TOWNSEND STREET,')].index)
# APT 1, 11 TO 12 SUNBURY GARDENS, DARTRY RD
df = df.drop(index=df[df['Address'].str.contains('APT 1, 11 TO 12 SUNBURY GARDENS, DARTRY RD')].index)

# APT 12, THE PINES, HERBERT PARK LANE DUBLIN 4
```

```

df = df.drop(index=213162)
# APT 43, THE PINES, HERBERT PARK LANE DUBLIN 4
df = df.drop(index=213169)
# APT 49, THE PINES, HERBERT PARK LANE DUBLIN 4
df = df.drop(index=213170)
# APT 5, THE PINES, HERBERT PARK LANE DUBLIN 4
df = df.drop(index=213171)
# south richmond st buildings
richmond_st = df[(df['Address'].str.contains('RICHMOND ST')) & (df['Price'] >= p_99)].index
df = df.drop(index=richmond_st)

for col, row in appts.iterrows():
    if row.Address not in error_properties:
        df = df.drop(index=df[df['Address'] == (row.Address)].index)

df = df.reset_index(drop=True)
df.shape

```

Out[75]: (591904, 8)

In [76]: # check 200 most expensive sales
df.nlargest(200, 'Price')

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Pro
232766	2016-12-15	WALFORD, SHREWSBURY RD, DUBLIN 4	Dublin	14250000.000		No	Second-Hand Dwelling house /Apartment	
69254	2013-03-29	WALFORD, SHREWSBURY ROAD, DUBLIN 4	Dublin	14000000.000		No	Second-Hand Dwelling house /Apartment	
474118	2021-05-19	UIMHI A NAOI, BÓTHA SRIÚSBAIRE	Dublin	13250000.000		No	Second-Hand Dwelling house /Apartment	
554051	2022-09-12	ST KILDAS SANDYCOVE AVE EAST, SANDYCOVE CO DUBLIN	Dublin	12500000.000		No	Second-Hand Dwelling house /Apartment	
519886	2022-02-16	73 AILESBUARY RD, BALLSBRIDGE, DUBLIN 4	Dublin	11700000.000		No	Second-Hand Dwelling house /Apartment	
422098	2020-05-25	CARRICKBRACK HOUSE, OLD CARRICKBRACK RD, HOWTH	Dublin	11500000.000		No	Second-Hand Dwelling house /Apartment	
481136	2021-07-02	SORRENTO HOUSE, 1 SORRENTO TERRACE, SORRENTO ROAD	Dublin	10650000.000		No	Second-Hand Dwelling house /Apartment	
167895	2015-08-31	SORRENTO HOUSE, 1 SORRENTO TERRACE, SORRENTO ROAD	Dublin	10000000.000		No	Second-Hand Dwelling house /Apartment	
260668	2017-07-10	GORSE HILL, VICO RD, KILLINEY	Dublin	9500000.000		No	Second-Hand Dwelling house /Apartment	
180395	2015-11-27	KENAH HILL, ST GEORGES AVENUE, KILLINEY	Dublin	9000000.000		No	Second-Hand Dwelling house /Apartment	
233768	2016-12-19	CLANCOOL, 15 SHREWSBURY RD, BALLSBRIDGE DUBLIN 4	Dublin	8761192.000		No	Second-Hand Dwelling house /Apartment	
201697	2016-05-18	19 TEMPLE RD, DARTRY, DUBLIN 6	Dublin	8725000.000		No	Second-Hand Dwelling house /Apartment	
421695	2020-05-20	SUMMERHILL, MARINO AVE WEST, KILLINEY	Dublin	8500000.000		No	Second-Hand Dwelling house /Apartment	
248799	2017-04-13	FINTRAGH, 11 SHREWSBURY ROAD, DUBLIN 4	Dublin	8450000.000		No	Second-Hand Dwelling house /Apartment	
479895	2021-06-25	GLENHEATHER, AVOCA AVE, BLACKROCK	Dublin	8220000.000		No	Second-Hand Dwelling house /Apartment	
270343	2017-09-08	DANES HOLLOW, THORMANBY RD, HOWTH	Dublin	8200000.000		No	Second-Hand Dwelling house /Apartment	
		CLONMORE, 15A SHREWSBURY RD,					Second-Hand Dwelling	

454861	2020-12-23	TUDOR HALL, THE HILL, MONKSTOWN	Dublin	4200000.000	No	No	Second-Hand Dwelling house /Apartment
466322	2021-03-26	1 COLIEMORE VILLAS, DALKEY, DUBLIN	Dublin	4200000.000	No	No	Second-Hand Dwelling house /Apartment
571439	2022-12-09	WOODFIELD MANOR, KILLELINE, NEWCASTLE WEST	Limerick	4191394.000	Yes	No	New Dwelling house /Apartment
88875	2013-11-26	ARENA CENTRE, TALLAGHT	Dublin	4180000.000	No	No	Second-Hand Dwelling house /Apartment
366060	2019-05-17	6 CLYDE RD, BALLBRIDGE, DUBLIN CITY	Dublin	4164000.000	No	No	Second-Hand Dwelling house /Apartment
81518	2013-09-06	THE PRIORY, MONKSTOWN ROAD, MONKSTOWN	Dublin	4155000.000	No	No	Second-Hand Dwelling house /Apartment

```
In [77]: error_properties = ['39A, MOUNTJOY STREET', 'PARKLANDS PLACE PARKLANDS PARADE', 'MARIST CONVENT LANDS, TUBBERCURRY', '38 ELDERWOOD PARK BOREENMANNA RD, CORK', '1 MILFORD MANOR GREEN, CLONDALKIN, DUBLIN 22', 'THE GREY GABLES, 34A GILFORD ROAD, SANDYMOULY', 'LOUGHION GREEN, LOUGHION', 'MAYNOOTH CAMPUS, MAYNOOTH', 'THE LAURELS, MAIN ST, TALLAGHT', 'CASTLETOWN HOUSE, CASTLETOWN DEMESNE', '1A 2AA YLMER DRIVE 5-6AYLMER GREEN', 'DUNEDIN TERRACE, MONKSTOWN FARM, DUN LAOGHAIRE', 'DALKEY MANOR & DALKEY LODGE, BARNHILL ROAD', 'SITE AT NEW ROAD, ENFIELD', 'BARRETTTS, DUBLIN', '37 CHURCH VIEW, BOYLE, CO ROSCOMMON', 'FLAG ALLEY, 58/59 MEATH STREET', 'VERITAS HOUSE, DOMINICAN CONVENT, MUCKROSS PARK', 'WILLOWVIEW, SOMERTON, BALLYBODEN ROAD', 'LANDS COMPRISED IN, FOLIO GY123014F', 'LANDS AT MILL LANE, OFF THE OLD NANGOR ROAD', 'ABINGTON WOOD, MALAHIDE, CO DUBLIN', '1, MILFORD MANOR WALK, CLONDALKIN', 'AUBURN LODGE, JOHNSTOWN ROAD, COUNTY DUBLIN', 'CORBALTON HALL, CORBALTON, TARA', '1 HALCYON PLACE, THE PARK, CASTLETROY', '2 BLUE COAT MEWS, KEVIN ST, CORK', '51 WILLOW AVENUE, PRIMROSE WAY, CELBRIDGE', 'ARENA CENTRE, TALLAGHT']
]

df = df.drop(index=df[df['Address'] == ('LANCASTER GATE, LANCASTER QUAY')].index)
df = df.drop(index=df[df['Address'] == ('FOUR SEASONS HOTEL, BALLSBRIDGE, DUBLIN')].index)
df = df.drop(index=df[df['Address'] == ('HEELEY'S VIEW, THE MALL, MALAHIDE')].index)
df = df.drop(index=df[df['Address'] == ('FOTA ISLAND RESORT, FOTA ISLAND')].index)
df = df.drop(index=df[df['Address'] == ('2 CLORAGH MILLS, EDMONDSTOWN RD, RATHFARNHAM DUBLIN 16')].index)
df = df.drop(index=df[df['Address'] == ('BARROWHALL, MOUNTGARRET, NEW ROSS')].index)
df = df.drop(index=df[df['Address'] == ('SAGGART LODGE COURT, SAGGART, COUNTY DUBLIN')].index)
df = df.drop(index=df[df['Address'] == ('WOODFIELD MANOR, KILLELINE, NEWCASTLE WEST')].index)

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)
# drop rows
df = df.reset_index(drop=True)
df.shape
```

Out[77]: (591863, 8)

```
In [78]: # check 200 most expensive sales
df.nlargest(200, 'Price')
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Property
232752	2016-12-15	WALFORD, SHREWSBURY RD, DUBLIN 4	Dublin	14250000.000	No	No	Second-Hand Dwelling house /Apartment	
69246	2013-03-29	WALFORD, SHREWSBURY ROAD, DUBLIN 4	Dublin	14000000.000	No	No	Second-Hand Dwelling house /Apartment	
474087	2021-05-19	UIMHI A NAOI, BÓTHA SRIÚSBAIRE	Dublin	13250000.000	No	No	Second-Hand Dwelling house /Apartment	
554012	2022-09-12	ST KILDAS SANDYCOVE AVE EAST, SANDYCOVE CO DUBLIN	Dublin	12500000.000	No	No	Second-Hand Dwelling house /Apartment	
519850	2022-02-16	73 AILESBOURY RD, BALLSBRIDGE, DUBLIN 4	Dublin	11700000.000	No	No	Second-Hand Dwelling house /Apartment	
422075	2020-05-25	CARRICKBRACK HOUSE, OLD CARRICKBRACK RD, HOWTH	Dublin	11500000.000	No	No	Second-Hand Dwelling house /Apartment	

122239	2014-10-03	DODDER GROUND, MILLTOWN BRIDGE RD, MILLTOWN DU...	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
250109	2017-04-27	STANFORD HOUSE, WESTMINSTER RD, FOXROCK DUBLIN 18	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
280895	2017-11-14	GLENARM, TORQUAY ROAD, FOXROCK	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
288051	2017-12-18	SANABRIA, KERRYMOUNT AVE, FOXROCK DUBLIN 18	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
304804	2018-04-23	ASHTOWN, CASTLEKNOCK	Meath	4000000.000	No	No	Second-Hand Dwelling house /Apartment
305438	2018-04-27	25 AILESBUARY ROAD, BALLSBRIDGE, DUBLIN 4	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
325072	2018-08-31	45 WOODLANDS PARK, BLACKROCK, DUBLIN	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
426906	2020-07-10	BRACKEN HILL, THORMANBY RD, BAILY	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
496463	2021-10-05	ANNACREEVY, ADELAIDE RD, GLENAGEARY	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
534363	2022-05-20	TURNBERRY, KERRYMOUNT AVE, FOXROCK DUBLIN 18	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
534370	2022-05-23	1 RAGLAN RD, BALLSBRIDGE, DUBLIN 4	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
589381	2023-04-06	TALL TREES, CLAREMONT RD, CARRICKMINES DUBLIN 18	Dublin	4000000.000	No	No	Second-Hand Dwelling house /Apartment
403983	2019-12-18	APT 1, 48 RANELAGH RD, DUBLIN 6	Dublin	3990000.000	No	No	Second-Hand Dwelling house /Apartment
110357	2014-07-02	16 TEMPLE VILLAS, PALMERSTOWN RD, DUBLIN	Dublin	3950000.000	No	No	Second-Hand Dwelling house /Apartment
408021	2020-01-23	7 HILLSIDE DR, CASTLE PARK DUBLIN 14, DUBLIN	Dublin	3950000.000	No	No	Second-Hand Dwelling house /Apartment
587577	2023-03-28	106 MERRION RD, BALLSBRIDGE, DUBLIN 4	Dublin	3950000.000	No	No	Second-Hand Dwelling house /Apartment
378709	2019-08-01	THE OAKS, PORTERSTOWN RD, CLONSILLA DUBLIN 15	Dublin	3946000.000	No	No	Second-Hand Dwelling house /Apartment

```
In [79]: error_properties = ['45 WOODLANDS PARK, BLACKROCK, DUBLIN', 'ARD NA GLAISE, STILLORGAN PARK, BLACKROCK',  
'COOPERS COURT, BOND STREET', 'THE MONASTERY, MOUNT ARGUS, HAROLDS CROSS',  
'2A-D 3A-D 4A-D, EAST ARRAN STREET', '98 EXCHANGE HALL, BELGARD SQ NORTH, TALLAGHT',  
'5 SPENCER HOUSE, CUSTOM HOUSE SQ,']  
  
df = df.drop(index=df[df['Address'] == ('ASHTOWN, CASTLEKNOCK')].index)  
  
for address in error_properties:  
    df = df.drop(index=df[df['Address'].str.contains(address)].index)  
# drop rows  
df = df.reset_index(drop=True)  
df.shape
```

Out[79]: (591850, 8)

- Within the 200 most expensive properties, Dublin properties are the most common. Many of these appear to be correct and so I will check the most expensive properties outside of Dublin.

```
In [80]: # check 1000 most expensive sales outside of Dublin  
df.nlargest(1000, 'Price')[df['County'] != 'Dublin']
```

```
/var/folders/bn/6x11zkt6n14vfwyb0jymgt00000gp/T/ipykernel_23318/618855605.py:1: UserWarning: Boolean Series key
will be reindexed to match DataFrame index.
df.nlargest(1000, 'Price')[df['County'] != 'Dublin']
```

Out[80]:	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	Prop
313958	2018-06-22	THE PLAZA CITY SQUARE, WATERCOURSE ROAD, BLAC...	Cork	7777973.570	No	No	Second-Hand Dwelling house /Apartment	
58921	2012-11-16	HUMEWOOD CASTLE, KILTEGAN	Wicklow	7225000.000	No	No	Second-Hand Dwelling house /Apartment	
141771	2015-02-09	BOYSTOWN HOUSE AND ESTATE, BLESSINGTON	Wicklow	6350000.000	No	No	Second-Hand Dwelling house /Apartment	
162669	2015-07-23	COURTTOWN HOUSE, COURTTOWN STUD, KILCOCK	Kildare	6250000.000	No	No	Second-Hand Dwelling house /Apartment	
66119	2013-02-05	BUSHY PARK, ENNISKERRY	Wicklow	5850000.000	No	No	Second-Hand Dwelling house /Apartment	
528403	2022-04-11	33 GORTBOY, CHARLEVILLE ROAD, KILMALLOCK	Limerick	5773794.080	Yes	Yes	New Dwelling house /Apartment	
352262	2019-02-13	KILFINNAN CASTLE, GLANDORE, CO CORK	Cork	5732000.000	No	No	Second-Hand Dwelling house /Apartment	
314413	2018-06-27	41 CAMDEN COURT, CARROLLS QUAY, CORK	Cork	5525000.000	No	No	Second-Hand Dwelling house /Apartment	
552206	2022-09-01	CONSTANTIA FARM, COMPASS HILL, KINSALE	Clare	5080000.000	No	No	Second-Hand Dwelling house /Apartment	
188150	2016-01-25	KILBRIDE HILL HOUSE, HERBERT ROAD, BRAY	Wicklow	5000000.000	No	No	Second-Hand Dwelling house /Apartment	
238487	2017-01-27	KILBRIDE HILL HOUSE, HERBERT ROAD, BRAY	Wicklow	5000000.000	No	No	Second-Hand Dwelling house /Apartment	
95500	2014-01-24	BOYSTOWN HOUSE, BOYSTOWN	Wicklow	4925000.000	No	No	Second-Hand Dwelling house /Apartment	
91449	2013-12-16	MILICENT HOUSE DEMESNE FARMHOUSE, AND ADJOIN...	Kildare	4875000.000	No	No	Second-Hand Dwelling house /Apartment	
451306	2020-12-15	LOUGHLIN GREEN, LOUGHLIN, KILDARE	Kildare	4875000.000	No	Yes	New Dwelling house /Apartment	
514204	2022-01-07	RAFFEEN, SCILLY, KINSALE	Cork	4750000.000	No	No	Second-Hand Dwelling house /Apartment	
112275	2014-07-15	AVONMORE HOUSE, LANDS GLEBE, WICKLOW	Wicklow	4400000.000	No	No	Second-Hand Dwelling house /Apartment	
499499	2021-10-21	KILCREENE LODGE, KILKENNY, KILKENNY	Kilkenny	4250000.000	No	No	Second-Hand Dwelling house /Apartment	
172638	2015-10-05	CAPARD HOUSE AND ADJOINING LANDS, ROSENALLIS	Laois	4246000.000	No	No	Second-Hand Dwelling house /Apartment	
415626	2020-03-19	BRAGANSTOWN HOUSE, CASTLEBELLINGHAM, DUNDALK	Louth	4200000.000	No	No	Second-Hand Dwelling house /Apartment	
278910	2017-11-01	AVOKA, BLACKROCK RD, CORK	Cork	4080000.000	No	No	Second-Hand Dwelling house /Apartment	
289290	2017-12-21	2 PALMYRA PARK, GALWAY, GALWAY	Galway	3925000.000	No	No	Second-Hand Dwelling house /Apartment	
447642	2020-12-01	AN CULU, DROMORE, GREENANE	Kerry	3850000.000	No	No	Second-Hand Dwelling house /Apartment	
454489	2020-12-22	KNOCKDRIN CASTLE, MULLINGAR, CO WESTMEATH	Westmeath	3850000.000	No	No	Second-Hand Dwelling house /Apartment	
573487	2022-12-16	PEBBLE BEACH, CURRAGH, ARDMORE	Waterford	3850000.000	No	No	Second-Hand Dwelling house /Apartment	
		CARMELITE MONASTERY,					Second-Hand Dwelling	

508917	2021-12-09	BARRETTSTOWN HOUSE, NEWBRIDGE, KILDARE	Kildare	2405000.000	No	No	Second-Hand Dwelling house /Apartment
473911	2021-05-18	GORT BUI, BALLYMAKENNY ROAD, DROGHEDA	Louth	2401594.720	Yes	Yes	New Dwelling house /Apartment
299116	2018-03-13	CRAIGNATHAN, SOUTH PLACE, THE BURNABY	Wicklow	2400000.000	No	No	Second-Hand Dwelling house /Apartment
236831	2017-01-13	OUGHTERANY VILLAGE, BRANGANSTOWN, KILCOCK	Kildare	2394950.000	No	Yes	New Dwelling house /Apartment
232674	2016-12-15	BALLISODARE, BALLISODARE	Sligo	2393000.000	No	Yes	New Dwelling house /Apartment
471824	2021-05-04	43 CARRIGBROOK, TULLOW ROAD, CARLOW	Carlow	2389213.070	Yes	Yes	New Dwelling house /Apartment
279762	2017-11-07	HOLY FAITH CONVENT, KILCOOLE, CO WICKLOW	Wicklow	2373000.000	No	No	Second-Hand Dwelling house /Apartment
17904	2010-11-15	DUNCARRIG HOUSE, SUMMERCOVE, KINSALE	Cork	2350000.000	No	No	Second-Hand Dwelling house /Apartment
78735	2013-08-06	TUNDUFF, ABBEYLEIX	Laois	2350000.000	No	No	Second-Hand Dwelling house /Apartment
497944	2021-10-13	DUNSALLAGH HOUSE, KILGARRON HILL, ENNISKERRY	Wicklow	2350000.000	No	No	Second-Hand Dwelling house /Apartment
546723	2022-07-29	5 NORTH SHORE, THE GROVE, GREYSTONES	Wicklow	2350000.000	No	No	Second-Hand Dwelling house /Apartment
547551	2022-08-04	AVALON, SANDYCOVE, KINSALE	Cork	2350000.000	No	No	Second-Hand Dwelling house /Apartment

```
In [81]: error_properties = ['LOUGHLIN GREEN, LOUGHLIN, KILDARE',
 'CARMELITE MONASTERY, DELGANY, CO. WICKLOW',
 'ARDBRACCAN HOUSE, CHURCH LODGES AND CHURCH, NAVAN',
 'THE VIEW, THE HAWTHORNS, TULLAMORE', '11 BRUACH NA SIONNA, CASTLECONNELL, LIMERICK',
 '2 SEANAMHUILEAN & ORS, LADYSWELL, NORTH CITY', 'HOUSE 1, MIDDLETOWN HOUSE RETIREMENT VILLAGE',
 '1 MOGEESHA HOUSE, WATERSEDGE, RIVERSIDE WAY', '17 PRIMROSE FORGE, CELBRIDGE, KILDARE',
 '1 MONASTERY CLOSE, TEMPLEMORE ROAD, THURLES', '66 CARRIGBROOK, TULLOW ROAD, CARLOW',
 'LEITIR BEAG, CRIEVESMITH, LETTERKENNY', '2 CLUAIN RAITE, BUNRATTY WEST, BUNRATTY',
 '1 MILITARY CLOSE, CREAGH ROAD, BALLINROBE', 'THE TUNNEL, ST VINCENTS RD, GREYSTONES',
 '10 CUAN NA RÍ OCEANFRONT RESORT, CARRIGART', '101 ALLIN'S QUAY, PIER HEAD, YOUGHAL',
 '25 THE LODGE, ABBEYCOURT HOTEL, DUBLIN ROAD', 'MAYDENHAYES LAWN, MORNINGTON, CO. MEATH',
 '5 OAKLANDS DR, LONGFORD, LONGFORD', 'GLEANN NA GREINE, MULLINDRAIT, STRANORLAR',
 '142 BELVEDERE HILLS, MULLINGAR, WESTMEATH', 'GORT BUI, BALLYMAKENNY ROAD, DROGHEDA',
 'BALLISODARE, BALLISODARE', '43 CARRIGBROOK, TULLOW ROAD, CARLOW', '24 HOUSES',
 'HOLY FAITH CONVENT, KILCOOLE, CO WICKLOW'
]

# 12 RACECOURSE HILL, CLIFTON, GALWAY & 3 COIS ABHAINN, RIVERSTICK, CORK & KENNEDY PLAZA, THE QUAY, NEW ROSS
# & 5 CASTLEGAR, CHAPEL LANE, MALLOW & 8 BASIN COURT, TRALEE, KERRY & 71 COIS LUACHRA, DOORADOYLE, LIMERICK
# & UNIT 172 (BLOCK Z) LEARGAN & 1 DUN UISCE, CAHIR, TIPPERARY & MULLAGHDUFF, BALLYCONNELL (x2) & TUNDUFF, ABBEY
rows = [360729, 236720, 454925, 349815, 539812, 295325, 22489, 392732, 536856, 566288, 78735]

df = df.drop(index=rows)
df = df.drop(index=df[df['Address'] == ('2 PALMYRA PARK, GALWAY, GALWAY')].index)
df = df.drop(index=df[df['Address'] == ('ABBEYWOOD, CLANE')].index)
df = df.drop(index=df[df['Address'] == ('DRUMMIN VILLAGE, NENAGH')].index)
df = df.drop(index=df[df['Address'] == ('GORESBRIDGE, KILKENNY')].index)
df = df.drop(index=df[df['Address'] == ('3 SLI NA COILLE, CROBALLY UPPER, TRAMORE')].index)
df = df.drop(index=df[df['Address'] == ('AN DUCHE, SCALLAGHEEN')].index)
df = df.drop(index=df[df['Address'] == ('OUGHTERANY VILLAGE, BRANGANSTOWN, KILCOCK')].index)
df = df.drop(index=df[df['Address'] == ('WOODLAWN PARK, BALLYSIMON')].index)

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)
# drop rows
df = df.reset_index(drop=True)
df.shape
```

Out[81]: (591803, 8)

In [82]: # check the most expensive sales in the top 1% outside of Dublin

```
df.nlargest(50000, 'Price')[!(df['County'] != 'Dublin') & (df['Price'] >= p_99)]['Address'].unique()
```

```
/var/folders/bn/6x11zkt6n14vfwyb0jymgt00000gp/T/ipykernel_23318/3478829852.py:1: UserWarning: Boolean Series ke  
y will be reindexed to match DataFrame index.  
df.nlargest(50000, 'Price')[!(df['County'] != 'Dublin') & (df['Price'] >= p_99)]['Address'].unique()
```

```
Out[82]: array(['THE PLAZA CITY SQUARE, WATERCOURSE ROAD, BLACKPOOL',  
   'HUMEWOOD CASTLE, KILTEGAN',  
   'BOYSTOWN HOUSE AND ESTATE, BLESSINGTON',  
   'COURTTOWN HOUSE, COURTTOWN STUD, KILCOCK',  
   'BUSHY PARK, ENNISKERRY',  
   '33 GORTBOY, CHARLEVILLE ROAD, KILMALLOCK',  
   'KILFINNAN CASTLE, GLANDORE, CO CORK',  
   '41 CAMDEN COURT, CARROLLS QUAY, CORK',  
   'CONSTANTIA FARM, COMPASS HILL, KINSALE',  
   'KILBRIDE HILL HOUSE, HERBERT ROAD, BRAY',  
   'BOYSTOWN HOUSE, BOYSTOWN',  
   'MILLICENT HOUSE DEMESNE FARMHOUSE, AND ADJOINING LANDS, SALLINS NAAS',  
   'RAFFEEN, SCILLY, KINSALE', 'AVONMORE HOUSE, LANDS GLEBE, WICKLOW',  
   'KILCREENE LODGE, KILKENNY, KILKENNY',  
   'CAPARD HOUSE AND ADJOINING LANDS, ROSENALLIS',  
   'BRAGANSTOWN HOUSE, CASTLEBELLINGHAM, DUNDALK',  
   'AVOKA, BLACKROCK RD, CORK', 'AN CULU, DROMORE, GREENANE',  
   'KNOCKDRIN CASTLE, MULLINGAR, CO WESTMEATH',  
   'PEBBLE BEACH, CURRAGH, ARDMORE',  
   'GRATTAN HOUSE, TINNEHINCH, ENNISKERRY', 'SEAMARK, GLANDORE',  
   'CASTLEGANGE, NEWCASTLE, WICKLOW',  
   'SANDYMOUNT HOUSE, BLACKROCK, LOUTH',  
   'RATHMORE HOUSE, RATHMORE, NAAS',  
   'FASTNET HOUSE, ARDBRACK, KINSALE',  
   'KERDUFFSTOWN HOUSE, JOHNSTOWN, NAAS',  
   'DOWDSTOWN HOUSE, STRAFFAN ROAD, MAYNOOTH',  
   'STONE HALL, GLANDORE, CO CORK',  
   'MAIN HOUSE, DOLLANSTOWN STUD LTD, KILCOCK',  
   'BRANDON HOUSE, CARBURY', '12 RACECOURSE HILL, CLIFTON, GALWAY',  
   '9 AN MÓINÉAR, MURROUGH, RENMORE',  
   'KENNEDY PLAZA, THE QUAY, NEW ROSS',  
   'RAHEEN, LISCAFFER RD, LAHINCH',  
   'GUEST ANNEX, STRAFFAN LODGE, STRAFFAN',  
   '1 INSE BEAG, DOUGHISKA RD, DOUGHISKA',  
   'LOUGH RUSHEEN HOUSE, BARNA ROAD, KNOCKNACARRA',  
   'SHELBOURNE SQUARE, NORTH CIRCULAR ROAD',  
   'THE PADDOCKS, CASTLEDILLON, STRAFFAN', 'WILFORD, DUBLIN RD, BRAY',  
   'WINTERWOOD, 2 THE DEMENSE, ADARE',  
   'BURRISHOOLE LODGE, NEWPORT, MAYO',  
   '3CHURCHFIELDS, THE K CLUB, STRAFFAN', 'COOLMAIN CASTLE, COOLMAIN',  
   '5 CASTLEGAR, CHAPEL LANE, MALLOW', '8 BASIN COURT, TRALEE, KERRY',  
   'SHEARWATER, TEMPLENOE, KENMARE', 'STACUMNY HOUSE, CELBRIDGE',  
   'MERVYN, COOKSTOWN RD, ENNISKERRY', 'DALEMOUNT, DUBLIN ROAD, NAAS',  
   'LISNALEE HOUSE, BARRINGTON S AVE, BLACKROCK',  
   'CASTLE OLIVER, ARDPATRICK, KILMALLOCK',  
   '1 DUN UISCE, CAHIR, TIPPERARY',  
   'FEDAMORE HOUSE, FEDAMORE, LIMERICK',  
   'PROSPECT HOUSE, NEWCASTLE, NR GREYSTONES',  
   'MAIN HOUSE, HORSE ISLAND, SCHULL', 'SCILLY HILL, SCILLY, KINSALE',  
   'THE ROCKET HOUSE, CASTLETOWNSHEND, SKIBBEREEN',  
   'LUGDINE PARK, GLENGARRIFF, CORK',  
   'ST JOSEPHS, MODEL FARM RD, CORK', 'FASSAROE HOUSE, BRAY',  
   'KINDESTOWN HOUSE, DELGANY, COUNTY WICKLOW',  
   'WINDY RIDGE, ROCHESTOWN RD, CORK', 'MULLAGHDUFF, BALLYCONNELL',  
   'WHALE HARBOUR COTTAGE, ERVALLAGH, ROUNDSTONE',  
   'MUIR WOOD, MARYBOROUGH ORCHARD, DOUGLAS',  
   'BARRETTSTOWN HOUSE, NEWBRIDGE, KILDARE',  
   'CRAIGNATHAN, SOUTH PLACE, THE BURNABY',  
   'DUNCARRIG HOUSE, SUMMERCOVE, KINSALE',  
   'DUNSALLAGH HOUSE, KILGARRON HILL, ENNISKERRY',  
   '5 NORTH SHORE, THE GROVE, GREYSTONES',  
   'avalon, SANDYCOVE, KINSALE',  
   'ASHLEIGH HOUSE, CORBALLY, CELBRIDGE',  
   'CORRELLA, ST VINCENT ROAD, GREYSTONES',  
   'CURRABEG, WOODVIEW, DOUGLAS RD',  
   'KILCORNAN, NORTH CIRCULAR RD, LIMERICK',  
   'DOYLES BARN, THE QUAY, NEW ROSS',  
   'COTTAGE 1 THE COURTYARD, CLONABREANY CROSSAKIELS, KELLS',  
   '7 GLENAIR MANOR, PIORY ROAD, DELGANY',  
   'AUBREY COURT, PARNELL ROAD, BRAY',  
   '83 BRU NA SIONNA, SHANNON, COUNTY CLARE',  
   'SALTERBRIDGE HOUSE, SALTERBRIDGE, CAPPOQUIN',  
   'KILLEEN, DUNANY, MEATH', 'LOUGHINSTOWN HOUSE, KILBRIDE, CLONEE',  
   '6 BRINDLEY PARK RD, ASHBOURNE, MEATH',  
   'CLARA HOUSE, LANDS AT STYLEBAWN, DELGANY',  
   'GLEBE HOUSE, LISHEEN, CHURCH CROSS',  
   '72 COLLEGE MANOR, HOEYS LANE, DUNDALK',  
   '3 SEAMOUNT, SALTHILL, GALWAY',
```

'ONAGH FARMLONGE, ONAGH, ENNISKERRY',
'INISFREE, 10 KING EDWARD ROAD, BRAY',
'RAVENSWOOD, BUNCLOUDY, CO WEXFORD',
'RATHCARRICK HOUSE, RATHCARRICK, STRANDHILL',
'8 SILLAN, CHURCH LANE, GREYSTONES',
'TAVISTOCK HOUSE, PORTLAND RD, GREYSTONES',
'2 CARRS HILL, DOUGLAS, CORK',
'50 RAMPART VIEW, RAHARDRUM, VIRGINIA',
'LINARO HOUSE, LINARO AVE, COLLEGE RD',
'REENMEEN EAST, GLENGARRIFF, CO CORK',
'BESSBOROUGH HOUSE, BALRATH', "CAMPBELL'S LANE, FARNAGH",
'OLD CHURCH STREET, ATHENRY', 'ST HELENS, TAYLORS HILL, GALWAY',
'1 AN GRIANÁN, RIVER ST, KILLENAULE',
'KNOCKMORE, ENNISKERRY, CO WICKLOW',
'GLENSIDE HOUSE, WOODSTOCK, KILQUADE',
'9 THE HALL, MEADOWFIELDS, THE MOYNE',
'YEOMANSTOWN LODGE, VICTORIA BRIDGE, NAAS',
'CHESNUT LANE, BUSHYPARK, GALWAY', 'MALLOW CASTLE, MALLOW',
'NEWBERRY HOUSE & STUD FARM, KILCULLEN', 'RAVENSWOOD, BUNCLOUDY',
'LISGAR, BAILIEBOROUGH', '10 BARTON CLOSE, CHURCHFIELDS, STRAFFAN',
'MOYARTA, NORTH CIRCULAR RD, LIMERICK',
'HAVEN HILL, SUMMERCOVE, KINSALE',
'SPRINGDALE, ENNIS RD, LIMERICK',
'LAVENDER HEIGHTS, ROCHESTOWN RD, CORK',
'HOTWELL BALLYMAN ROAD, ENNISKERRY, COUNTY WICKLOW',
'LOWTOWN, ROBERTSTOWN, COUNTY KILDARE',
'LONG ACRE, COOKSTOWN, ENNISKERRY',
'SUNCROFT, MODEL FARM RD, CORK',
'4 CLANWILLIAM, GOLF LINK ROAD, CASTLETROY',
'GROVE HOUSE, ROSS RD, KILLARNEY',
'25 LINDVILLE, BLACKROCK RD, BLACKROCK',
'CHESAPEAKE, 2 KENDALSTOWN RISE, DELGANY',
'RATHELLIN, LEIGHLINBRIDGE, CARLOW',
'ROCKFIELD, MARLINSTOWN, MULLINGAR',
'1 GLEANN CHUILLE, KILCULLY, CO CORK',
'OAKFIELD COTTAGE, KILGARRON HILL, ENNISKERRY',
'PREMISES AT GATE LODGE, BLACKROCK, CORK',
'15 MARYBOROUGH ORCHARD, MARYBOROUGH HILL, DOUGLAS',
'LINDEN, ST VINCENT RD, GREYSTONES',
'CHESTNUT LANE, DANGAN, GALWAY',
'3 COLPE DR, DEEPFORDE, DUBLIN RD',
'THE RECTORY, PARSON ST, MAYNOOTH',
'CLONBRYN, KINLEN RD, GREYSTONES',
'GREENWAYS MANOR AVE, GREYSTONES, CO WICKLOW',
'101 EAGLE VALLEY, ENNISKERRY, WICKLOW',
'KILFANE GLEBE, THOMASTOWN, KILKENNY', 'IONA PARK STUD, STRAFFAN',
'MOUNT NEPHIN, BISHOPSTOWN AVE, MODEL FARM RD',
'ISLAND HOUSE, CARRIGILLIHY, UNION HALL',
'6 ROCKY VALLEY CRESCENT, KILMACANOQUE, WICKLOW',
'BALLINACOOLA HOUSE, GLENEALY', 'SLEVOIR, TERRYGLASS, NENAGH',
'LADYSWELL, WHITCHURCH, STRAFFAN',
'OLD COASTGUARD STATION, DROMARD, COASTGUARD STATION',
'CARRIGWEIR, TUAM', '11 LANCASTER QUAY, CORK, CORK',
'CARRIGFOYLE, MARYBOROUGH ORCHARD, DOUGLAS',
'NIRVANA, FENNELLS BAY, CROSSHAVEN',
'24 BURNABY PARK, GREYSTONES, WICKLOW',
'TELUK ANSON, DOCK RD, DUNMORE EAST',
'BALLYKNOCKAN HOUSE, BALLYKNOCKAN UPPER, RATHDRUM'] , dtype=object)

In [130]:

```
error_properties = ['THE PLAZA CITY SQUARE, WATERCOURSE ROAD, BLACKPOOL',
                    '41 CAMDEN COURT, CARROLLS QUAY, CORK',
                    'DOYLES BARN, THE QUAY, NEW ROSS', 'COTTAGE 1 THE COURTYARD, CLONABREANY',
                    '83 BRU NA SIONNA, SHANNON, COUNTY CLARE', 'LOUGHINSTOWN HOUSE, KILBRIDE, CLONEE',
                    '72 COLLEGE MANOR, HOEYS LANE, DUNDALK', 'NEWBOROUGH, BALLINACLASHET, BELGOOLY',
                    '1 CRESCENT VILLAS, THE CRESCNET, GALWAY', 'GLOUNAGILLAGH, CARAGH LAKE, KILLORGLIN',
                    '1 ST MICHAELS MEWS, BARRINGTON ST, LIMERICK', 'ROS ALAINN HEIGHTS, GURTEENROE, MACROOM',
                    '88, THE CAMPION, MARINA VILLAGE', '19 MILLTOWN MEADOWS, ASHBOURNE',
                    '6 ARD NA DARA, ARDMORE ROAD, MULLINGAR', '11 PARK STREET, COULTER PLACE, MOORLAND ARMAGH RD',
                    'EMMET PLACE DRAWBRIDGE STREET, PERRY STREET & LAVITTS QUAY, CORK', 'HEATHHILL, NEWTOWNCUNI',
                    '10A 10B 11A 11B 12A 12B, 13A 13B 14 15 16 17 18 AND 19, MULLAGH BÁN BALLAGHDERO',
                    '\OAKMEADOWS\', BROOKFIELD, DONEGAL TOWN', 'ST KILDA#039;S, WESTERN ROAD, CORK',
                    'WOODHOUSE ESTATE, STRADBALLY', '1 KIMBERLY LANE, CARRIG EDEN, GREYSTONES', 'GLENA HOUSE, MI',
                    '5 HAZEL CLOSE, GREENHILLS, CARRICK-ON-SUIR', '38 CLONMULLION RETIREMENT VILLAGE, MONEEN LAI',
                    '1 LELIA HOUSE, ST LELIA ST, LIMERICK', 'URSULINE CONVENT, BLACKROCK, CORK',
                    '1 LINTOWN MEWS, LINTOWN HALL, JOHN SWELL ROAD', '1 BLACKMORE COURT, BLACKMORE LANE, MEADE ST',
                    'SAN PAULA, ORCHARD RD, CORK', '111 PARKLANDS, PIERSE ROAD, TRALEE',
                    '1 BALLINTAGGART HOLIDAY HOMES, DINGLE, KERRY', '38 THE MAPLES, OAKLEIGH WOOD, TULLA ROAD',
                    'DOONBEG PROPERTY COMPANY, KILRUSH, COTTAGE 158', '64 FIONNUISCE, DOUGHISKA, GALWAY',
                    '1 PENNEFEATHER COURT, HEBRON ROAD, KILKENNY CITY', 'CLONMACKEN, KNOCKLYON',
                    '27 THE MALTINGS, LOUGHREA, CO. GALWAY', 'MAIGH RUA, DOUGHISKA, GALWAY',
                    'COVA DA IRIA, UPPER SALTHILL, GALWAY', 'RANNOCH, BETTYSTOWN, MEATH',
                    'SPRINGFIELD HOUSE, CLANE ROAD, CELBRIDGE', 'WELLS, GOREY, WEXFORD',
                    '35 GORTONORA, DINGLE, CO KERRY', 'ST CATHERINES CONVENT, HIGH FIELD LAWN, MODEL FARM ROAD']
```

```

'1 PARK STREET, COULTER PLACE, ARMAGH ROAD', 'GREVILLE COURT, MARKET STREET, GRANARD',
'THE K CLUB, STRAFFEN, CO KILDARE', 'ST PATRICKS HOUSE, LOWER GLANNMIRE ROAD',
'1 CAMPBELLS LANE, LONGFORD, CO LONGFORD', 'RATHDRUM, BALLINACLAUGH',
'170 THE GLEN, KILNACOURT WOODS, PORTARLINGTON', 'ST ANTHONY, DWYER PARK, BRAY',
'ABBEYBROOK, KILBEGGAN, COUNTY WESTMEATH', '42 SUMMERCOVE, LAHINCH', 'KILCLEIGH, MAYNOOTH RO
'1 KNIGHTSBROOK SQ, TRIM, MEATH', '1 GLEN ABHAINN VILLAGE, DUBLIN RD, ATHLONE',
'CARRICK PLAZA SUITES, TOWN CENTRE, CARRICK ON SHANNON', 'MELLON STUD, MELLON, KILDIMO',
'MEADOW COURT STUD, MADDENSTOWN, THE CURRAGH', 'RELDARE HOUSE, MODEL FARM RD, CARRIGROHANE'
'1 PLUNKETT COURT, OLIVER PLUNKETT STREET, BANDON', 'WATERFORD CASTLE, THE ISLAND, BALLINA
'PREMISES AT RIVERCHAPEL WOOD, COURNTOWN', '1 GROVE COURT, EYRE ST, NEWBRIDGE',
'1 BIRCH HOUSE, RIVERVIEW, MAIN ST', 'CHURCH LANE, BALLINAMORE', '2 COURTYARD, MURANNY, MAYO
'6 CHERRYVALLEY COURT, RATHMOYLAN COUNTY MEATH', '13 GREENFIELDS, LANESBORO ROAD, ROSCOMMON
'WYGRAM PLACE, WEXFORD', 'BALLINTESKIN STUD, OLD LONG HILL, ENNISKERRY',
'1 RIVERSIDE COURT, JOHN STREET, DROGHEDA', '124 CLUAIN BUI, QUINAGH, CARLOW',
'14 DONEGAN COURT, NEW RD, GALWAY', 'BUN NA LEACA, NEWCASTLE, CO GALWAY',
'GLENEALY HALL, CUL NA COILLE, GLENEALY', '37 HOOKLESS HOLIDAY HOMES, FETHARD ON SEA, CO.WE
'CARNMUGGAGH UPPER, LETTERKENNY', 'NO72 CAISLEAN RI, ATHENERY, CO GALWAY',
'1 DINGLE MARINA COTTAGES, THE WOOD, DINGLE', 'ACHILL HOUSE, WESTERN RD, CORK',
'25 QUAY STREET, GALWAY CITY, COUNTY GALWAY', 'CARRIG VIEW, GREENHILLS ROAD, WICKLOW',
'ARDROSTIG, WATERFALL, BISHOPSTOWN', 'RYEVALE MILL, LEIXLIP', 'UNIT 28 AYRESFIELD, GRANGE RO
'NORELANDS, STONEYFORD, CO KILKENNY', '31 COLLEGE RD, GALWAY, GALWAY',
'CAMPBELL\S LANE, FARNAGH', 'OLD CHURCH STREET, ATHENERY', '9 THE HALL, MEADOWFIELDS, THE MO
'MALLOW CASTLE, MALLOW', 'PREMISES AT GATE LODGE, BLACKROCK, CORK',
'SITE AT GATE LODGE, BLACKROCK, CORK', '11 LANCASTER QUAY, CORK, CORK'

]

# 6 BRINDLEY PARK RD, ASHBOURNE, MEATH & 11 GEATA NA CATHRACH, FAIRGREEN RD, GALWAY
# 2 CLUAIN DARA, KINGSCOURT, CAVAN & 1 BROOKES MILL, PROSPEROUS, COUNTY KILDARE
# 9 LORRO GATE, ATHENERY, GALWAY & 1 BRAMBLE HILL, GREEN VALLEY, CASTLETREASURE
# 18 CLUAIN CAIRN, CORK, CORK & 1 COIS TEAMPAILL, NEWCASTLEWEST, LIMERICK
# 59 ROWAN HILL, MOUNT OVAL VILLAGE, ROCHESTOWN & 4 THE WRITINGS, LISTOWEL, COUNTY KERRY
# 1 CORAL GROVE, DUNMORE EAST, CO WATERFORD & 9 ASHBROOK, ORANMORE, GALWAY
# 1 LIMEWOOD COURT, CURRAHEEN RD, BISHOPSTOWN & 1 CASTLE COURT, ABBEY ST, PORTUMNA
# 2 THE LAWN, ASHWOOD, ASHBOURNE & 50 RAMPART VIEW, RAHARDRUM, VIRGINIA
# LISGAR, BAILIEBOROUGH & LOWTOWN, ROBERTSTOWN, COUNTY KILDARE
# 3 COLPE DR, DEEPFORDE, DUBLIN RD

rows = [467681, 210251, 231533, 268445, 280106, 183532, 350657, 233345, 285160, 424954, 344639, 580349, 272478,
       308136, 400479, 144756, 423837, 109602]

df = df.drop(index=rows)
df = df.drop(index=df[df['Address'] == ('AUBREY COURT, PARNELL ROAD, BRAY')].index)
df = df.drop(index=df[df['Address'] == ('DE VESCI HILL, ABBEYLEIX')].index)
df = df.drop(index=df[df['Address'] == ('KINGSTON ROAD, GALWAY')].index)
df = df.drop(index=df[df['Address'] == ('THE GRANGE, LURGANBOY')].index)
df = df.drop(index=df[df['Address'] == ('NEWTOWN, TRAMORE')].index)
df = df.drop(index=df[df['Address'] == ('OAKLEY WOOD, TULLOW')].index)
df = df.drop(index=df[df['Address'] == ('1 PARKLANDS HOLIDAY HOMES, PORT RD, KILLARNEY')].index)
df = df.drop(index=df[df['Address'] == ('POYNTON PLACE, BALLYVOR')].index)
df = df.drop(index=df[df['Address'] == ('5 CARRIGBROOK, TULLOW ROAD, CARLOW')].index)
df = df.drop(index=df[df['Address'] == ('3 GARRAI SHEAIN, ROSCAM, GALWAY')].index)
df = df.drop(index=df[df['Address'].str.contains('COIS NA MONA')] & (df['Address'].str.contains('ORANMORE'))).index
df = df.drop(index=df[df['Address'] == ('ALDERFIELD, FARNAGH HILL, LONGFORD')].index)
df = df.drop(index=df[df['Address'] == ('DRAKES POINT, CROSSHAVEN')].index)
df = df.drop(index=df[df['Address'] == ('ASCAIL ROIS, CARRICKMACROSS')].index)
df = df.drop(index=df[df['Address'] == ('CARRAIGH ABHAINN, KNOXPARK, BALISODARE')].index)
df = df.drop(index=df[df['Address'] == ('TULLOW ROAD, CARLOW')].index)
df = df.drop(index=df[df['Address'] == ('GARRYCASTLE, ATHLONE, CO WESTMEATH')].index)
df = df.drop(index=df[df['Address'] == ('RIVERSIDE, PORTARLINGTON, OFFALY')].index)
df = df.drop(index=df[df['Address'] == ('7 HILLVIEW, CREGGAUN NA HILLA, CLARECASTLE')].index)
df = df.drop(index=df[df['Address'] == ('ACHA BHILE, LAHINCH ROAD, ENNIS')].index)
df = df.drop(index=df[df['Address'] == ('LACKABANE, KILLARNEY')].index)
df = df.drop(index=df[df['Address'] == ('REENMEEN EAST, GLENGARRIFF, CO CORK')].index)
df = df.drop(index=df[df['Address'] == ('CARRIGWEIR, TUAM')].index)

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)
# drop rows
df = df.reset_index(drop=True)
df.shape

```

Out[130]: (591635, 8)

- Now I will check the most expensive properties within Dublin. I'm assuming there will be a large number of addresses and so I will filter out addresses from certain areas that I know are expensive.

```

In [131]: for address in df.nlargest(50000, 'Price')[ (df['County'] == 'Dublin') & (df['Price'] >= p_99) &
           (~df['Address'].str.contains('MONKSTOWN')) & (~df['Address'].str.contains('BALLSBRIDG
           (~df['Address'].str.contains('AILESBUY')) & (~df['Address'].str.contains('DUBLIN 4'
           (~df['Address'].str.contains('DALKEY')) & (~df['Address'].str.contains('BAGGOT')) &
           (~df['Address'].str.contains('LEESON')) & (~df['Address'].str.contains('KILLLINEY')) &
           (~df['Address'].str.contains('DUBLIN4')) & (~df['Address'].str.contains('GRAFTON ST'
           (~df['Address'].str.contains('SHREWSBUY')) & (~df['Address'].str.contains('DARTRY'))
           (~df['Address'].str.contains('HOWTH')) & (~df['Address'].str.contains('FOXROCK')) &

```

```

        (~df['Address'].str.contains('BLACKROCK')) & (~df['Address'].str.contains('GLENAGEAR'))
        (~df['Address'].str.contains('SANDYCOVE')) & (~df['Address'].str.contains('CLONTARF'))
        (~df['Address'].str.contains('LANSDOWNE')) & (~df['Address'].str.contains('HOTEL'))
        (~df['Address'].str.contains('SUTTON')) & (~df['Address'].str.contains('RATHGAR')) &
        (~df['Address'].str.contains('RANELAGH')) & (~df['Address'].str.contains('RATHMINES'))
        (~df['Address'].str.contains('DONNYBROOK')) & (~df['Address'].str.contains('HANOVER'))
        (~df['Address'].str.contains('DUBLIN 2')) & (~df['Address'].str.contains('DUBLIN2'))
print(address)

```

UIMHI A NAOI, BÓTHA SRIÚSBAIRE
 SORRENTO HOUSE, 1 SORRENTO TERRACE, SORRENTO ROAD
 SEAFIELD HOUSE, KILCREA, DONABATE
 AUBURN HOUSE, MALAHIDE, RIGHT OF WAY AT DN226834F
 HAMPTON MONASTERY, GRACEPARK ROAD, DRUMCONDRA
 ROCKALL, THE BIRCHES, TORQUAY RD DUBLIN 18
 CLONTRA, QUINNS RD, SHANKILL
 APT 1, 138A SANDFORD RD, DUBLIN
 CARNALEA, THORMANBY RD, BAILY DUBLIN 13
 LADYWELL, CORRYBEG, TEMPLEOGUE
 MARIST COMMUNITY, MOUNT ST. MARY'S, MILLTOWN
 BOLTON HALL, RATHFARNHAM, DUBLIN 16
 6 BALLYBRIDE MANOR, BALLYBRIDE, RATHMICHAEL
 6 CLYDE RD, BALLBRIDGE, DUBLIN CITY
 KNOCKRATH, CASTLEKNOCK ROAD, DUBLIN 15
 11 MOYOLA COURT, CHURCHTOWN RD LOWER, DUBLIN 14
 DODDER GROUND, MILLTOWN BRIDGE RD, MILLTOWN DUBLIN 14
 BRACKEN HILL, THORMANBY RD, BAILY
 TALL TREES, CLAREMONT RD, CARRICKMINES DUBLIN 18
 16 TEMPLE VILLAS, PALMERSTOWN RD, DUBLIN
 7 HILLSIDE DR, CASTLE PARK DUBLIN 14, DUBLIN
 THE OAKS, PORTERSTOWN RD, CLONSILLA DUBLIN 15
 1 PRIORY GARDENS, PRIORY GROVE, STILLORGAN
 THE MEWS, CITYWEST, DUBLIN
 2 CHERRYGARTH, MOUNT MERRION, DUBLIN
 76 BUTTERFIELD AVE, RATHFARNHAM, DUBLIN 14
 86 BUTTERFIELD AVENUE, RATHFARNHAM, DUBLIN 14
 LINN DUBH, HAYSTOWN RD, RUSH
 APARTMENT 14, CASTLE COURT, 5-6 CASTLE COURT
 DUNMORE, PALMERSTON RD, DUBLIN 6
 HILLSIDE, THE HILL, MALAHIDE
 18 CASTLEGANGE GARDENS, CASTLEGANGE, DUBLIN
 ST ANNES, SYDENHAM RD, DUNDRUM DUBLIN 14
 BRENTWOOD, SYDENHAM RD, DUNDRUM DUBLIN 14
 WINWARD, CEANCHOR RD, BAILY
 115 SEVILLE PLACE &, 175-176 NORTH STRAND ROAD, FIVE LAMPS
 34 MAYOR STREET UPPER, DUBLIN 1
 SAXAROON, GLENAMUCK RD, CARRICKMINES DUBLIN 18
 WHINSFIELD, LAMBS CROSS, SANDYFORD DUBLIN 18
 GRANGE HOUSE, GRANGE, BALLYBOUGHAL
 17 TEMPLE GARDENS, DUBLIN 6, DUBLIN
 BRETT COURT, FOLEY STREET, DUBLIN 1
 10 BRENNANSTOWN VALE, CARRICKMINES, DUBLIN 18
 16 TEMPLE VILLAS, PALMERSTOWN ROAD
 ROSENEATH, CHURCH RD, MALAHIDE DUBLIN
 2 DISWELLSTOWN HEATH, CASTLEKNOCK, DUBLIN
 4 DISWELLSTOWN HEATH, CASTLEKNOCK
 1 CARLISLE TCE, CHURCH RD, MALAHIDE
 18 LOMOND AVE, DUBLIN 3, DUBLIN
 BALURE HOUSE, BALURE LANE, CHURCH RD
 1 HARBOUR HOUSE, BOND ST, DUBLIN 8
 1 EMMET LODGE, EMMET COURT, INCHICORE
 7 LANDORE HALL, 7 LANDORE, ORWELL RD DUBLIN 7
 LEVANDA, 14 YELLOW WALLS RD, MALAHIDE
 FLORAVILLE COTTAGE, MONASTERY ROAD, CLONDALKIN
 APARTMENT 9 THE IVORY BUILDING, HANOVER STREET EAST, SIR JOHN ROGERSONS QUAY
 ASHTOWN LODGE, RIVER RD, ASHTOWN DUBLIN 15
 SOMALI, CARRICKBRACK ROAD, BAILY
 72 PARK AVENUE, SANDYMOUNT
 INDEPENDENT HOUSE AND BRETT COURT, TALBOT STREET
 69 HIGHFIELD RD, DUBLIN 6, DUBLIN
 WOODBERRY, THE BIRCHES, TORQUAY RD DUBLIN 18
 9 MAPLE ROAD, CLONSKEAGH
 42F PALMERSTON RD, DUBLIN 6, DUBLIN
 ST COLUMBANS, GRANGE ROAD, DONAGHMEDE
 THE LEYS, GLENAMUCK RD, CARRICKMINES DUBLIN 18
 RACHRA, COAST RD, MALAHIDE
 FLAT 1, 97 NORTH CIRCULAR RD, DUBLIN 7
 ORMISTON, CHURCH RD, MALAHIDE
 QUICHON, CEANCHOR RD, BAILY DUBLIN 13
 1 DARTMOUTH ROAD, LESSON PARK, DUBLIN 6
 FLAT 1, 46 SOUTH CIRCULAR ROAD, D8
 TANGLEWOOD, FALLS RD, RATHMICHAEL

ANYA, ANASTASIA LANE, SORRENTO RD
 7 CHAMLEY PARK, SWORDS RD, MALAHIDE
 2 GRANGE MANOR RD, RATHFARNHAM, DUBLIN 16
 HALL FLOOR APT, 15 ROYAL TERRACE EAST, DUN LAOGHAIRE
 5 CAIRNBOOK, CARRICKMINES, DUBLIN
 HAVEN, CROSTHWAITE PARK SOUTH, DUN LAOGHAIRE
 1 FARNEY PARK, SANDYOUNT, DUBLIN
 103 BRAEMOR RD, CHURCHTOWN, DUBLIN 14
 8 WYNNSWARD PARK, CLONSKEAGH, DUBLIN 14
 9 SANDYOUNT CASTLE COURT, SANDYOUNT
 4 GROVE HOUSE GARDENS, GROVE AVENUE, STILLORGAN
 1 PROBY PLACE, PROBY SQUARE, CARYSFORT AVENUE
 19 DEERPARK DRIVE, CASTLEKNOCK
 SAN JOSE, HAZELBROOK, MALAHIDE
 10 A BALLYBRIDE, RATHMICHAEL, DUBLIN 18
 DUNSANDLE MEWS, CORBAWN LANE, SHANKILL DUBLIN 18
 PADUA, CARRICKHILL ROAD UPPER, PORTMARNOCK
 9 A HOME LEIGH, CASTLE KNOCK
 4 ROEBUCK AVE, MOUNT MERRION, DUBLIN
 10 HILLFIELD, FERNDALE ROAD, RATHMICHAEL
 54 DEERPARK RD, CASTLEKNOCK, DUBLIN 15
 26 ROSMEEN GARDENS, DUN LAOGHAIRE, DUBLIN
 MUTIARA, 5 KILGOBBIN HEIFGTS, STEPASIDE
 153 TEMPLEOGUE RD, TERENURE, DUBLIN 6W
 3 BALURE, BALURE LANE, CHURCH ROAD
 5 MEADOW AVENUE, DUNDREUM, DUBLIN 16
 HOMESTEAD, 19 RATHDOWN PARK, TERENURE
 9 DEERPARK LAWN, CASTLEKNOCK, DUBLIN 15
 1 DEERPARK CLOSE, CASTLEKNOCK, DUBLIN 15
 11 BRAEMOR PARK, CHURCHTOWN, DUBLIN 14
 STOCKHOLE, CLOUGHRAIN, DUBLIN
 6 PECKS LANE, CASTLEKNOCK
 16 CLAREMONT PINES, CARRICKMINES, DUBLIN 18
 CLAREHAVEN, DUBLIN ROAD, MALAHIDE
 10 LAMBOURNE WOOD, BRENNANSTOWN RD, CABINTEELY DUBLIN 18
 6 SYCAMORE GREEN, THE PARK, CABINTEELY DUBLIN 18
 6 6A & 6B SYCAMORE GREEN, THE PARK, CABINTEELY DUBLIN 18
 6 LOWER CHURCHTOWN ROAD, MILLTOWN, DUBLIN 14
 20 CLARINDA PARK WEST, DUN LAOGHAIRE, CO DUBLIN
 48 CRANNAGH PARK, RATHFARNHAM, DUBLIN 14
 12 DORNDEN PARK, Booterstown, CO DUBLIN
 2 CLAREMONT PINES, CARRICKMINES, DUBLIN 18
 ELMHURST, BALLYBRIDE RD, RATHMICHAEL
 APARTMENT 4 WESTFIELD, BLOOMFIELD HOUSE, BLOOMFIELD AVENUE
 10A GLINBURY, WHITECHURCH ROAD
 23 BALLYMUN RD, GLASNEVIN, DUBLIN 9
 27 LEINSTER LAWN, CLONSKEAGH, DUBLIN 14
 29 HAMPTON PARK, Booterstown, DUBLIN
 20 CASTLEKNOCK LODGE, CASTLEKNOCK, DUBLIN 15
 3 BERESFORD GREEN, GRIFFITH AVE, DRUMCONDRA DUBLIN 9
 THE APPLEHOUSE, HOLMWOOD, CABINTEELY DUBLIN 18
 41 HEIDELBERG, ARDILEA, CLONSKEAGH
 24 PARKVIEW, CASTLEKNOCK, DUBLIN 15
 42 BRAEMOR RD, CHURCHTOWN, DUBLIN 14
 FAIRWINDS, 53 QUAY ST, SKERRIES
 1 CHURCH COTTAGE, CHURCH RD, MALAHIDE
 73 RATHDOWN PARK, TERENURE, DUBLIN 6W
 SANTESSA, OLD BRAY RD, CABINTEELY
 7 CARRICKMINES DALE, CARRICKMINES WOOD, DUBLIN 18
 8 SALAMANCA, ARDILEA, DUBLIN 14

```
/var/folders/bn/6x11zkt6n14vfwyb0jymgt00000gp/T/ipykernel_23318/3945560878.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
```

```
for address in df.nlargest(50000, 'Price')[ (df['County'] == 'Dublin') & (df['Price'] >= p_99) &
```

In [147...]

```
error_properties = ['INDEPENDENT HOUSE AND BRETT COURT, TALBOT STREET', 'CONVENT OF MERCY, BEAUMONT, DUBLIN 9',  

'HAMPTON MONASTERY, GRACEPARK ROAD, DRUMCONDRA', 'MARIST COMMUNITY, MOUNT ST. MARY\S, MILL'  

'11 MOYOLA COURT, CHURCHTOWN RD LOWER, DUBLIN 14', 'THE OAKS, PORTERSTOWN RD, CLONSILLA DUBLIN 14',  

'THE MEWS, CITYWEST, DUBLIN 9', '2 CHERRYGARTH, MOUNT MERRION, DUBLIN 9', 'LINN DUBH, HAYSTOWN 1',  

'ST ANNES, SYDENHAM RD, DUNDREUM DUBLIN 14', '34 MAYOR STREET UPPER, DUBLIN 1',  

'SAXAROON, GLENAMUCK RD, CARRICKMINES DUBLIN 18', 'BRETT COURT, FOLEY STREET, DUBLIN 1',  

'2 DISWELLSTOWN HEATH, CASTLEKNOCK, DUBLIN 14', '4 DISWELLSTOWN HEATH, CASTLEKNOCK',  

'PROPERTIES AT BREMORE PASTURES, BALBRIGGAN, CO DUBLIN 14', '18 LOMOND AVE, DUBLIN 3, DUBLIN 14',  

'1 HARBOUR HOUSE, BOND ST, DUBLIN 8', 'INDEPENDENT HOUSE AND BRETT COURT, TALBOT STREET',  

'ST COLUMBANS, GRANGE ROAD, DONAGHMEDE', '2 BELMONT, GARDINER ST, DUBLIN 1',  

'CONVENT OF MERCY, BEAUMONT, DUBLIN 9',  

'GENAZZANO, HAROLDS GRANGE ROAD, RATHFARNHAM', 'GLENMALURE, CASTLEKNOCK ROAD, CASTLEKNOCK',  

'UNIT 1, 416 HAROLDS CROSS ROAD, DUBLIN 6W', 'KARUNA, SANDYFORD RD, SANDYFORD DUBLIN 18',  

'BALROY, CARPENTERSTOWN RD, CASTLEKNOCK DUBLIN 15',  

'DRUMKEEN, GLENAMUCK RD, DUBLIN 18',  

'1 MILLRACE COURT, CLONSKEAGH RD, CLONSKEAGH', '24A 25 25A, HILL STREET, DUBLIN 1',  

'4 EDWARD SQ, BLOOMFIELD AVE, DUBLIN 14', '1A LINN DUBH, HAYSTOWN RD, RUSH',  

'1 WOLFE TONE COURT, 53 MARY ST, DUBLIN 1',  

'ST. JOSEPH\S, TIVOLI ROAD, DUN LAOGHAIRE', 'OLDMEADOW, PRIORY AVE, STILLORGAN',
```

```

'THORNHILL, CHERRY GARTH, MOUNT MERRION', 'THE LODGE, BESIDE ST MARYS CHURCH, CLONSILLA DUBLIN',
'PALMERSGATE, KENNELSFORT ROAD LOWER, PALMERSTOWN',
'DUN NA RI, SWORD RD, MALAHIDE', '14 WARRINGTON, PLACE', 'CALDERWOOD COURT, CALDERWOOD RD AND UNIT 13 KENILWORTH ROAD, DUBLIN 6',
'GRAGARNAGH, CARPENTERSTOWN RD, CASTLEKNOCK', '1 ST. AGNES CRESCENT, ST. AGNES ROAD, CRUMLINE',
'KILLAMONAN HOUSE, THE WARD, DUBLIN', 'DERRIANA, CASTLEKNOCK ROAD, DUBLIN 15',
'UNIT 1, 110 SEVILLE PLACE, DUBLIN 1', '3 CNOC DUBH CLOSE, 3 CNOC DUBH LAWN, 4 5 & 8 CNOC DUBLIN',
'BEDFORD HOUSE, CHURCH STREET, BALBRIGGAN', '71 GRAYDON CRESCENT, NEWCASTLE, DUBLIN',
'MARYFIELD, SCHOLARSTOWN RD, TEMPLEOGUE DUBLIN 16', 'BLOOMFIELD HOUSE, CAR SPACES 11 AND 12',
'WESTMINISTER LODGE MEWS, WESTMINISTER GROVE, OLD BRAY RD', '75 DORSET STREET, DUBLIN 1',
'8 NORTH FREDERICK ST, DUBLIN 1, DUBLIN', '\THE LAURELS\ MAIN STREET, TALLAGHT',
'73/73A SWEETMOUNT AVENUE, DUNDRUM', 'ASHFIELD HOUSE, 5 CLONSKEAGH ROAD, CLONSKEAGH',
'FUMBALLY, FUMBALLY LANE', 'HOLY FAITH CONVENT, ST BRIGID\ S ROAD, KILLESTER',
'2 FRANKFORT CASTLE, FRANKFORT, DUNDRUM DUBLIN 14', '17/17A LONGWOOD AVENUE, SOUTH CIRCULAR',
'ST ANNES CONVENT, KILMACUD ROAD UPPER, STILLORGAN', '1 MOBHI COURT, GLASNEVIN, DUBLIN 9',
'76 UPPER GARDINER ST, DUBLIN 1', '6 TIVOLI TERRACE EAST, DUN LAOGHAIRE, DUBLIN',
'3RD FLOOR, ST MONICA\ S NURSING HOME, 28 BELVEDERE PLACE', '21 HARRINGTON STREET, DUBLIN 8',
'BRAMBLEDOWN, GLENAMUCK RD, CARRICKMINES', 'SCHOOL PREMISES, CITYWEST', 'GEORGE\ S COURT, TOWER',
'ST ANNES CONVENT, UPPER KILMACUD RD, STILLORGAN', 'DUCALLA, CASTLEKNOCK RD, CASTLEKNOCK',
'1 FITZROY COURT, FITZROY AVE, DUBLIN 9', 'WENDOVER, 2 BEECHPARK AVENUE, CASTLEKNOCK',
'WYNNDHAM COURT, CHANCERY LANE, DUBLIN 8', '21 LAD LANE &, 3RD FLOOR 21 FITZWILLIAM PLACE, DUBLIN 2',
'7 CNOC DUBH WALK, 16 17 18 CNOC DUBH GREEN, BALLYBOUGHAL', 'GLENAYON, 46 BALLYMUN ROAD, DUBLIN',
'GROVE MEWS, REAR OF 8-12 MONK PLACE, PHIBSBORO', '32A 32B McCUALEY AVENUE, 55A LEIN ROAD,
'SITE 4, MALIN, BALLYGIHEN AVENUE', 'KINVARA, LEOPARDSTOWN RD, DUBLIN 18', 'STOCKHOLE, CLOGHOREE',
'6 6A & 6B SYCAMORE GREEN, THE PARK, CABINTEELY DUBLIN 18'
]

# 189 KILLINARDEN EST, TALLAGHT, DUBLIN
# 18 CASTLEGRANGE GARDENS, CASTLEGRANGE, DUBLIN & ST COLUMBAS, CHURCH AVE, BALLYBRACK
# 8 CLORAGH MILLS, EDMONDSTOWN RD, RATHFARNHAM DUBLIN 16 & 39 ELLENSBOROUGH DRIVE, KILTIPPER, TALLAGHT
# 120 COLLEGE VIEW, BALLYMUN, DUBLIN 11 & 97 LATCHFORD SQ, CASTAHEANY, DUBLIN 15
# 6 LIMEKILN COURT, WALKINSTOWN, DUBLIN 12 & 64 CAMAC WALK, 7 MOUNT BROWN, KILMAINHAM
rows = [327077, 245243, 162928, 575205, 12246, 390741, 554451, 383782, 571001]

# df = df.drop(index=rows)
df = df.drop(index=df[df['Address'] == ('ABBERLEY SQUARE, HIGH STREET, TALLAGHT')].index)
df = df.drop(index=df[df['Address'] == ('19 VICTORIA ST, SOUTH CIRCULAR RD')].index)
df = df.drop(index=df[df['Address'] == ('WATERSIDE ESTATE, MALAHIDE')].index)
df = df.drop(index=df[df['Address'] == ('GRACEPARK HOUSE, DRUMCONDRA')].index)
df = df.drop(index=df[df['Address'] == ('JOHNSTOWN RD, DUN LAOGHAIRE, DUBLIN')].index)
df = df.drop(index=df[df['Address'] == ('ARD EADRÁD, BLAKESTOWN ROAD, MULHUDDART')].index)

for address in error_properties:
    df = df.drop(index=df[df['Address'].str.contains(address)].index)
# drop rows
df = df.reset_index(drop=True)
df.shape

```

Out[147]: (591536, 8)

- Recheck Price data with outliers removed

```
In [7]: # show Price statistics
df['Price'].describe().apply(lambda x: format(x, 'f'))
```

```
Out[7]: count      591536.000000
mean      260060.541003
std       247532.349081
min       5001.000000
25%      125991.187500
50%      212067.750000
75%      320000.000000
max      14250000.000000
Name: Price, dtype: object
```

```
In [8]: # check the price threshold of the highest 99% of prices and the number of rows
p_99 = np.percentile(df['Price'], 99)
p_99_rows = (df['Price'] < p_99).sum()
p_1_rows = df['Price'].size - (df['Price'] < p_99).sum()

# print the results
print("99% of data is <= " + str(p_99) + ', containing ' + str(p_99_rows) + ' rows')
print("Top 1% percent of data is <= " + str(p_99) + ', containing ' + str(p_1_rows) + ' rows')
```

99% of data is <= 1145000.0, containing 585610 rows
Top 1% percent of data is <= 1145000.0, containing 5926 rows

- After removing outliers:
 - We have identified and removed 4,059 properties from the full dataset.
 - The maximum price of a property has been reduced from €225,000,000 to €14,250,000.
 - There are 83 less properties in the most expensive 1%.

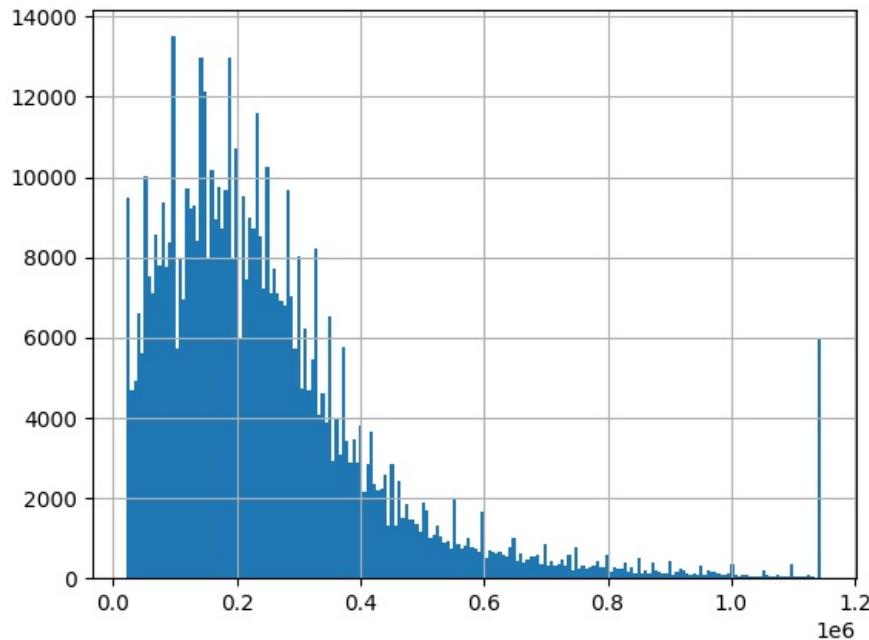
- 99% of data is now less than or equal to €1,145,000 compared to €1,275,000 previously.

- It makes sense to remove properties above the top 1% solely for the purpose of viewing both the histogram and box plot in order to better understand the data distribution.

```
In [9]: #clamp data to 99%
data_clamped = df['Price'].clip(lower=df['Price'].quantile(0.01), upper=df['Price'].quantile(0.99))
```

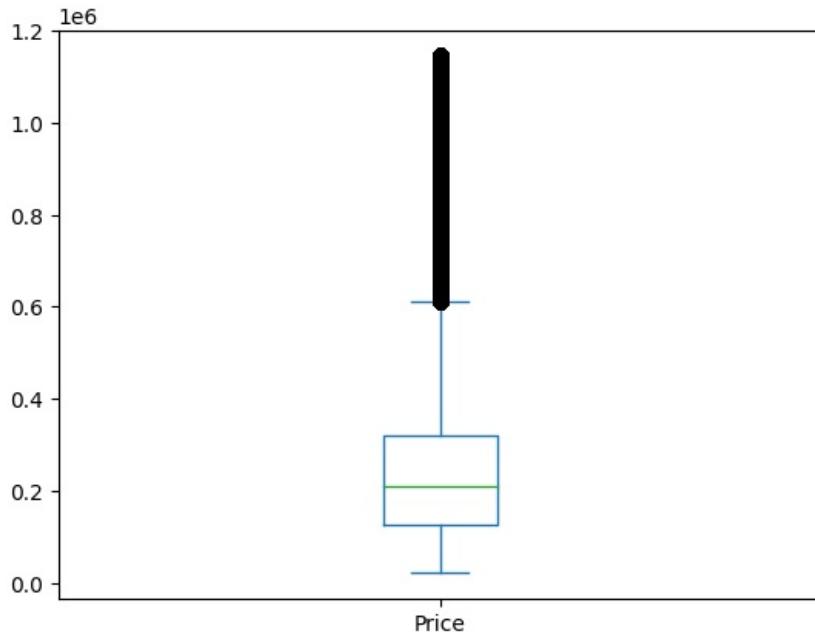
```
In [10]: #show the clamped price data as a histogram
data_clamped.hist(bins=200)
```

Out[10]: <Axes: >



```
In [11]: #show the clamped price data as a boxplot
data_clamped.plot(kind='box')
```

Out[11]: <Axes: >



- Viewing the clamped data as a histogram, we can see that Price data is negatively skewed. From the boxplot we can still see there are outliers but we have a better understanding of the interquartile range and median compared to the maximum and minimum prices. The median price for a property appears to be around €200,000.

Summary of data quality plan:

Variable Names	Data Quality Issue	Handling Strategy
DateofSale(dd/mm/yyyy)	----	----

Address	----	----
Eircode	Missing Values	Drop feature
County	----	----
Price	Outliers	Drop rows
Price	Irrelevant Data	Drop rows
NotFullMarketPrice	----	----
VATExclusive	----	----
DescriptionofProperty	----	----
PropertySizeDescription	Missing Values	----

- Save the data quality plan to a single PDF file.

```
In [ ]: # saving cleaned dataframe to a csv
df.to_csv("PPR-ALL_cleaned_section2.csv")
```

3 - Exploring relationships between feature pairs

```
In [80]: # load cleaned data
# file should be "PPR-ALL_cleaned_section2.csv"
df = pd.read_csv("PPR-ALL_cleaned_section3.csv", index_col=0)
df.dtypes
```

```
Out[80]: DateofSale          object
Address            object
County             object
Price              float64
NotFullMarketPrice    object
VATExclusive       object
DescriptionofProperty    object
PropertySizeDescription    object
dtype: object
```

```
In [67]: # # convert feature types
# df['DateofSale'] = df['DateofSale'].astype('datetime64[ns]')
# df['County'] = df['County'].astype('category')
# df['NotFullMarketPrice'] = df['NotFullMarketPrice'].astype('category')
# df['VATExclusive'] = df['VATExclusive'].astype('category')
# df['DescriptionofProperty'] = df['DescriptionofProperty'].astype('category')
# df['PropertySizeDescription'] = df['PropertySizeDescription'].astype('category')
# df.dtypes
```

```
Out[67]: DateofSale          datetime64[ns]
Address            object
County             category
Price              float64
NotFullMarketPrice    category
VATExclusive       category
DescriptionofProperty    category
PropertySizeDescription    category
dtype: object
```

Choose a subset of features you find promising and plot pairwise feature interactions

- Categorical vs Categorical

- First I will compare all of the categorical features I find promising with the DescriptionofProperty column.

```
In [71]: import matplotlib.patches as mpatches
# Function to generate Stacked Bar Charts for Categorical Columns

def stacked_bar_chart(col):
    # will check the percentage of new vs old properties sold in the selected column by values in the column
    array = pd.unique(df[col].ravel())
    df['percent'] = 0

    for i in array:
        count = df[df[col] == i].count()['DescriptionofProperty']
        count_percentage = (1 / count) * 100
        print(i, "\t", count)
        index_list = df[df[col] == i].index.tolist()
        for ind in index_list:
            df.loc[ind, 'percent'] = count_percentage
```

```

group = df[['percent', col, 'DescriptionofProperty']].groupby([df[col], 'DescriptionofProperty']).sum()

my_plot = group.unstack().plot(kind='bar', stacked=True, title="Description of Property based on " + str(col))

blue_patch = mpatches.Patch(color='blue', label='New Dwelling house /Apartment')
red_patch = mpatches.Patch(color='orange', label='Second-Hand Dwelling house /Apartment')
my_plot.legend(handles=[red_patch, blue_patch], frameon = True)

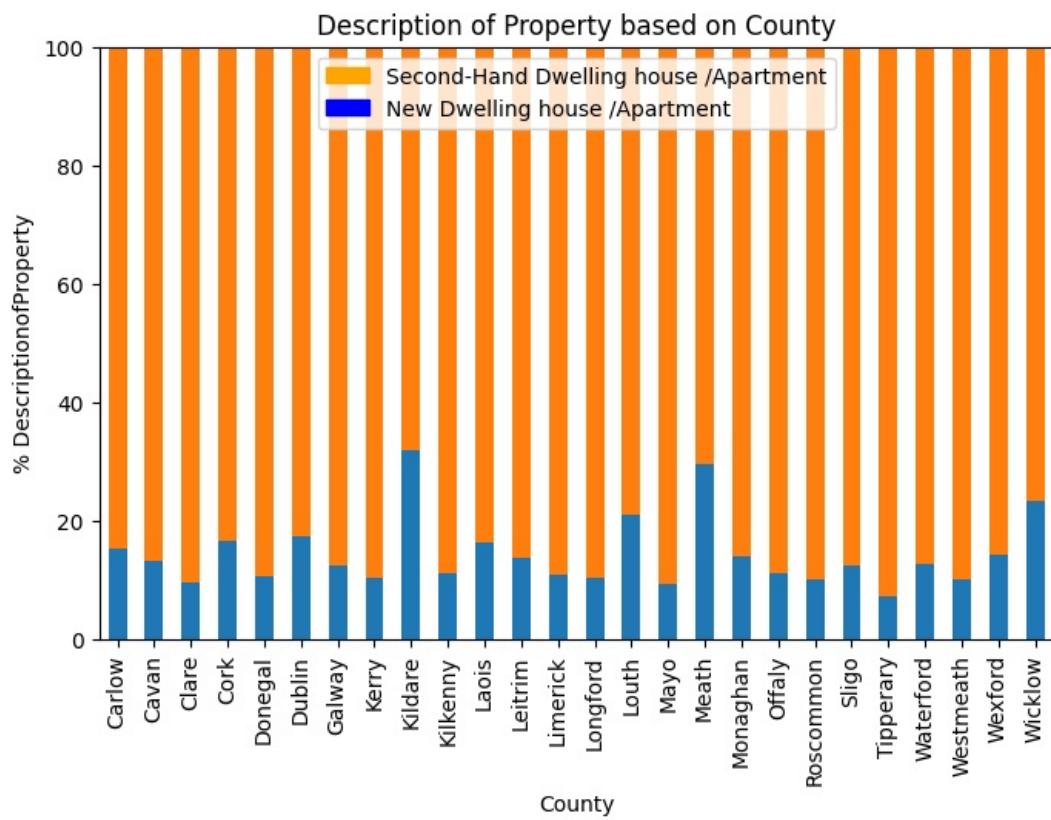
my_plot.set_xlabel(col)
my_plot.set_ylabel("% DescriptionofProperty")
my_plot.set_ylim([0,100])

df.drop("percent", axis=1, inplace = True)

```

In [72]: #check new vs old homes sold based on county
stacked_bar_chart("County")

Dublin	185607
Laois	9260
Meath	24202
Kilkenny	9291
Limerick	22024
Carlow	6574
Cork	64646
Clare	13571
Sligo	8882
Cavan	9064
Tipperary	16014
Wicklow	19011
Roscommon	8604
Wexford	20564
Mayo	14284
Donegal	16205
Longford	5151
Galway	29289
Offaly	7396
Kildare	31309
Waterford	16200
Louth	15990
Kerry	16981
Westmeath	11391
Monaghan	4669
Leitrim	5357

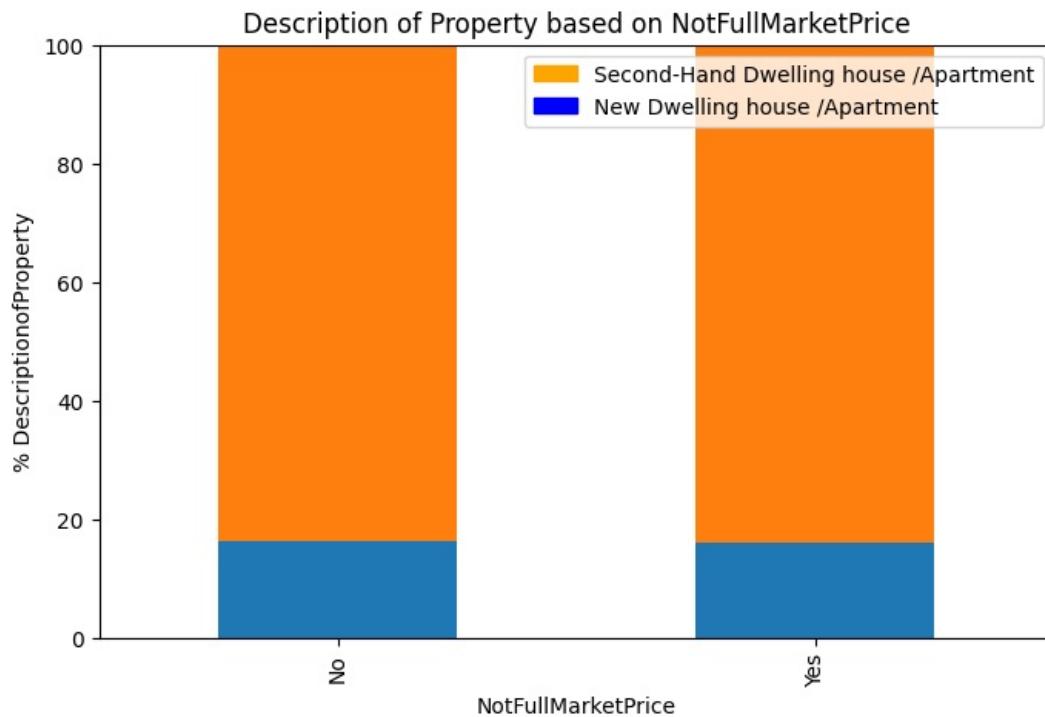


- I suspected the highest proportion of new to old properties sold would be in either Dublin or Cork since they are the most populous counties. This does not appear to be the case as Kildare, Meath and Wicklow had the highest amount of new homes sold compared to second hand houses, each one neighbouring Dublin.

In [73]: #check new vs old homes based on whether they are sold for full market price

```
stacked_bar_chart("NotFullMarketPrice")
```

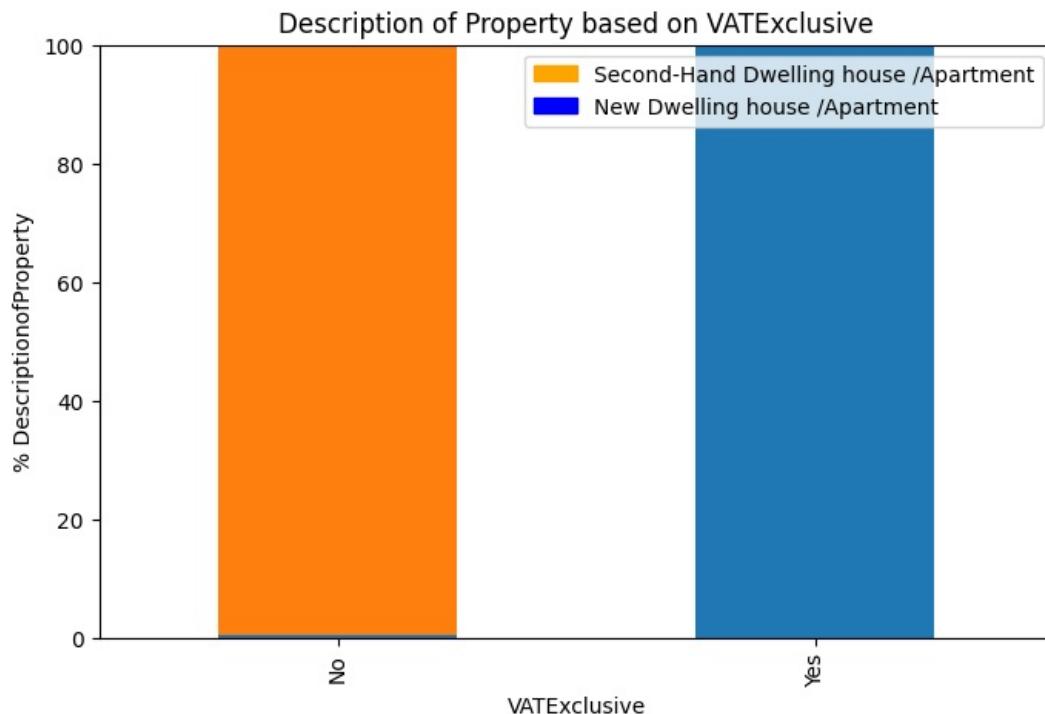
No	563161
Yes	28375



- The number of new and second-hand properties sold for full market price is similar and so the age of the property does not appear to affect whether the seller will get full price for it.

```
In [74]: # check new vs old homes based on whether their prices include VAT  
stacked_bar_chart("VATExclusive")
```

No	496650
Yes	94886



- Almost all new properties do not have VAT included in the price.

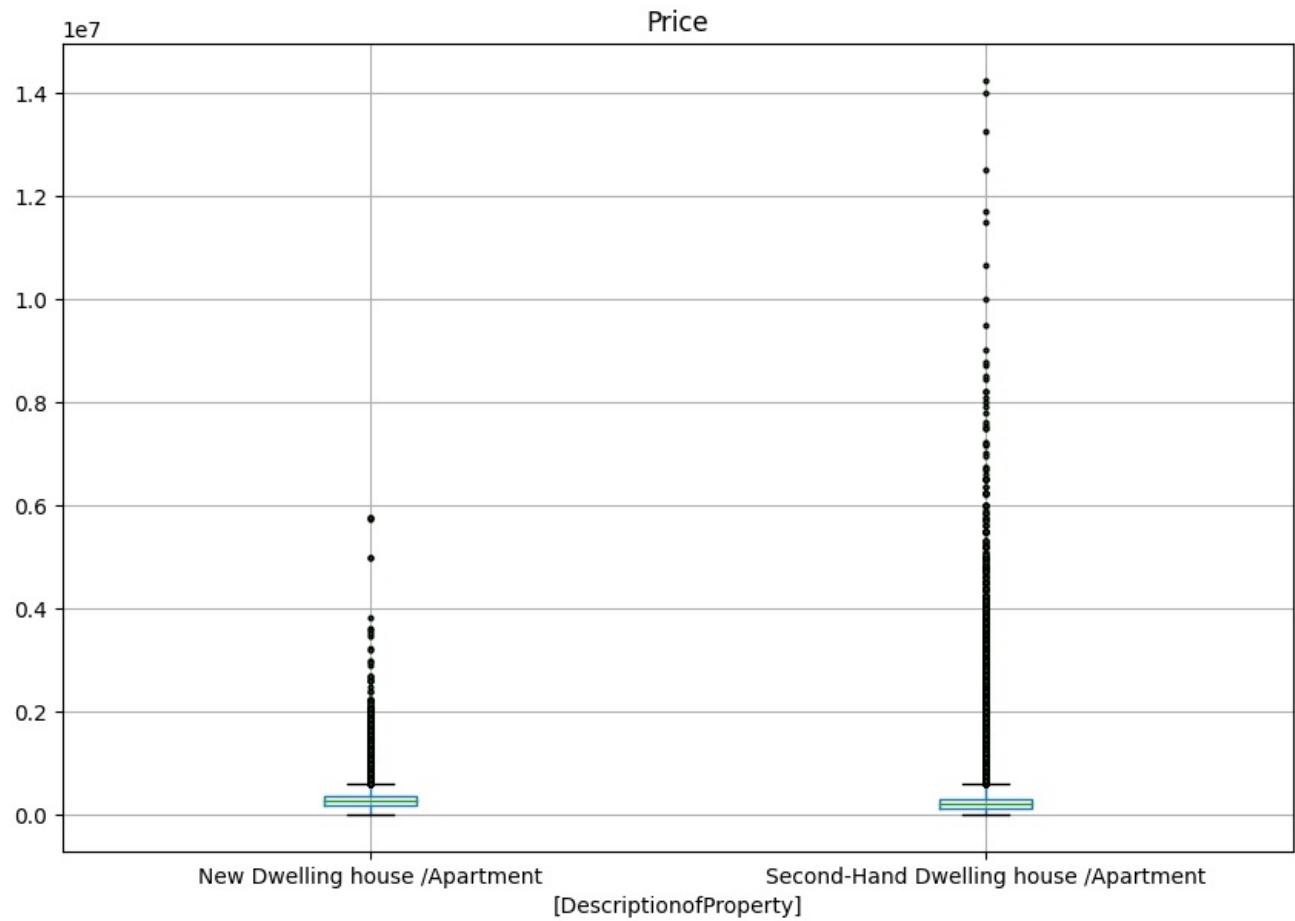
- Categorical vs Continuous

```
In [75]: # code for box plot Price vs DescriptionofProperty  
plt.figure()  
flierprops = dict(marker='o', markerfacecolor='green', markersize=2,  
                  linestyle='none')  
df.boxplot(column=['Price'], by=['DescriptionofProperty'], flierprops=flierprops, figsize=(10,7))
```

```
Out[75]: <Axes: title={'center': 'Price'}, xlabel='[DescriptionofProperty]'>
```

```
<Figure size 640x480 with 0 Axes>
```

Boxplot grouped by DescriptionofProperty



- The box plot above shows a similar interquartile range for price, while second-hand houses appear to have a higher number of outliers over a larger range. This would suggest that price is not dependent on the age of the property, while there could be factors exclusive to second-hand houses that could cause a higher price, such as being located in desirable areas to live where new homes are not being built.

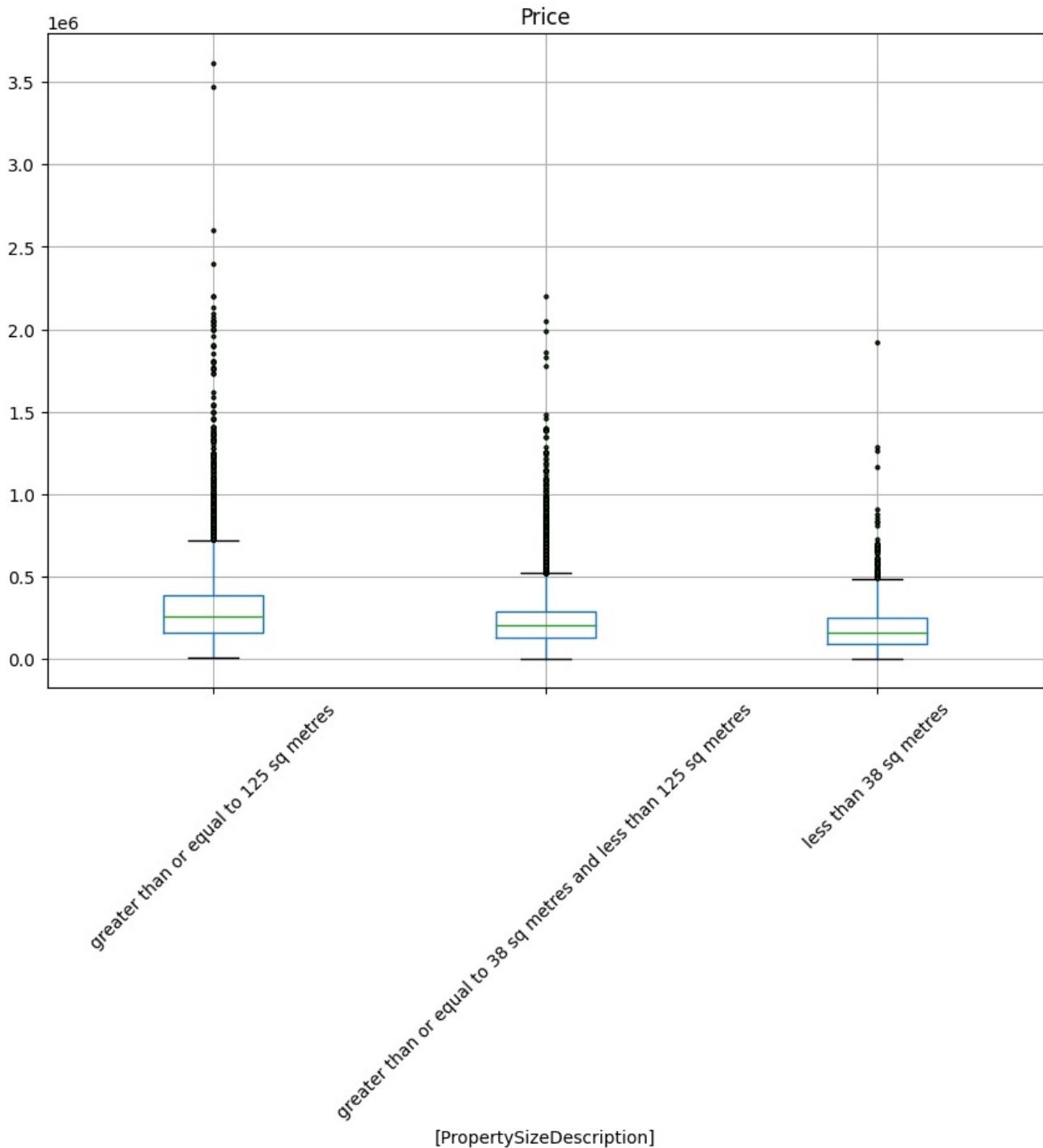
```
In [76]: # code for box plot Price vs PropertySizeDescription
```

```
plt.figure()
flierprops = dict(marker='o', markerfacecolor='green', markersize=2,
                  linestyle='none')
df[df['PropertySizeDescription'] != 'other'].boxplot(column=['Price'], by=['PropertySizeDescription'], rot=45,
```

```
Out[76]: <Axes: title={'center': 'Price'}, xlabel='[PropertySizeDescription]'>
```

```
<Figure size 640x480 with 0 Axes>
```

Boxplot grouped by PropertySizeDescription

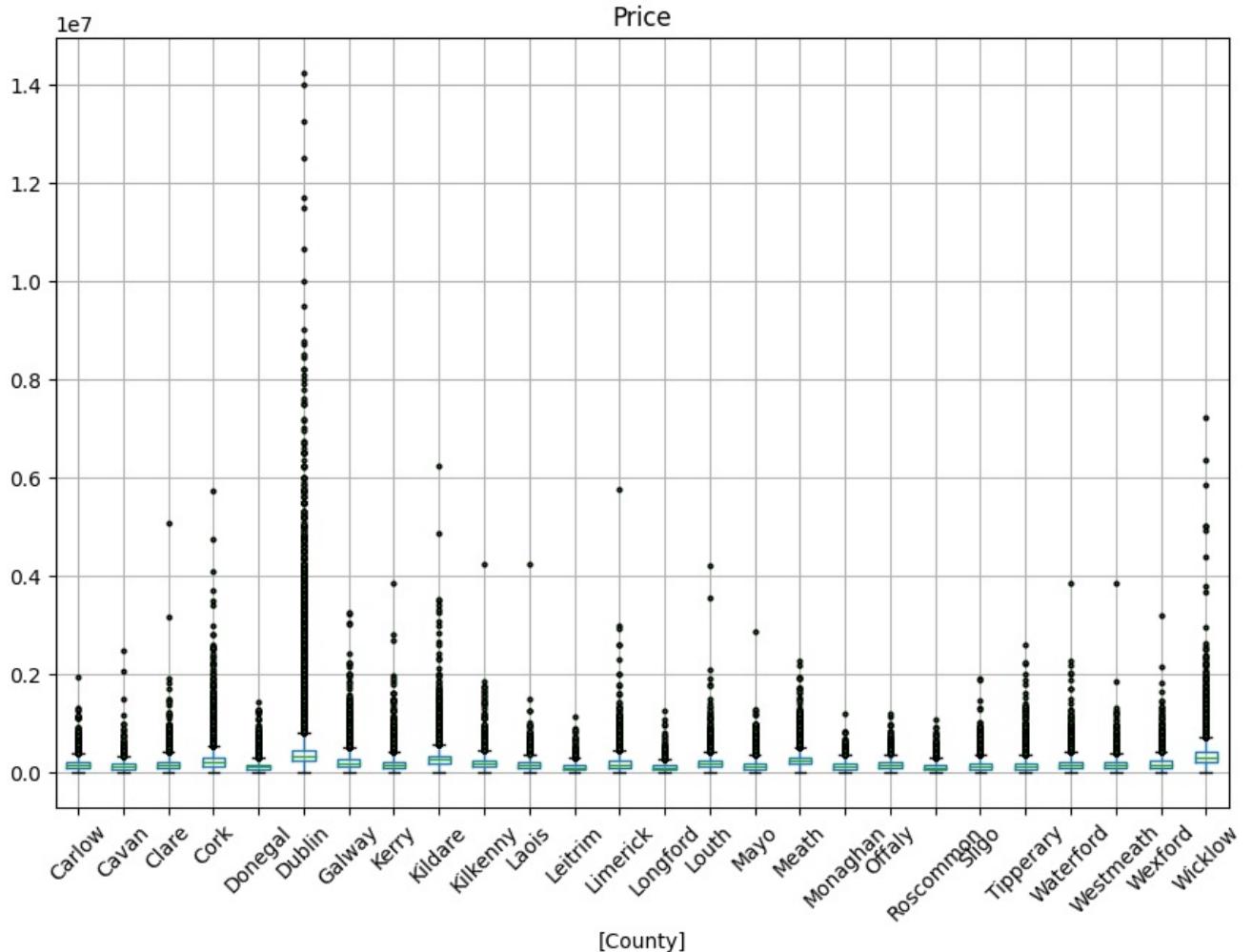


- As expected, house size has an influence on the price, with smaller properties having a lower median price, while larger properties have the largest median price. Outliers in price may suggest competitive buying, or high demand, meaning that other factors such as location influence prices.

```
In [77]: # code for box plot Price vs County
plt.figure()
flierprops = dict(marker='o', markerfacecolor='green', markersize=2,
                  linestyle='none')
df.boxplot(column=['Price'], by=['County'], rot=45, flierprops=flierprops, figsize=(10,7))
```

```
Out[77]: <Axes: title={'center': 'Price'}, xlabel='[County]'>
<Figure size 640x480 with 0 Axes>
```

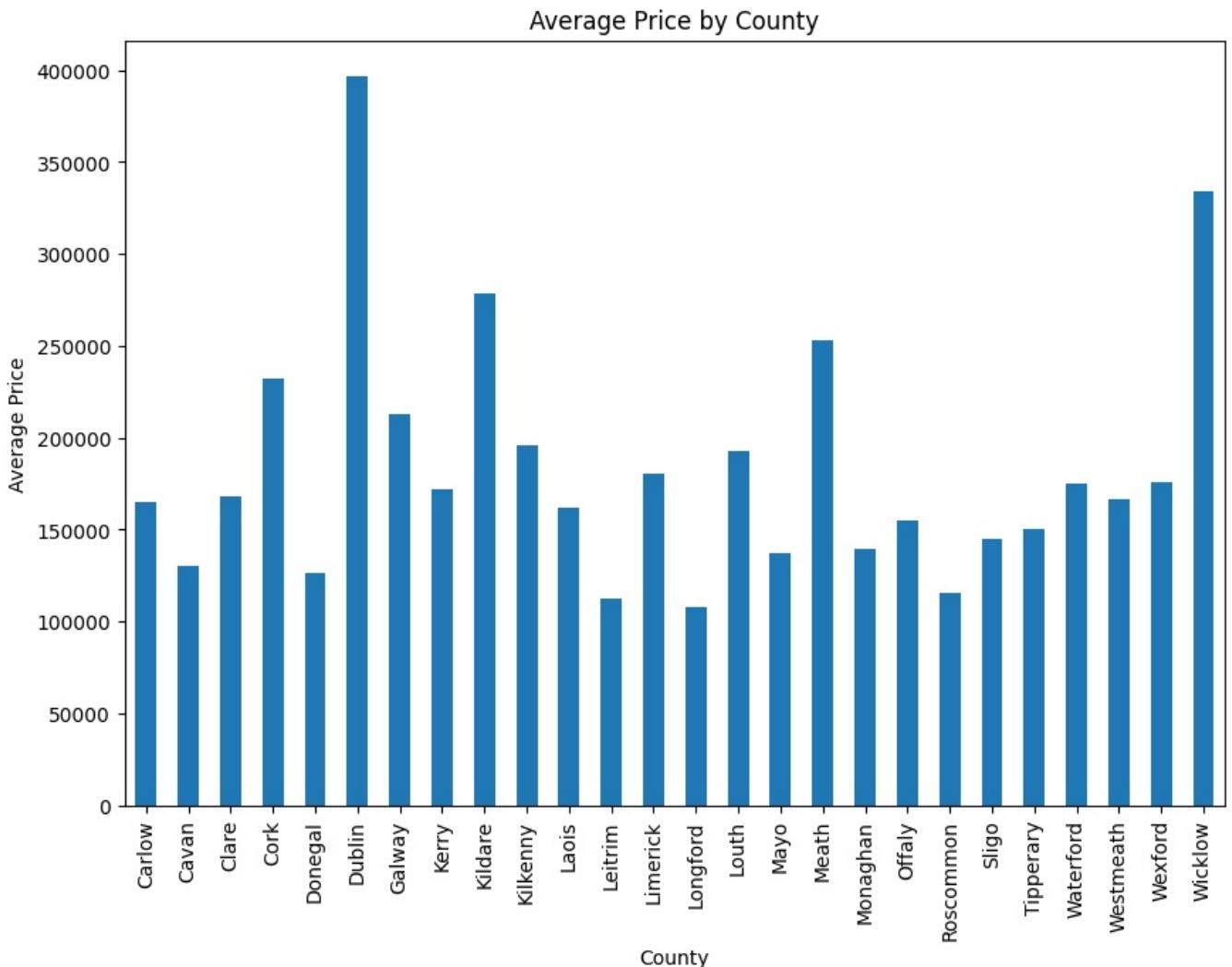
Boxplot grouped by County



- Dublin and Wicklow property prices have the largest interquartile range than other counties, and have the highest number of outliers meaning the demand for houses in each county is quite high, with a high variability in prices. It may be useful to look at average prices for each county.

```
In [78]: # bar chart of mean prices per county
plt.figure(figsize=(10,7))
df[["Price"]].groupby(df['County']).mean().plot(kind='bar')
plt.title("Average Price by County")
plt.ylabel("Average Price")
plt.xlabel("County")
```

```
Out[78]: Text(0.5, 0, 'County')
```



- Here we can clearly see Dublin has the most expensive average price, with three neighbouring counties being the next highest, Wicklow, Kildare and Meath. Cork is the fourth highest, while not geographically close to Dublin, it is the second most populous county in Ireland. The average price per county seems to indicate that properties in, or close to, Dublin have the highest value.

- Discuss your findings from the plots above.

Categorical vs Categorical:

- Over 80% of properties sold since 2010 have been second-hand. Dublin has the highest amount of properties sold, implying that there is a high demand for Dublin properties, yet it doesn't have one of the highest percentages of new houses sold. The two counties with the rates for new properties are Kildare and Meath, which neighbour Dublin, suggesting that if people can't get housing in Dublin, they will try to live nearby.
- The age of a property does not have an effect on whether a seller will get full market price for it.
- The prices of new properties do not have VAT included in the price.

Categorical vs Continuous:

- There does not appear to be a relationship to the price of a property and its age.
- Property prices appear to increase with the size of the property.
- There appears to be a relationship between the price a property and its location in relation to Dublin, with surrounding

counties have the next highest average prices per county after Dublin.

4 - Transform, extend or combine the existing features to create a few new features

```
In [81]: # load cleaned data
# file should be "PPR-ALL_cleaned_section2.csv"
df = pd.read_csv("PPR-ALL_cleaned_section3.csv", index_col=0)
df.shape
```

```
Out[81]: (591536, 8)
```

New features we will implement:

- Postal Code
- Month of Sale
- Year of Sale
- Geolocation

Feature 1

- The first feature I want to create is one which shows which postal code each address was sold. This will only be possible for Dublin properties since it is the only county with multiple postal codes. Since Dublin has the most expensive average property prices, it will be interesting to see if there is a correlation between property prices and certain areas within Dublin.

```
In [82]: postal_codes = ['DUBLIN 1', 'DUBLIN 2', 'DUBLIN 3', 'DUBLIN 4', 'DUBLIN 5', 'DUBLIN 6', 'DUBLIN 6 W',
                    'DUBLIN 6W', 'DUBLIN 7', 'DUBLIN 8', 'DUBLIN 9', 'DUBLIN 10', 'DUBLIN 11', 'DUBLIN 12', 'DUBLIN
                    'DUBLIN 14', 'DUBLIN 15', 'DUBLIN 16', 'DUBLIN 17', 'DUBLIN 18', 'DUBLIN 20', 'DUBLIN 22',
                    'DUBLIN 24', 'D 1', 'D 2', 'D 3', 'D 4', 'D 5', 'D 6', 'D 6 W', 'D 7', 'D 8', 'D 9',
                    'D 10', 'D 11', 'D 12', 'D 13', 'D 14', 'D 15', 'D 16', 'D 17', 'D 18', 'D 20',
                    'D 22', 'D 24', 'CO. DUBLIN', 'CO DUBLIN', 'COUNTY DUBLIN']
df['PostalCode'] = np.nan

for row in range(0, len(df)):
    if df.loc[row]['County'] == 'Dublin':
        address = df.loc[row]['Address'].split(',')
        for code in postal_codes:
            if (code == address[-1]) | ((code.replace(" ", "") == address[-1])):
                print(row, code)
                df.at[row, 'PostalCode'] = code
            elif (code in address[-1]) | ((code.replace(" ", "") in address[-1])):
                df.at[row, 'PostalCode'] = code
```

```
In [83]: df['County'].value_counts()
```

```
Out[83]: County
Dublin      185607
Cork        64646
Kildare     31309
Galway      29289
Meath       24202
Limerick    22024
Wexford     20564
Wicklow     19011
Kerry        16981
Donegal      16205
Waterford    16200
Tipperary    16014
Louth        15990
Mayo         14284
Clare        13571
Westmeath    11391
Kilkenny     9291
Laois        9260
Cavan        9064
Sligo        8882
Roscommon    8604
Offaly       7396
Carlow       6574
Leitrim      5357
Longford     5151
Monaghan     4669
Name: count, dtype: int64
```

```
In [84]: df[df['County'] == 'Dublin']['PostalCode'].isnull().sum()
```

Out[84]: 9677

- We can see that we were able to determine the postal codes of just under half of Dublin properties. I will look closer at the properties that have no postal code information.

```
In [85]: pd.set_option('display.max_rows', len(df))

df[(df['County'] == 'Dublin') & (df['PostalCode'].isnull())]['Address']
```

```
Out[85]: 5          12 SALLYMOUNT AVENUE, RANELAGH
13          23 ELTON PARK, SANDYCove
16          29 CARYSFORT WOODS, BLACKROCK
21          49 BALLYNAKELLY GREEN, NEWCASTLE
23          54., BRIDES GLEN AVENUE, SWORDS
42          THE MEWS, 11A WYCKHAM PARK ROAD, DUNDRUM
47          134 LONGBOAT QUAY NORTH, SIR JOHN ROGERTSONS QUAY
50          15 SPENCE, VILLAS, GLENGEAR
56          3 MYRTLE SQUARE, THE COAST, BALDOYLE
61          6 CLONASLEIGH LANE, SHANKILL
63          66 VILLA PARK GARDENS, NAVAN ROAD
65          87 MANGERTON RAOD, DRIMNAGH
67          APT 1 LINDEN COURT, BLACKROCK
76          11 TRITONVILLE COURT, SANDYMOUNT
87          58 PALMER PARK, BALLYBODEN, RATHFARNHAM
95          84 CASTLELAND PARK VIEW, BALBRIGGAN
97          9 COLTHURST GREEN, HUNTINGTON GLEN, LUCAN
115         15 GRIFFITH DRIVE, FINGLAS EAST
117         16 LANDSCAPE GARDENS, CHURCHTOWN
122         32 GRIFFEEN GLEN AVENUE, LUCAN
124         37 CASTLE COURT, KILLINEY HILL ROAD, KILLINEY
128         7 MOYCLARE AVENUE, BALDOYLE
133         UNIT 55 HEATHFIELD PARK, HEATHFIELD, CAPPAGH ...
134         109 DOWLAND ROAD, WALKINSTOWN
140         14 CASTLEVIEW PARK, MALAHIDE
141         14 ST. AONGHUS ROAD, TALLAGHT
142         15 BEECHDALE PLACE, BALLYCULLEN
143         15 DESELBY LAWNS, BLESSINGTON ROAD
149         2 CASTLEGATE HOUSE, ADAMSTOWN
152         21 CHARNWOOD GARDENS, CLONSILLA
153         21 MAPLE AVENUE, CASTLEKNOCK
155         22 GRANGEVIEW ROAD, CLONDALKIN
157         24 ABBEY PARK, KILLESTER AVENUE
158         26 SANDYFORD HALL PARK, SANDYFORD
160         27 NORTH PRIORY, NAVAN ROAD
162         29 SYCAMORE DRIVE, KINGSWOOD HEIGHTS
167         38 ELLENSBOROUGH COPSE, KILTIPPER, DUBLIN
172         44 BAYSIDE WALK, SUTTON
176         62 WATSON ROAD, KILLINEY
177         64 THE WAY, HUNTERS ROAD, CLONEE
178         67 BELMAYNE PARK, NORTH BELMAYNE, BALGRiffin
180         7 EDEN GROVE, DONABATE
183         75 LOWER BEECHWOOD AVENUE, RANELAGH
186         9 SUNDALE MEADOWS, MOUNTAINVIEW, TALLAGHT
188         APARTMENT 26, PORTABELLO HARBOUR, GROVE ROAD
209         10 HARRISON COVE, SKERRIES
210         108, CASTLEFIELD COURT, CLONSILLA
215         16 OAKDALE DRIVE, BALLYCULLEN, FIRHOUSE
225         29 MYRTLE HOUSE, LONGFIELD ROAD THE COAST, BA...
232         56 CHERRYWOOD PARK, CLONDALKIN
234         73 COLLEGE PARK, TERENURE
238         APARTMENT 18 THE GALLERY, TURVEY WALK, DONABATE
250         1 COIS INBHIR, DONABATE
261         21 FOXWOOD, SWORDS
262         25 BELTON PARK ROAD, DONNYCARNEY
266         32 KILL ABBEY, DEANSGRANGE, BLACKROCK
269         5 TEMPLE VIEW PARK, CLARE HALL
272         7 CASTLEVIEW GROVE, SWORDS
274         APARTMENT 15 DAWSON COURT, CROSS AVENUE, BLACK...
277         GLENDOWER, OLD STATION ROAD, PORTMARNOCK
289         11 BALALLY DRIVE, DUNDRUM
291         17 SWARNWARD COURT, PARNELL ROAD
296         231 SEAPARK, MALAHIDE, DUBLIN
298         26 THE CROFT, PARC NA SILLA LOUGHlinSTOWN
301         3 CASTLEGATE HOUSE, CASTLEGATE, ADAMSTOWN
302         3 MERRION PARK, SOUTH HILL AVENUE, BLACKROCK
304         36 THE MEADOWS, GRANGE ROAD, CLONSILLA
309         5 ROSS STREET, STONEYBATTER
313         APARTMENT 37 BLOCK D, CLARE VILLAGE, CLARE HALL
314         APARTMENT 8 THE SYCAMORE, ELMFIELD BALLYOGAN ...
330         13 ROS MOR VIEW, SCHOLARSTOWN ROAD, KNOCKLYON
337         279 CAPTAIN'S ROAD, CRUMLIN
```

```

591455    17 BARN ELMS, UPPER CHURCHTOWN ROAD, CHURCHTOWN
591459          2 MARINERS AVENUE, MARINERS WAY, RUSH
591462    21 CRONAN'S WELL GROVESWORDS, MOORETOWN, DUBLIN
591463          21 HALLIDAY SQ, DUBLIN CITY
591483          74 ARD NA MARA, MALAHIDE, DUBLIN
591484    8 THE LAURELS, RATHFARNHAM VILLAGE, RATHFARNHAM
591496   10 BARN ELMS, UPPER CHURCHTOWN ROAD, CHURCHTOWN
591507   23 LLEWELLYN CLOSE, GRANGE VALLEY, RATHFARNHAM
591521       67 WOODVIEW HEIGHTS, LUCAN, DUBLIN
591533      SPRING COTTAGE, SHALLON, THE WARD
Name: Address, dtype: object

```

```

In [86]: areas = {'DUBLIN 1' : ['ABBEY ST', 'AMIENS ST', 'CAPEL ST', 'DORSET ST', 'HENRY ST', 'MARY ST', 'MOUNTJOY SQ',
                           'MARLBOROUGH ST', 'NORTH WALL', 'O\CONNELL ST', 'PARNELL SQ', 'TALBOT ST'],
              'DUBLIN 2' : ['DAME ST', 'MERRION SQ', 'TEMPLE BAR', 'GRAFTON ST', 'STEPHEN\S GR', 'STEPHENS GR',
                           'LOWER LEESON ST', 'AUNGIER ST', 'WEXFORD ST', 'CAMDEN ST'],
              'DUBLIN 3' : ['BALLYBOUGH', 'NORTH STRAND', 'DOLLYMOUNT', 'CLONTARF', 'EAST WALL', 'FAIRVIEW', 'MARINO',
                           'CLONLIFFE', 'KILLESTER'],
              'DUBLIN 4' : ['UPPER LEESON ST', 'BALLSBIDGE', 'DONNYBROOK', 'IRISHTOWN', 'BELFIELD', 'MERRION', 'PEMI',
                           'RINGSEND', 'SANDYMOUNT'],
              'DUBLIN 5' : ['ARTANE', 'COOLOCK', 'HARMONSTOWN', 'KILBARRACK', 'KILLESTER', 'RAHENY'],
              'DUBLIN 6' : ['MILLTOWN', 'RANELAGH', 'TERENURE', 'RATHMINES', 'DARTRY', 'RATHGAR'],
              'DUBLIN 6W' : ['HAROLD\S CROSS', 'HAROLDS CROSS', 'HAROLDS CROSS', 'TEMPLEOGUE', 'KIMMAGE', 'TERENURE'],
              'DUBLIN 7' : ['ARBOUR HILL', 'ASHTOWN', 'BROADSTONE', 'CABRA', 'GRANGEGORMAN', 'OXMANTOWN', 'PHIBSBORO',
                           'SMITHFIELD', 'STONEYBATTER'],
              'DUBLIN 8' : ['DOLPHIN\S BARN', 'DOLPHINS BARN', 'INCHICORE', 'ISLANDBRIDGE', 'KILMAINHAM', 'MERCHANTS',
                           'PORTOBELLO', 'SOUTH CIRCULAR ROAD', 'SOUTH CIRCULAR RD', 'LIBERTIES', 'PHOENIX PARK'],
              'DUBLIN 9' : ['SHANGAN', 'COULTRY', 'BEAUMOUNT', 'DONNYCARNEY', 'DRUMCONDRA', 'ELM MOUNT', 'GRIFFITH A',
                           'SANTRY', 'WHITEHALL'],
              'DUBLIN 10' : ['BALLYFERMOT', 'SARSFIELD ROAD', 'SARSFIELD RD', 'CHERRY ORCHARD'],
              'DUBLIN 11' : ['SILLOGUE', 'BALBUTCHER', 'BALCURRIS', 'POPINTREE', 'SANDYHILL', 'WADELEI', 'DUBBER CROS',
                           'FINGLAS', 'CAPPAGH', 'CREMORE', 'ADDISON', 'VIOLET HILL', 'FINGLAS ROAD', 'FINGLAS RD',
                           'KILSHANE CROSS', 'THE WARD', 'COOLQUAY'],
              'DUBLIN 12' : ['BLUEBELL', 'CRUMLIN', 'TERENURE', 'KIMMAGE', 'DRIMNAGH', 'GREENHILLS', 'PERRYSTOWN',
                           'WALKINSTOWN'],
              'DUBLIN 13' : ['BALDOYLE', 'BAYSIDE', 'DONAGHMEDE', 'CLONGRIGGIN', 'SUTTON', 'HOWTH', 'AYRFIELD'],
              'DUBLIN 14' : ['CHURCHTOWN', 'CLONSKEAGH', 'DUNDRUM', 'GOASTOWN', 'RATHFARNHAM', 'WINDY ARBOUR'],
              'DUBLIN 15' : ['ASHTOWN', 'BLANCHARDSTOWN', 'CASTLEKNOCK', 'COOLMINE', 'CLONSILLA', 'CORDUFF', 'MULHUDI',
                           'TYRRELSTOWN', 'ONGAR', 'CLONEE'],
              'DUBLIN 16' : ['BALLINTEER', 'BALLYBODEN', 'KILMASHOGUE', 'KNOCKLYON', 'ROCKBROOK'],
              'DUBLIN 17' : ['BALGRIFFON', 'COOLOCK', 'BELCAMP', 'DARNDALE', 'PRIORSWOOD'],
              'DUBLIN 18' : ['CABINTEELY', 'CARRICKMINES', 'FOXROCK', 'KILTERNAN', 'SANDYFORD', 'SHANKILL', 'TICKNOCK',
                           'BALLYEDMONDUFF', 'STEPASIDE', 'LEOPARDSTOWN'],
              'DUBLIN 20' : ['CHAPELIZOD', 'PALMERSTOWN'],
              'DUBLIN 22' : ['CLONDALKIN', 'LIFFEY VALLEY', 'NEWCASTLE', 'NEILSTOWN'],
              'DUBLIN 24' : ['FIRHOUSE', 'JOBSTOWN', 'OLD BAWN', 'TALLAGHT', 'BALLYMOUNT']]
```

```

In [87]: for code in areas:
    for c in areas[code]:
        print('*****')
        print(c)
        print('Number of addresses in this area with no postal code: ')
        print(df[((df['Address'].str.contains(c)) & (df['County'] == 'Dublin'))]['PostalCode'].isnull().sum())
        print('Number of postal codes with this address: ')
        print(df[((df['Address'].str.contains(c)) & (df['County'] == 'Dublin'))]['PostalCode'].value_counts())
        print('*****')

*****
```

ABBEY ST
Number of addresses in this area with no postal code:

36

Number of postal codes with this address:

PostalCode

DUBLIN 1 40

DUBLIN 13 1

CO. DUBLIN 1

CO DUBLIN 1

Name: count, dtype: int64

AMIENS ST

Number of addresses in this area with no postal code:

6

Number of postal codes with this address:

PostalCode

DUBLIN 1 28

Name: count, dtype: int64

CAPEL ST

Number of addresses in this area with no postal code:

24

Number of postal codes with this address:

```
*****
```

BALLYMOUNT

Number of addresses in this area with no postal code:

27

Number of postal codes with this address:

PostalCode

DUBLIN 24 31

DUBLIN 12 18

Name: count, dtype: int64

```
*****
```

- Here we can see that some areas can have multiple postal codes. Some of these contain the area name in the first line of the address and may be in a different area. I will investigate further and only assign postal codes to addresses in areas that have one definitive code.

In [88]:

```
areas = {'DUBLIN 1' : ['ABBEY ST', 'AMIENS ST', 'MOUNTJOY SQ',
                      'MARLBOROUGH ST', 'PARNELL SQ', 'TALBOT ST'],
         'DUBLIN 2' : ['DAME ST', 'MERRION SQ', 'GRAFTON ST', 'STEPHEN\S GR', 'STEPHENS GR',
                      'LOWER LEESON ST', 'CAMDEN ST'],
         'DUBLIN 3' : ['BALLYBOUGH', 'DOLLYMOUNT', 'CLONTARF', 'EAST WALL', 'FAIRVIEW',
                      'CLONLIFFE'], #clonliffe
         'DUBLIN 4' : ['UPPER LEESON ST', 'BALLSBRIDGE', 'DONNYBROOK', 'IRISHTOWN', 'BELFIELD', 'MERRION', 'PEMI',
                      'RINGSEND', 'SANDYMOUNT'], #ringsend
         'DUBLIN 5' : ['ARTANE', 'HARMONSTOWN', 'RAHENY'], #raheny
         'DUBLIN 6' : ['RANELAGH', 'RATHMINES', 'DARTRY', 'RATHGAR'],
         'DUBLIN 6W' : ['HAROLD\S CROSS', 'HAROLDS CROSS'], #HAROLDS CROSS
         'DUBLIN 7' : ['ARBOUR HILL', 'BROADSTONE', 'CABRA', 'GRANGEORGORMAN', 'OXMANTOWN', 'PHIBSBOROUGH',
                      'SMITHFIELD', 'STONEYBATTER'],
         'DUBLIN 8' : ['DOLPHIN\S BARN', 'DOLPHINS BARN', 'INCHICORE', 'ISLANDBRIDGE', 'KILMAINHAM',
                      'PORTOBELLO', 'SOUTH CIRCULAR ROAD', 'SOUTH CIRCULAR RD', 'LIBERTIES'],
         'DUBLIN 9' : ['BEAUMOUNT', 'DRUMCONDRA', 'ELM MOUNT', 'GRIFFITH AVE',
                      'SANTRY', 'WHITEHALL'],
         'DUBLIN 10' : ['BALLYFERMOT', 'SARSFIELD ROAD', 'SARSFIELD RD', 'CHERRY ORCHARD'],
         'DUBLIN 11' : ['SILLOGUE', 'BALBUTCHER', 'SANDYHILL', 'DUBBER CROSS',
                      'FINGLAS', 'CAPPAGH', 'CREMORE', 'ADDISON', 'VIOLET HILL', 'FINGLAS ROAD', 'FINGLAS RD',
                      'COOLQUAY'],
         'DUBLIN 12' : ['BLUEBELL', 'CRUMLIN', 'DRIMNAGH', 'GREENHILLS', 'PERRYSTOWN',
                      'WALKINSTOWN'],
         'DUBLIN 13' : ['BALDOYLE', 'BAYSIDE', 'DONAGHMEDE', 'CLONGRIGGIN', 'SUTTON', 'HOWTH', 'AYRFIELD'],
         'DUBLIN 14' : ['CHURCHTOWN', 'CLONSKEAGH', 'GOATSTOWN', 'WINDY ARBOUR'],
         'DUBLIN 15' : ['ASHTOWN', 'BLANCHARDSTOWN', 'CASTLEKNOCK', 'COOLMINE', 'CLONSILLA', 'CORDUFF', 'MULHUDI',
                      'TYRRELSTOWN', 'ONGAR', 'CLONEE'],
         'DUBLIN 16' : ['BALLINTEER', 'BALLYBODEN', 'KNOCKLYON'],
         'DUBLIN 17' : ['BELCAMP', 'DARNDALE', 'PRIORSWOOD'],
         'DUBLIN 18' : ['CABINTEELY', 'CARRICKMINES', 'FOXROCK', 'KILTERNAN', 'SHANKILL', 'TICKNOCK',
                      'BALLYEDMONDUFF', 'STEPASIDE', 'LEOPARDSTOWN'],
         'DUBLIN 20' : ['CHAPELIZOD', 'PALMERSTOWN'],
         'DUBLIN 22' : ['CLONDALKIN', 'NEILSTOWN'],
         'DUBLIN 24' : ['FIRHOUSE', 'JOBSTOWN', 'OLD BAWN', 'TALLAGHT', 'BALLYMOUNT']}
```

In [89]:

```
key_list = list(areas.keys())
val_list = list(areas.values())

for row in range(0, len(df)):
    if (df['County'][row] == 'Dublin') & (pd.isnull(df['PostalCode'][row])):
        address = df['Address'][row].split(',')
    #       print(address)

        for count, values in enumerate(val_list):
            for count2, place in enumerate(values):
                if (address[-1].strip(' ') == place):
                    position = val_list.index(values)
                    df.at[row, 'PostalCode'] = key_list[position]

                elif (address[-2].strip(' ') == place):
                    position = val_list.index(values)
                    df.at[row, 'PostalCode'] = key_list[position]

df[df['County'] == 'Dublin']['PostalCode'].isnull().sum()
```

Out[89]: 71914

In [90]:

```
df['PostalCode'].unique()
```

```
Out[90]: array(['CO. DUBLIN', nan, 'DUBLIN 14', 'DUBLIN 6', 'DUBLIN 3',
   'DUBLIN 15', 'DUBLIN 4', 'COUNTY DUBLIN', 'DUBLIN 11', 'DUBLIN 24',
   'DUBLIN 13', 'DUBLIN 18', 'DUBLIN 12', 'DUBLIN 16', 'DUBLIN 22',
   'DUBLIN 9', 'DUBLIN 1', 'DUBLIN 7', 'DUBLIN 17', 'DUBLIN 10',
   'DUBLIN 6 W', 'CO DUBLIN', 'DUBLIN 5', 'DUBLIN 8', 'DUBLIN 6W',
   'DUBLIN 20', 'D 15', 'D 16', 'DUBLIN 2', 'D 24', 'D 7',
   'D 18', 'D 13', 'D 4', 'D 8', 'D 9', 'D 6 W', 'D 5', 'D 2',
   'D 14', 'D 12'], dtype=object)
```

```
In [91]: df = df.replace({'PostalCode': {'D 1' : 'DUBLIN 1',
                                         'D 2' : 'DUBLIN 2',
                                         'D 3' : 'DUBLIN 3',
                                         'D 4' : 'DUBLIN 4',
                                         'D 5' : 'DUBLIN 5',
                                         'D 6' : 'DUBLIN 6',
                                         'D 6W' : 'DUBLIN 6W',
                                         'D 6 W' : 'DUBLIN 6W',
                                         'D 7' : 'DUBLIN 7',
                                         'D 8' : 'DUBLIN 8',
                                         'D 9' : 'DUBLIN 9',
                                         'D 10' : 'DUBLIN 10',
                                         'D 11' : 'DUBLIN 11',
                                         'D 12' : 'DUBLIN 12',
                                         'D 13' : 'DUBLIN 13',
                                         'D 14' : 'DUBLIN 14',
                                         'D 15' : 'DUBLIN 15',
                                         'D 16' : 'DUBLIN 16',
                                         'D 17' : 'DUBLIN 17',
                                         'D 18' : 'DUBLIN 18',
                                         'D 20' : 'DUBLIN 20',
                                         'D 22' : 'DUBLIN 22',
                                         'D 24' : 'DUBLIN 24',
                                         'CO DUBLIN' : 'COUNTY DUBLIN',
                                         'CO. DUBLIN' : 'COUNTY DUBLIN'
                                         }})
df[df['County'] == 'Dublin']['PostalCode'].isnull().sum()
# df['PostalCode'].unique()
```

```
Out[91]: 71914
```

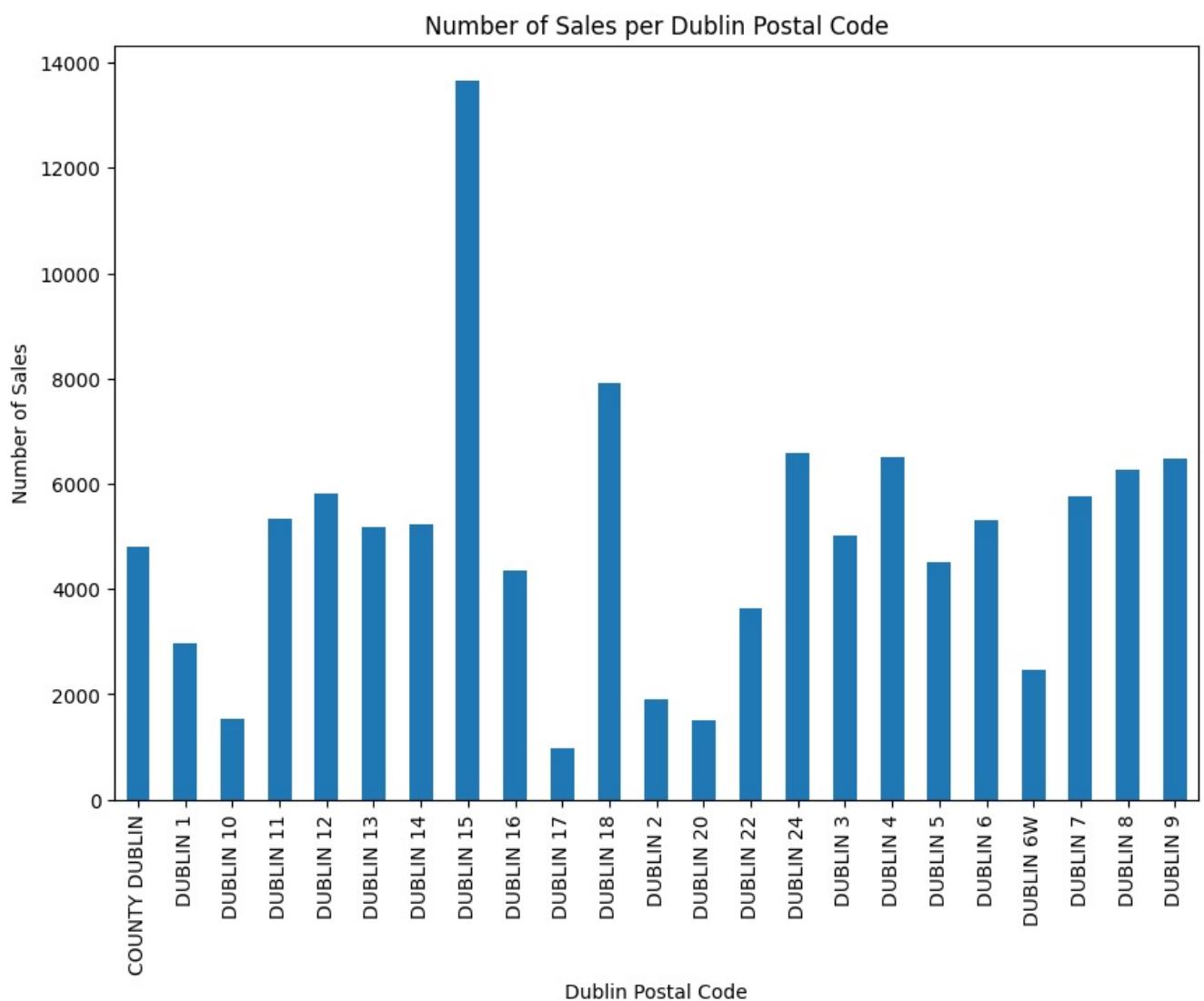
```
In [92]: df['PostalCode'].unique()
```

```
Out[92]: array(['COUNTY DUBLIN', nan, 'DUBLIN 14', 'DUBLIN 6', 'DUBLIN 3',
   'DUBLIN 15', 'DUBLIN 4', 'DUBLIN 11', 'DUBLIN 24', 'DUBLIN 13',
   'DUBLIN 18', 'DUBLIN 12', 'DUBLIN 16', 'DUBLIN 22', 'DUBLIN 9',
   'DUBLIN 1', 'DUBLIN 7', 'DUBLIN 17', 'DUBLIN 10', 'DUBLIN 6W',
   'DUBLIN 5', 'DUBLIN 8', 'DUBLIN 20', 'DUBLIN 2'], dtype=object)
```

- Number of Sales per Dublin Postal Code

```
In [96]: # Plot bar chart for Number of Sales in each Dublin area
plt.figure(figsize=(10,7))
df["Price"].groupby(df['PostalCode']).count().plot(kind='bar')
plt.title("Number of Sales per Dublin Postal Code")
plt.ylabel("Number of Sales")
plt.xlabel("Dublin Postal Code")
```

```
Out[96]: Text(0.5, 0, 'Dublin Postal Code')
```



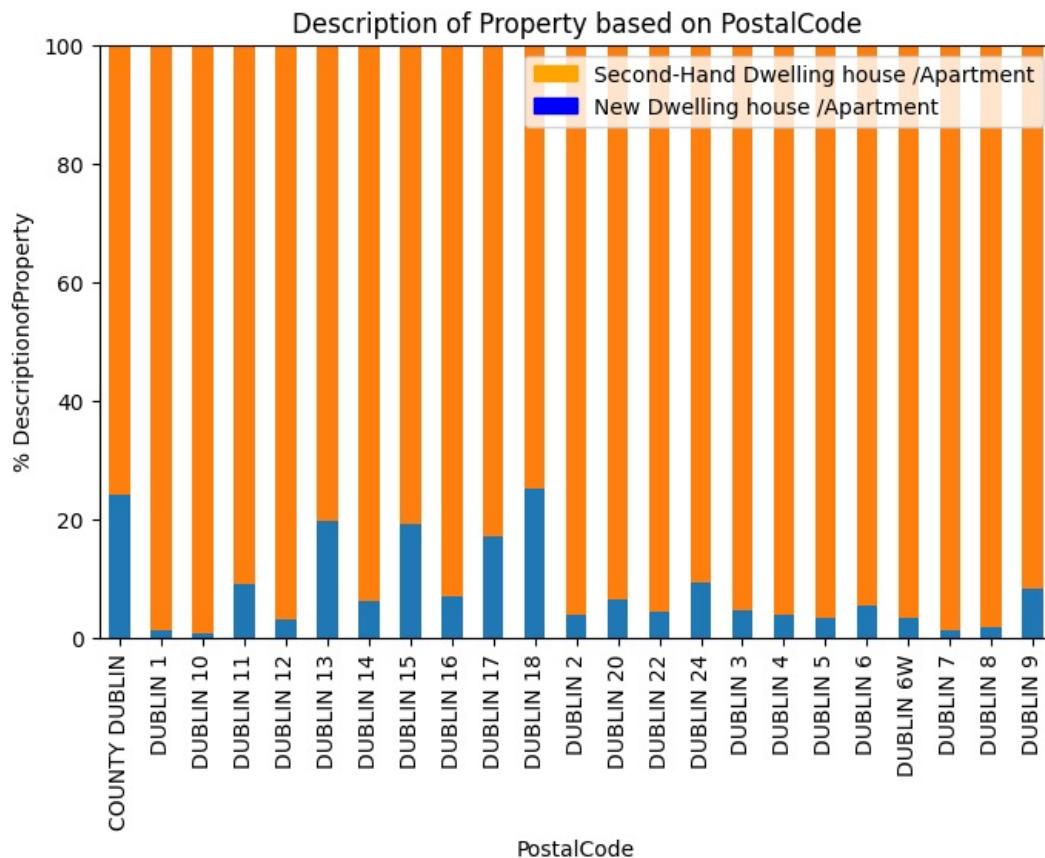
- Description of Property per Dublin Postal Code

```
In [93]: # check new vs old homes sold in Dublin based on postal code
stacked_bar_chart("PostalCode")
```

```
COUNTY DUBLIN      4803
nan          0
DUBLIN 14      5218
```

```
/var/folders/bn/6x11zkt6n14vfwyb0jymgt00000gp/T/ipykernel_31017/406172032.py:12: RuntimeWarning: divide by zero
encountered in long_scalars
  count_percentage = (1 / count) * 100
```

DUBLIN 6	5302
DUBLIN 3	5030
DUBLIN 15	13646
DUBLIN 4	6496
DUBLIN 11	5340
DUBLIN 24	6576
DUBLIN 13	5169
DUBLIN 18	7924
DUBLIN 12	5822
DUBLIN 16	4363
DUBLIN 22	3623
DUBLIN 9	6473
DUBLIN 1	2971
DUBLIN 7	5765
DUBLIN 17	988
DUBLIN 10	1530
DUBLIN 6W	2457
DUBLIN 5	4523
DUBLIN 8	6260
DUBLIN 20	1517
DUBLIN 2	1897

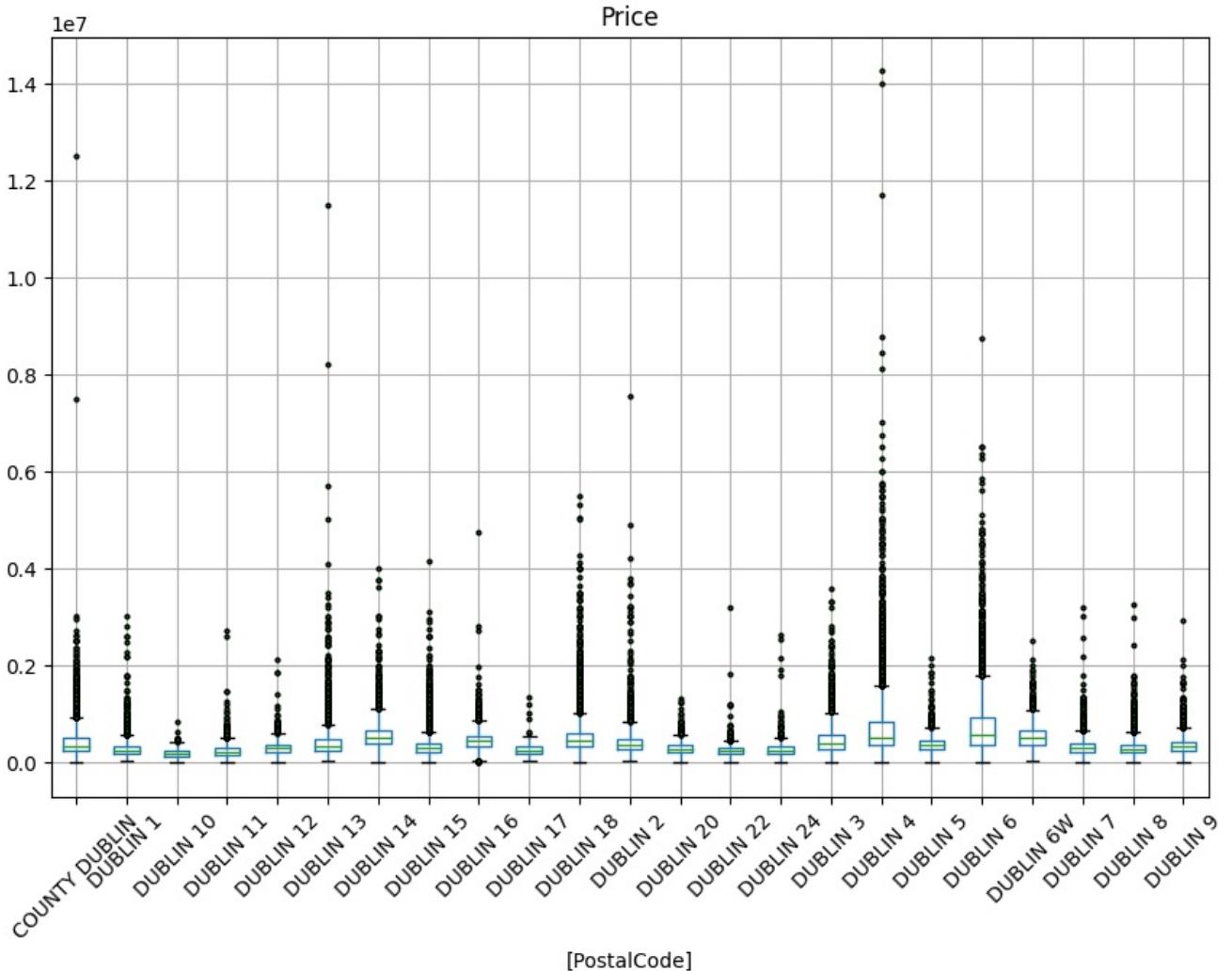


- Dublin 15 appears to be a newly desirable area to live as it has the highest number of properties sold and the fourth highest rate of new properties sold in Dublin.
- Here we can see Dublin 18 and County Dublin have the largest proportions of new houses sold compared to second hand houses. These areas are further out from the city centre than most other postal code areas which may indicate it is easier to build new properties further outside the city center.

```
In [97]: # code for box plot grouped by postal code
plt.figure()
flierprops = dict(marker='o', markerfacecolor='green', markersize=2,
                  linestyle='none')
df.boxplot(column=['Price'], by=['PostalCode'], rot=45, flierprops=flierprops, figsize=(10,7))
```

```
Out[97]: <Axes: title={'center': 'Price'}, xlabel='[PostalCode]'>
<Figure size 640x480 with 0 Axes>
```

Boxplot grouped by PostalCode

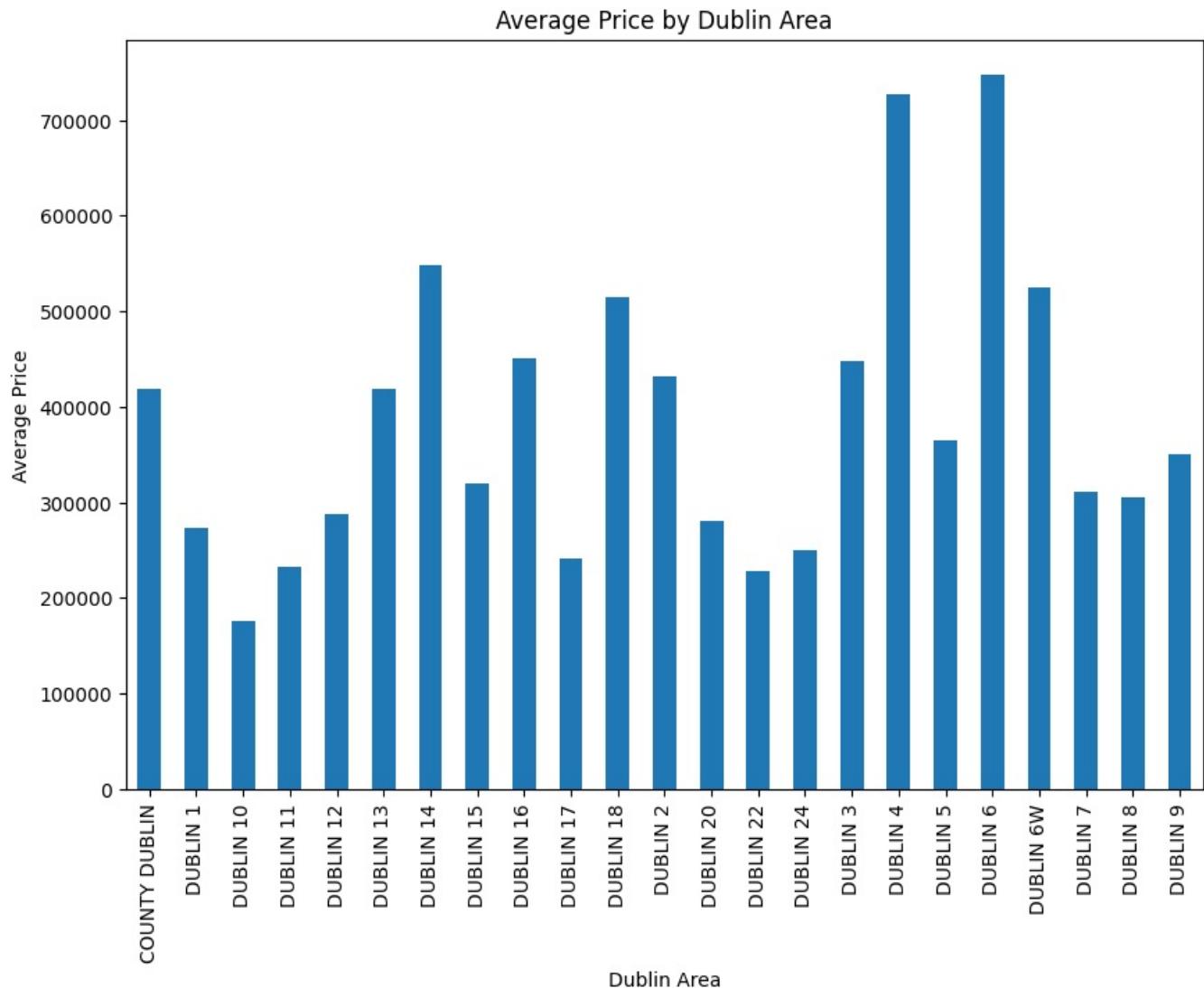


- Dublin 4 appears to have the largest range of prices for Dublin properties, showing a high variability in price. High outliers seem to suggest a high demand with buyers more than willing to pay a high price. I will look at the average price per Dublin area to get a better understanding of Dublin property prices.

- Average Price per Dublin Postal Code

```
In [98]: # bar chart for Average Price per Dublin Postal Code
plt.figure(figsize=(10,7))
df["Price"].groupby(df['PostalCode']).mean().plot(kind='bar')
plt.title("Average Price by Dublin Area")
plt.ylabel("Average Price")
plt.xlabel("Dublin Area")
```

```
Out[98]: Text(0.5, 0, 'Dublin Area')
```



- Here we can see Dublin 6 properties have the highest average Price, followed by Dublin 4, Dublin 14 and Dublin 6W. Interestingly, these are all areas south of the River Liffey. The highest prices in the northside are located in Dublin 3, which has the 7th highest average price per area code.

Feature 2

- The second feature I want to create is one which examines the month each address was sold. It will be interesting to see if sales peak at a certain time of year.

```
In [100]: #create a new column listing the month of each sale
df['MonthofSale'] = pd.DatetimeIndex(df['DateofSale']).month
```

```
In [101]: #check the column has been added by checking first 5 rows
df.head(5)
```

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescript
0	2010-01-01	5 BRAEMOR DRIVE, CHURCHTOWN, CO.DUBLIN	Dublin	343000.0		No	No	Second-Hand Dwelling house /Apartment
1	2010-01-03	134 ASHEWOOD WALK, SUMMERHILL LANE, PORTLAOISE	Laois	185000.0		No	Yes	New Dwelling house /Apartment
2	2010-01-04	1 MEADOW AVENUE, DUNDRUM, DUBLIN 14	Dublin	438500.0		No	No	Second-Hand Dwelling house /Apartment
3	2010-01-04	1 THE HAVEN, MORNINGTON	Meath	400000.0		No	No	Second-Hand Dwelling house /Apartment
4	2010-01-04	11 MELVILLE HEIGHTS, KILKENNY	Kilkenny	160000.0		No	No	Second-Hand Dwelling house /Apartment

In [102]: # plot the number of sales per month to look for a trend

```
plt.figure(figsize=(10,7))
df['MonthofSale'].value_counts().sort_index(ascending=True).plot(kind='bar')
plt.ylabel('Number of Sales')
plt.xlabel('Month by Number')
plt.title('Property Sales by month')
```

Out[102]: Text(0.5, 1.0, 'Property Sales by month')



- It appears the amount of properties sold per month trends upwards throughout the year with total properties sold peaking in December.

Feature 3

- The third feature I want to create is one which examines the year each address was sold. I would like to see whether house prices have changed from year to year.

In [103]: # create a new column with the year of sale

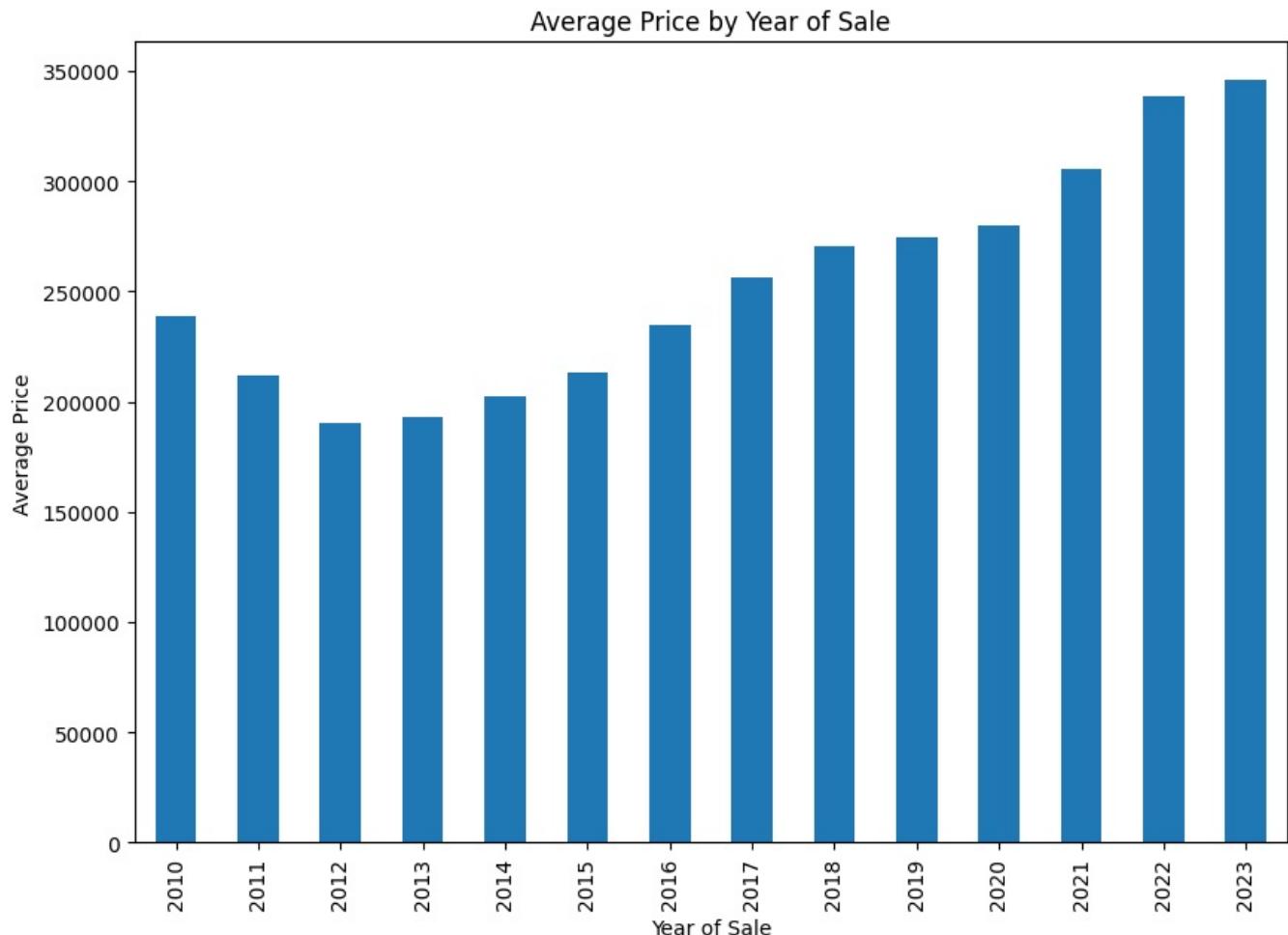
```
df['YearofSale'] = pd.DatetimeIndex(df['DateofSale']).year
```

In [104]: # check the column has been added
df.head(5)

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescripti
0	2010-01-01	5 BRAEMOR DRIVE, CHURCHTOWN, CO.DUBLIN	Dublin	343000.0		No	No	Second-Hand Dwelling house /Apartment
1	2010-01-03	134 ASHEWOOD WALK, SUMMERHILL LANE, PORTLAOISE	Laois	185000.0		No	Yes	New Dwelling house /Apartment greater than or equal 38 sq metres and less
2	2010-01-04	1 MEADOW AVENUE, DUNDRUM, DUBLIN 14	Dublin	438500.0		No	No	Second-Hand Dwelling house /Apartment
3	2010-01-04	1 THE HAVEN, MORNINGTON	Meath	400000.0		No	No	Second-Hand Dwelling house /Apartment
4	2010-01-04	11 MELVILLE HEIGHTS, KILKENNY	Kilkenny	160000.0		No	No	Second-Hand Dwelling house /Apartment

In [105]: # plot the mean price of properties sold per year to look for a trend
plt.figure(figsize=(10,7))
df["Price"].groupby(df['YearofSale']).mean().plot(kind='bar')
plt.title("Average Price by Year of Sale")
plt.ylabel("Average Price")
plt.xlabel("Year of Sale")

Out[105]: Text(0.5, 0, 'Year of Sale')

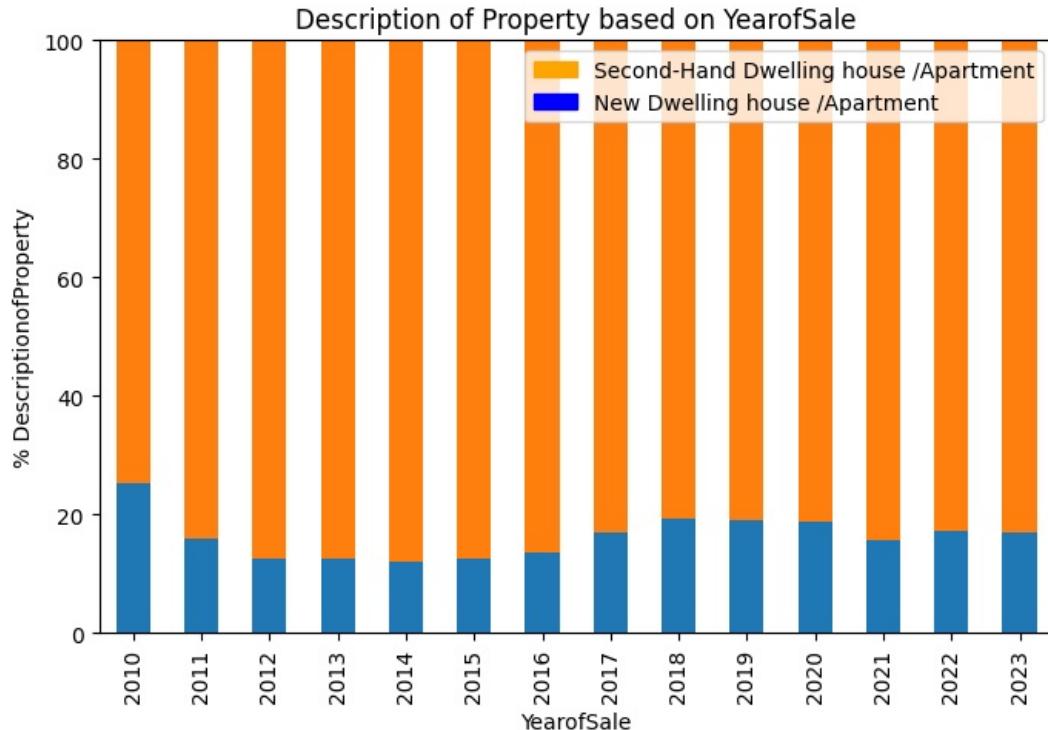


- We can see that the average property price in Ireland has been increasing since 2012. Next I would like to check the proportion of new houses to second-hand houses sold each year.

In [106]: #check proportion of new vs old homes sold in Dublin each year

```
stacked_bar_chart("YearofSale")
```

2010	20820
2011	18320
2012	25223
2013	29835
2014	43169
2015	48782
2016	49518
2017	54544
2018	56978
2019	58491
2020	48944
2021	58930
2022	61826
2023	16156



- There does not seem to be much correlation between the rise of property prices and new houses sold. My thinking was that rising property prices might be a result from a lack of new houses. This does not appear to be the case.

```
In [107]: # saving cleaned dataframe to a csv
df.to_csv("PPR-ALL_cleaned_3newFeatures.csv")
```

```
In [3]: # load cleaned data
df = pd.read_csv("PPR-ALL_cleaned_3newFeatures.csv", index_col=0)
df.shape
```

```
Out[3]: (591536, 11)
```

Feature 4

- The fourth feature I want to create is the geolocation of each address in Dublin. I think it is useful to only look at Dublin data since it is the county with the highest property prices and the county we have the most data on. We saw that South Dublin properties appear to have a higher average price and adding the geolocation of each address to the data allows for a more detailed view of how this price is affected by its location. For this I will add 2 columns for latitude and longitude.

```
In [4]: #import relevant packages
from geopy.geocoders import Nominatim
from geopy.extra.rate_limiter import RateLimiter
from pprint import pprint
import requests
import json

#Creating an instance of Nominatim Class
geolocator = Nominatim(user_agent='my_request')

#applying the rate limiter wrapper
geocode = RateLimiter(geolocator.geocode, min_delay_seconds=1)

#create two new columns for latitude and longitude
df['Lat'] = np.nan
```

```
df['Lon'] = np.nan
```

```
df.tail(5)
```

Out[4]:

		DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeE
591531	2023-05-05		MISTLETOE COTTAGE, GREENCLOYNE, YOUGHAL	Cork	300000.0		No	No	Second-Hand Dwelling house /Apartment
591532	2023-05-05		PARK RD, SWINFORD, MAYO	Mayo	190000.0		No	No	Second-Hand Dwelling house /Apartment
591533	2023-05-05		SPRING COTTAGE, SHALLON, THE WARD	Dublin	740000.0		No	No	Second-Hand Dwelling house /Apartment
591534	2023-05-05		THE ORCHARD, GARRYHILL, CARLOW	Carlow	375000.0		No	No	Second-Hand Dwelling house /Apartment
591535	2023-05-05		TULLYCROSS MOYDRUM, ATHLONE, WESTMEATH	Westmeath	381000.0		No	No	Second-Hand Dwelling house /Apartment

In [5]:

```
#iterate through each row, find geolocation of address if the county is Dublin
for row in range(0, len(df)):
    if df['County'][row] == 'Dublin':
        if pd.isnull(df.at[row, 'Lat']):
            addr = df.at[row, 'Address']
            try:
                location = geolocator.geocode(addr).raw
                df.at[row, 'Lat'] = location.get('lat')
                df.at[row, 'Lon'] = location.get('lon')
                print('Row No. ' + str(row) + ' success')
                df.to_csv("PPR-ALL_cleaned_section4.csv")
            except:
                df.at[row, 'Lat'] = 'other'
                df.at[row, 'Lon'] = 'other'
                print('Row No. ' + str(row) + ' failed')
        else:
            df.at[row, 'Lat'] = 'other'
            df.at[row, 'Lon'] = 'other'
```

```
df.tail(5)
#check rows with county as Dublin
# df[df['County'] == 'Dublin']
```

Row No. 0 success
Row No. 2 success
Row No. 5 failed
Row No. 11 success
Row No. 12 success
Row No. 13 success
Row No. 16 success
Row No. 17 failed
Row No. 19 success
Row No. 21 failed
Row No. 22 success
Row No. 23 success
Row No. 27 success
Row No. 42 failed
Row No. 47 failed
Row No. 50 failed
Row No. 56 failed
Row No. 58 success
Row No. 61 failed
Row No. 63 failed
Row No. 65 failed
Row No. 66 failed
Row No. 67 success
Row No. 76 success
Row No. 79 success
Row No. 84 success
Row No. 87 failed
Row No. 95 success
Row No. 97 failed
Row No. 115 success
Row No. 117 success

	DateofSale	Address	County	Price	NotFullMarketPrice	VATExclusive	DescriptionofProperty	PropertySizeDescription
591531	2023-05-05	MISTLETOE COTTAGE, GREENCLOYNE, YOUHAL	Cork	300000.0	No	No	Second-Hand Dwelling house /Apartment	
591532	2023-05-05	PARK RD, SWINFORD, MAYO	Mayo	190000.0	No	No	Second-Hand Dwelling house /Apartment	
591533	2023-05-05	SPRING COTTAGE, SHALLON, THE WARD	Dublin	740000.0	No	No	Second-Hand Dwelling house /Apartment	
591534	2023-05-05	THE ORCHARD, GARRYHILL, CARLOW	Carlow	375000.0	No	No	Second-Hand Dwelling house /Apartment	
591535	2023-05-05	TULLYCROSS MOYDRUM, ATHLONE, WESTMEATH	Westmeath	381000.0	No	No	Second-Hand Dwelling house /Apartment	

```
In [9]: #check number of rows with a geolocation
df[df['Lat'] != 'other'].count()
```

```
Out[9]: DateofSale      7176
Address        7176
County         7176
Price          7176
NotFullMarketPrice  7176
VATExclusive   7176
DescriptionofProperty 7176
PropertySizeDescription 7176
PostalCode      4801
MonthofSale     7176
YearofSale      7176
Lat            7176
Lon            7176
dtype: int64
```

```
In [10]: #saving cleaned dataframe to a csv
df.to_csv("PPR-ALL_cleaned_section4.csv")
```

- The geolocation of some addresses weren't found and so we still have missing coordinates for Dublin addresses. A quick google shows that this happens from time to time using geopy and so I will use the data I have rather than drop the new features. Since we are looking at Dublin addresses only, I will calculate new percentiles for this price data. Using this data, I will pin each address to a map and use colors to highlight each price category they fall under.

```
In [109...]: #calculate percentiles and median for Dublin property prices
dublin_data = df[df['County'] == 'Dublin']['Price']

p_25 = np.percentile(dublin_data, 25)
p_50 = np.median(dublin_data)
p_75 = np.percentile(dublin_data, 75)
print(p_25, p_50, p_75)
```

226000.0 320000.0 458000.0

```
In [ ]: import folium

#new list for index of rows with latitude and longitude data
geo = []

#new map
m = folium.Map(location = [53.3498, -6.2603], zoom_start = 10)

#add row indexes to list
for i in range(0,len(df)):
    if (df.loc[i, 'Lat'] != 'other') & (pd.notnull(df.loc[i, 'Lat'])):
#        if pd.notnull(df.loc[i, 'Lat']):
            geo += [i]

#iterate through rows in geo list, check the price range of each address and pin to the map
for coord in geo[:5000]:
    if df['Price'][coord] <= p_25:
        color = 'blue'
    elif p_25 < df['Price'][coord] <= p_50:
        color = 'green'
    elif p_50 < df['Price'][coord] <= p_75:
        color = 'orange'
```

```
    elif df['Price'][coord] > p_75:
        color = 'red'
    tooltip = df['Address'][coord]
    folium.Marker([df['Lat'][coord], df['Lon'][coord]], popup = f"<strong>coord</strong>", tooltip = tooltip, icon=icon)
#display map
m
```

- Here we have a better view of Dublin property prices. Most of the properties above or equal to the 75th percentile in price appear to be South Dublin. Most properties within this price range appear to be closer to the coast, regardless of whether the area is North or South of the River Liffey.

```
In [107]: #saving cleaned dataframe to a csv
df.to_csv("ppr-13391711_cleaned_section4.csv")
```

```
In [ ]:
```