price_on_zip.ipynb | Clean_House_Price_Index.csv | property_price.csv

Code

[46]:
```
import pandas as pd
import numpy as np
from pathlib import Path
```

[55]:
```
#import data and create data frame
price_data = Path("../data_collection/property/property_price.csv")
HP = pd.read_csv(price_data, index_col= "Five-Digit ZIP Code")
#data cleanup
HP['HPI'] = HP['HPI'].replace(['.'],'0')
HP['Annual Change (%)'] = HP['Annual Change (%)'].replace(['.'],'0')
HP['HPI with 1990 base'] = HP['HPI with 1990 base'].replace(['.'],'0')
HP['HPI with 2000 base'] = HP['HPI with 2000 base'].replace(['.'],'0')
HP = HP.astype({"Annual Change (%)": float, "HPI": float,"HPI with 1990 base": float,"HPI with 2000 base": float})
HP['HPI'] = HP['HPI'].div(50).round(0)
HP['HPI with 1990 base'] = HP['HPI with 1990 base'].div(50).round(0)
HP['HPI with 2000 base'] = HP['HPI with 2000 base'].div(50).round(0)
HP['Annual Change (%)'] = HP['Annual Change (%)'].round(0)
maxVal = 10
HP.loc[HP['HPI'] >= maxVal, 'HPI'] = maxVal
HP.loc[HP['HPI with 1990 base'] >= maxVal, 'HPI with 1990 base'] = maxVal
HP.loc[HP['HPI with 2000 base'] >= maxVal, 'HPI with 2000 base'] = maxVal
HP.head()
```

[55]:

Five-Digit ZIP Code	Year	Annual Change (%)	HPI	HPI with 1990 base	HPI with 2000 base
1001	1985	0.0	2.0	1.0	1.0
1001	1986	14.0	2.0	1.0	1.0
1001	1987	21.0	3.0	2.0	2.0
1001	1988	17.0	3.0	2.0	2.0
1001	1989	1.0	3.0	2.0	2.0

[56]:
```
#export cleansed dat to csv
HP.to_csv(r'..\data_collection\property\Clean_House_Price_Index.csv')
```