In [ ]: ```python
import pandas as pd
import numpy as np
```

In [ ]: ```python
df = pd.read_csv('kc_house_data.csv')
df.sample(10)
```

Out[ ]:

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors
4082	3751601501	20140716T000000	382450.0	3	2.50	2220	20531	2.0
605	240000058	20150408T000000	469000.0	4	2.75	3550	13938	1.0
4993	9274203190	20140611T000000	650000.0	2	1.00	1030	5750	1.0
10864	7140600055	20150411T000000	207000.0	3	1.00	990	10800	1.0
15028	7889100020	20150414T000000	270000.0	3	2.50	1720	8550	1.0
4134	7129302615	20150304T000000	292000.0	3	1.75	1090	4500	1.5
17280	2473380920	20140813T000000	206325.0	5	2.50	1720	10202	1.5
21499	6891100090	20141014T000000	850000.0	5	3.50	4200	5400	2.0
17425	480000170	20140627T000000	480000.0	2	1.00	1500	3420	1.0
16288	1549500585	20150427T000000	585000.0	3	2.00	2220	209523	1.0

10 rows × 21 columns

In [ ]: ```python
df.isna().sum()
```

Out[ ]:
```
id   0
date   0
price   0
bedrooms   0
bathrooms   0
sqft_living   0
sqft_lot   0
floors   0
waterfront   0
view   0
condition   0
grade   0
sqft_above   0
sqft_basement   0
yr_built   0
yr_renovated   0
zipcode   0
lat   0
long   0
sqft_living15   0
sqft_lot15   0
dtype: int64
```

In [ ]: ```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21613 entries, 0 to 21612
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   id             21613 non-null  int64
 1   date           21613 non-null  object
 2   price          21613 non-null  float64
 3   bedrooms       21613 non-null  int64
 4   bathrooms      21613 non-null  float64
 5   sqft_living    21613 non-null  int64
 6   sqft_lot       21613 non-null  int64
 7   floors         21613 non-null  float64
 8   waterfront     21613 non-null  int64
 9   view           21613 non-null  int64
 10  condition      21613 non-null  int64
 11  grade          21613 non-null  int64
 12  sqft_above     21613 non-null  int64
 13  sqft_basement  21613 non-null  int64
 14  yr_built       21613 non-null  int64
 15  yr_renovated   21613 non-null  int64
 16  zipcode        21613 non-null  int64
 17  lat            21613 non-null  float64
 18  long           21613 non-null  float64
 19  sqft_living15  21613 non-null  int64
 20  sqft_lot15     21613 non-null  int64
dtypes: float64(5), int64(15), object(1)
memory usage: 3.5+ MB
```

```python
In [ ]:  for column in df.columns:
             print(column,'\t\t', df[column].astype(str).str.len().max(), )
```

```
id               10
date             15
price            9
bedrooms                 2
bathrooms                4
sqft_living              5
sqft_lot                 7
floors           3
waterfront               1
view             1
condition                1
grade            2
sqft_above               4
sqft_basement            4
yr_built                 4
yr_renovated             4
zipcode                  5
lat              7
long             8
sqft_living15            4
sqft_lot15               6
```