Why 1.5 is used in IQR rule:

Empirical Rule of Thumb:

The 1.5 factor is an empirically derived value, largely popularized by John Tukey, the inventor of the box plot. When asked about the choice of 1.5, he reportedly stated that "2 was too big, and 1 was too small." This suggests it was chosen as a reasonable compromise to identify potential outliers without being overly sensitive or insensitive.

Balancing Inclusivity and Exclusivity:

- Using a multiplier of 1 (i.e., just the IQR) would be too exclusive, potentially classifying too many valid data points as outliers.
- Using a multiplier of 2 would be too inclusive, potentially missing genuine outliers.
- 1.5 provides a balance, aiming to capture data points that are significantly far from the central mass of the data.

Central Tendency:

	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	sqft_above	sqft_basement	yr_built	yr_renovated
Mean	551962.988473	3.40087	2.160815	2139.346957	14852.516087	1.512065	0.007174	0.240652	3.451739	1827.265435	312.081522	1970.786304	808.608261
Median	460943.461539	3.0	2.25	1980.0	7683.0	1.5	0.0	0.0	3.0	1590.0	0.0	1976.0	0.0
Mode	0.0	3.0	2.5	1720	5000	1.0	0	0	3	1010	0	2006	0

Summary of the above central tendency descriptive data

- Most of the homes have 3 and 2 bedrooms
- The average build year of the house is 1970s
- Median house size is 2000 sq ft lot
- Most of the homes do not have waterfront area

Percentile

	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	sqft_above	sqft_basement	yr_built	yr_renovated
Mean	551962.988473	3.40087	2.160815	2139.346957	14852.516087	1.512065	0.007174	0.240652	3.451739	1827.265435	312.081522	1970.786304	808.608261
Median	460943.461539	3.0	2.25	1980.0	7683.0	1.5	0.0	0.0	3.0	1590.0	0.0	1976.0	0.0
Mode	0.0	3.0	2.5	1720	5000	1.0	0	0	3	1010	0	2006	0
01:25%	322875.0	3.0	1.75	1460.0	5000.75	1.0	0.0	0.0	3.0	1190.0	0.0	1951.0	0.0
Q2:50%	460943.461539	3.0	2.25	1980.0	7683.0	1.5	0.0	0.0	3.0	1590.0	0.0	1976.0	0.0
Q3:75%	654962.5	4.0	2.5	2620.0	11001.25	2.0	0.0	0.0	4.0	2300.0	610.0	1997.0	1999.0
99%	2005220.0	6.0	4.5	5180.1	203455.49	3.0	0.0	4.0	5.0	4460.1	1700.2	2014.0	2014.0
Q4:100%	26590000.0	9.0	8.0	13540.0	1074218.0	3.5	1.0	4.0	5.0	9410.0	4820.0	2014.0	2014.0

- Most homes are under $655K, with outliers in the millions.
- Typical homes have 3–4 bedrooms. Upto 75th %tile are 4 bedroom house only
- Majority of homes are in average (3) to good (4) condition. Yes Good conditional homes are having in Q4 only.
- Waterfront properties are rare.
- A large range, but 75% of homes are under 2,620 sq ft.
- Most homes have no designated view rating. Upto 75th %tile are no review
- Most homes were built between 1950–2000

Skewness & Kurtosis

	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	sqft_above	sqft_basement	yr_built	yr_renovated
skew	24.790933	0.456447	0.616033	1.723513	11.307139	0.551441	11.682901	3.341586	0.959068	1.494211	1.642732	-0.502155	0.385919
kurtosis	1044.352151	1.235377	1.865905	8.291683	219.872987	-0.538852	134.548673	10.464178	0.19773	4.070138	4.08238	-0.670076	-1.851111

1. **Price**, **sqft_lot**, and **waterfront** are **extremely skewed and have extreme kurtosis** because of many outliers
2. **Bedrooms**, **bathrooms**, **floors**, and **condition** are close to normal but slightly right-skewed
3. **Year renovated** has very low kurtosis. Because homes are not renovated