Домашнее задание

Пример идеального решения:

Задание 1:

```
import pandas as pd

df = pd.read_csv('kc_house_data.csv', sep=',')
df.head()
```

sqft_living	sqft_lot	floors	waterfront	view	...	grade	sqft_above	sqft_basement	yr_built	yr_renovated
1180	5650	1.0	0	0	...	7	1180	0	1955	0
2570	7242	2.0	0	0	...	7	2170	400	1951	1991
770	10000	1.0	0	0	...	6	770	0	1933	0
1960	5000	1.0	0	0	...	7	1050	910	1965	0
1680	8080	1.0	0	0	...	8	1680	0	1987	0

Задание 2:

Все столбцы датасета состоят из числовых типов данных (int64, float64), за исключением столбца дата (тип данных object).

Таблица не содержит пустых ячеек.

```
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
```

```
df.describe()
```

	id	price	bedrooms	bathrooms
count	2.161300e+04	2.161300e+04	21613.000000	21613.000000
mean	4.580302e+09	5.400881e+05	3.370842	2.114757
std	2.876566e+09	3.671272e+05	0.930062	0.770163
min	1.000102e+06	7.500000e+04	0.000000	0.000000
25%	2.123049e+09	3.219500e+05	3.000000	1.750000
50%	3.904930e+09	4.500000e+05	3.000000	2.250000
75%	7.308900e+09	6.450000e+05	4.000000	2.500000
max	9.900000e+09	7.700000e+06	33.000000	8.000000

Задание 3:

3.1

```
df['price'].min(), df['price'].max()
```

```
(75000.0, 7700000.0)
```

3.2

```
df['sqft_living'].mean() * 100 / df['sqft_lot'].mean()
```

```
13.76781757959227
```

3.3

```
df['floors'].value_counts()
```

```
1.0    10680
2.0     8241
1.5     1910
3.0      613
2.5      161
3.5        8
Name: floors, dtype: int64
```

3.4

```
df['condition'].value_counts()
```

```
3    14031
4     5679
5     1701
2      172
1       30
Name: condition, dtype: int64
```

3.5

```
df['yr_built'].min(), df['yr_built'].max()
```

```
(1900, 2015)
```

Задание 4:

4.1

```
df[df['bedrooms'] == 2]['price'].mean()
```

```
401372.681884058
```

4.2

```
df[df['price'] > 600000]['sqft_lot'].mean()
```

```
20442.524776214832
```

4.3

```
df[(df['yr_renovated'] != 0)].shape[0]
```

```
914
```

4.4

```
price_10 = df[df['grade'] > 10]['price'].mean()
price_10
```

```
1678635.1175298805
```

```
price_4 = df[df['grade'] < 4]['price'].mean()
price_4
```

```
189750.0
```

```
price_10 - price_4
```

```
1488885.1175298805
```

Задание 5:

5.1

```
df[(df['waterfront'] == 1) & (df['bathrooms'] > 3) & (df['sqft_basement'] != 0)].shape[0]
```

```
36
```

5.2

```
condition = ((df['view'] == 4) | (df['waterfront'] == 1)) & (df['condition'] == 5) & (df['yr_built'] >= 1980)
client_choice = df[condition]

client_choice['price'].min(), client_choice['price'].max()
```

```
(1295000.0, 3000000.0)
```

5.3

```
client_choice = df[(df['sqft_basement'] == 0) & (df['floors'] == 2) & (df['price'] < 150000)]
client_choice['condition'].mean()
```

```
2.8333333333333335
```