In [39]:
```
import pandas as pd
import sys
import sklearn.neighbors._base
sys.modules['sklearn.neighbors.base'] = sklearn.neighbors._base
import numpy as np
import re
import pandas as pd
from missingpy import MissForest

houseData = pd.read_csv(r'C:\Users\VAIO\Documents\houseData.csv')
#houseData
```

In [40]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2230 entries, 0 to 2229
Data columns (total 26 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   num_half_bathrooms              172 non-null    float64
 1   total_taxes                     584 non-null    object
 2   walk_score                      2230 non-null   int64
 3   listing_price_to_nearest_1000   1696 non-null   object
 4   sq_footage                      1020 non-null   float64
 5   sale_price                      528 non-null    object
 6   pct_tax_deductibl               476 non-null    float64
 7   parking_charges                 559 non-null    object
 8   num_total_rooms                 2228 non-null   float64
 9   num_full_bathrooms              2230 non-null   int64
 10  num_floors_in_building          1580 non-null   float64
 11  num_bedrooms                    2115 non-null   float64
 12  model_type                      2151 non-null   object
 13  maintenance_cost                1607 non-null   object
 14  kitchen_type                    2214 non-null   object
 15  garage_exists                   404 non-null    object
 16  full_address_or_zip_code        2230 non-null   object
 17  fuel_type                       2118 non-null   object
 18  dogs_allowed                    2230 non-null   object
 19  dining_room_type                1782 non-null   object
 20  date_of_sale                    528 non-null    object
 21  coop_condo                      2230 non-null   object
 22  community_district_num          2211 non-null   float64
 23  common_charges                  546 non-null    object
 24  cats_allowed                    2230 non-null   object
 25  approx_year_built               2190 non-null   float64
dtypes: float64(8), int64(2), object(16)
memory usage: 453.1+ KB
```

In [41]:
```
# Before combining costs to get a more comprehensive feature, lets record missingness
for x in houseData.index:
    if pd.isnull(houseData.loc[x, 'total_taxes'] ):
        houseData.loc[x, 'Missing_taxes'] = 1
    else:
        houseData.loc[x, 'Missing_taxes'] = 0

    if pd.isnull(houseData.loc[x, 'maintenance_cost'] ):
        houseData.loc[x, 'Missing_maintenance_cost'] = 1
    else:
        houseData.loc[x, 'Missing_maintenance_cost'] = 0

    if pd.isnull(houseData.loc[x, 'common_charges'] ):
        houseData.loc[x, 'Missing_common_charges'] = 1
```

```
    else:
        houseData.loc[x, 'Missing_common_charges'] = 0

    if pd.isnull(houseData.loc[x, 'parking_charges'] ):
        houseData.loc[x, 'Missing_parking_charges'] = 1
    else:
        houseData.loc[x, 'Missing_parking_charges'] = 0
```

In [42]:

```
#get rid of dollar signs and commas
#combine maintenence cost, park charge, common charge, tot taxes
houseData['sale_price'] = houseData['sale_price'].replace({'\$': '', ',': ''}, regex=True)
houseData['sale_price_$'] = houseData['sale_price']

houseData[['total_taxes',
           'maintenance_cost',
           'common_charges',
           'parking_charges' ,
           'listing_price_to_nearest_1000'
          ]
         ] = houseData[['total_taxes',
                        'maintenance_cost',
                        'common_charges',
                        'parking_charges' ,
                        'listing_price_to_nearest_1000'
                       ]
                      ].replace({'\$': '', ',': ''}, regex=True)

houseData[['parking_charges',
           'total_taxes',
           'maintenance_cost',
           'common_charges',
           'num_half_bathrooms',
           'garage_exists'
          ]
         ] = houseData[['parking_charges',
                        'total_taxes',
                        'maintenance_cost',
                        'common_charges',
                        'num_half_bathrooms',
                        'garage_exists'
                       ]
                      ].fillna(0)

for x in houseData.index:
    houseData.loc[x,"additional_costs_$"] = (int(houseData.loc[x,"parking_charges"]) +
                                             int(houseData.loc[x,"total_taxes"]) +
                                             int(houseData.loc[x,'maintenance_cost']) +
                                             int(houseData.loc[x,'common_charges'])
                                            )
```

In [43]:

```
#drop the features with the dollar signs that have now been removed and combined into one
houseData.drop(['total_taxes',
                'sale_price',
                "parking_charges",
                'maintenance_cost',
                'common_charges'
               ], inplace = True, axis = 1)
```

In [44]:

```
#Clean up the categorical features
for x in houseData.index:
    # kitchen type has a ton of typos
```

```
    if houseData.loc[x, "kitchen_type"] == "Eat In":
        houseData.loc[x, "kitchen_type"] = "eat in"
    if houseData.loc[x, "kitchen_type"] == "eatin":
        houseData.loc[x, "kitchen_type"] = "eat in"
    if houseData.loc[x, "kitchen_type"] == "Eat in":
        houseData.loc[x, "kitchen_type"] = "eat in"
    if houseData.loc[x, "kitchen_type"] == "Combo":
        houseData.loc[x, "kitchen_type"] = "combo"
    if houseData.loc[x, "kitchen_type"] == "1955":
        houseData.loc[x, "kitchen_type"] = "none"
    if houseData.loc[x, "kitchen_type"] == "efficiency kitchen":
        houseData.loc[x, "kitchen_type"] = "efficiency"
    if houseData.loc[x, "kitchen_type"] == "efficiency kitchene":
        houseData.loc[x, "kitchen_type"] = "efficiency"
    if houseData.loc[x, "kitchen_type"] == "efficiency ktchen":
        houseData.loc[x, "kitchen_type"] = "efficiency"
    if houseData.loc[x, "kitchen_type"] == "efficiemcy":
        houseData.loc[x, "kitchen_type"] = "efficiency"
    # fuel type needs 'Other' - 'other'
    if houseData.loc[x, "fuel_type"] == "Other":
        houseData.loc[x, "fuel_type"] = "other"
    #garage exists needs 'Yes' - 'yes'
    #'Underground' - 'yes'
    # 'UG' - 'yes'
    # '1' - 'yes'
    # 'eys' - 'yes'
    if houseData.loc[x, "garage_exists"] == "Yes":
        houseData.loc[x, "garage_exists"] = "1"
    if houseData.loc[x, "garage_exists"] == 'Underground':
        houseData.loc[x, "garage_exists"] = "1"
    if houseData.loc[x, "garage_exists"] == 'UG':
        houseData.loc[x, "garage_exists"] = "1"
    if houseData.loc[x, "garage_exists"] == 'eys':
        houseData.loc[x, "garage_exists"] = "1"
    if houseData.loc[x, "garage_exists"] == '1':
        houseData.loc[x, "garage_exists"] = "1"
    if houseData.loc[x, "garage_exists"] == 'yes':
        houseData.loc[x, "garage_exists"] = "1"
    #dogs needs 'yes89' -'yes
    if houseData.loc[x, "dogs_allowed"] == "yes89":
        houseData.loc[x, "dogs_allowed"] = "1"
    if houseData.loc[x, "dogs_allowed"] == "yes":
        houseData.loc[x, "dogs_allowed"] = "1"
    if houseData.loc[x, "dogs_allowed"] == "no":
        houseData.loc[x, "dogs_allowed"] = "0"
    #cats needs 'y' - 'yes'
    if houseData.loc[x, "cats_allowed"] == "y":
        houseData.loc[x, "cats_allowed"] = "1"
    if houseData.loc[x, "cats_allowed"] == "yes":
        houseData.loc[x, "cats_allowed"] = "1"
    if houseData.loc[x, "cats_allowed"] == "no":
        houseData.loc[x, "cats_allowed"] = "0"
```

In [45]:

```
# can't figure out a great way to use full address
houseData['zip_code'] = houseData['full_address_or_zip_code'].str.extract(r'(\d{5}\-?\d{0,
```

In [46]:

```
import sklearn.neighbors._base
from sklearn.preprocessing import LabelEncoder
import sys
sys.modules['sklearn.neighbors.base'] = sklearn.neighbors._base
from missingpy import MissForest

#how to impute categorical features with missingpy's MissForest https://stackoverflow.com/
```

```python
def label_encoding(df, columns):
    encoders = dict()
    for col_name in columns:
        series = df[col_name]
        label_encoder = LabelEncoder()
        df[col_name] = pd.Series(
            label_encoder.fit_transform(series[series.notnull()]),
            index=series[series.notnull()].index
        )
        encoders[col_name] = label_encoder
    return encoders

# adding to be imputed global category along with features
features = ['num_half_bathrooms',
            'num_total_rooms',
            'num_full_bathrooms',
            'num_floors_in_building',
            'num_bedrooms',
            'kitchen_type',
            'full_address_or_zip_code',
            'fuel_type',
            'dining_room_type',
            'date_of_sale',
            'coop_condo',
            'community_district_num'
           ]
# label encoding features
encoders = label_encoding(houseData, features)
# categorical imputation using random forest
# parameters can be tuned accordingly
imp_cat = MissForest(criterion = "gini")
houseData[features] = imp_cat.fit_transform(houseData[features], cat_vars=[0,1,2,3,4,5,6,7
```

```
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
```

```
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
Iteration: 0
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
Iteration: 1
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
```

```
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
Iteration: 2

C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
```

```
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
Iteration: 3
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
Iteration: 4
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
```

```
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also
the default value for RandomForestClassifiers and ExtraTreesClassifiers.
  warn(
Iteration: 5
```

In [56]:

```
pd.set_option('max_columns', None)
pd.set_option("max_rows", None)
houseData.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2230 entries, 0 to 2229
Data columns (total 28 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   num_half_bathrooms              2230 non-null   float64
 1   walk_score                      2230 non-null   int64  
 2   listing_price_to_nearest_1000   1696 non-null   object 
 3   sq_footage                      1020 non-null   float64
 4   pct_tax_deductibl               476 non-null    float64
 5   num_total_rooms                 2230 non-null   float64
 6   num_full_bathrooms              2230 non-null   float64
 7   num_floors_in_building          2230 non-null   float64
 8   num_bedrooms                    2230 non-null   float64
 9   model_type                      2151 non-null   object 
 10  kitchen_type                    2230 non-null   float64
 11  garage_exists                   2230 non-null   object 
 12  full_address_or_zip_code        2230 non-null   float64
 13  fuel_type                       2230 non-null   float64
 14  dogs_allowed                    2230 non-null   object 
 15  dining_room_type                2230 non-null   float64
 16  date_of_sale                    2230 non-null   float64
 17  coop_condo                      2230 non-null   float64
 18  community_district_num          2230 non-null   float64
 19  cats_allowed                    2230 non-null   object 
 20  approx_year_built               2190 non-null   float64
 21  Missing_taxes                   2230 non-null   float64
 22  Missing_maintenance_cost        2230 non-null   float64
 23  Missing_common_charges          2230 non-null   float64
 24  Missing_parking_charges         2230 non-null   float64
 25  sale_price_$                    528 non-null    object 
 26  additional_costs_$              2230 non-null   float64
 27  zip_code                        2215 non-null   object 
dtypes: float64(20), int64(1), object(7)
memory usage: 487.9+ KB
```

In [48]:

```
# Make an instance and perform the imputation
imp_cont = MissForest()
```

```
X = houseData.drop(['sale_price_$', 'model_type'], axis = 1)
#X.info()
```

In [49]:

```
#cat_vars=[0,5,6,7,8,10,11,13,14,15,16]
#[['listing_price_to_nearest_1000', 'sq_footage', 'pct_tax_deductibl', 'zip_code']]
X_imputed = imp_cont.fit_transform(X)
X_imputed
```

```
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
Iteration: 0
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
```

```
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
Iteration: 1
```

```
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
```

```
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
Iteration: 2
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:400: FutureWarning:
Criterion 'mse' was deprecated in v1.0 and will be removed in version 1.2. Use `criterion
='squared_error'` which is equivalent.
  warn(
C:\Users\VAIO\anaconda3\lib\site-packages\sklearn\ensemble\_forest.py:416: FutureWarning:
`max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the p
ast behaviour, explicitly set `max_features=1.0` or remove this parameter as it is also th
e default value for RandomForestRegressors and ExtraTreesRegressors.
  warn(
Iteration: 3
```

```
Out[49]: array([[0.0000e+00, 8.2000e+01, 2.7435e+02, ..., 1.0000e+00, 7.6700e+02,
```

```
        1.1355e+04],
       [0.0000e+00, 8.9000e+01, 2.6641e+02, ..., 1.0000e+00, 6.0400e+02,
        1.1354e+04],
       [0.0000e+00, 9.0000e+01, 4.6933e+02, ..., 1.0000e+00, 5.6670e+03,
        1.1368e+04],
       ...,
       [0.0000e+00, 9.6000e+01, 8.5000e+02, ..., 1.0000e+00, 5.0000e+02,
        1.1385e+04],
       [0.0000e+00, 9.6000e+01, 8.5000e+02, ..., 1.0000e+00, 5.0000e+02,
        1.1385e+04],
       [0.0000e+00, 8.2000e+01, 8.9900e+02, ..., 1.0000e+00, 4.5770e+03,
        1.1360e+04]])
```

In [50]:

```
X_imputed = pd.DataFrame(X_imputed, columns = X.columns)
X_imputed = pd.concat([X_imputed,houseData['sale_price_$']], axis = 1)
# decoding features
#for variable in features:
#    X_imputed[variable] = encoders[variable].inverse_transform(X_imputed[variable].astype
#X_imputed
```

In [51]:

```
X_imputed.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2230 entries, 0 to 2229
Data columns (total 27 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   num_half_bathrooms              2230 non-null   float64
 1   walk_score                      2230 non-null   float64
 2   listing_price_to_nearest_1000   2230 non-null   float64
 3   sq_footage                      2230 non-null   float64
 4   pct_tax_deductibl               2230 non-null   float64
 5   num_total_rooms                 2230 non-null   float64
 6   num_full_bathrooms              2230 non-null   float64
 7   num_floors_in_building          2230 non-null   float64
 8   num_bedrooms                    2230 non-null   float64
 9   kitchen_type                    2230 non-null   float64
 10  garage_exists                   2230 non-null   float64
 11  full_address_or_zip_code        2230 non-null   float64
 12  fuel_type                       2230 non-null   float64
 13  dogs_allowed                    2230 non-null   float64
 14  dining_room_type                2230 non-null   float64
 15  date_of_sale                    2230 non-null   float64
 16  coop_condo                      2230 non-null   float64
 17  community_district_num          2230 non-null   float64
 18  cats_allowed                    2230 non-null   float64
 19  approx_year_built               2230 non-null   float64
 20  Missing_taxes                   2230 non-null   float64
 21  Missing_maintenance_cost        2230 non-null   float64
 22  Missing_common_charges          2230 non-null   float64
 23  Missing_parking_charges         2230 non-null   float64
 24  additional_costs_$              2230 non-null   float64
 25  zip_code                        2230 non-null   float64
 26  sale_price_$                    528 non-null    object 
dtypes: float64(26), object(1)
memory usage: 470.5+ KB
```

In [52]:

```
# We can't impute sale price so we're done imputing. We must drop any missing sale_price 
# to attempt to train models on this data

houseData5 = X_imputed
houseData5 = houseData5.dropna(axis = 0)
```

In [ ]:
```
houseData5
```

In [55]:
```
houseData5.to_csv(r'C:\Users\VAIO\Documents\houseData5.csv', index = False)
```

In [ ]:
```

```