

Critical Summary:

Uncertainty in dataset

This dataset has high uncertainty in the categorical columns like Built in Year, No of bedrooms , Floors. I removed outliers from Built in Year and kept it between 1980 to 2025 , For the Bedrooms and Floors there are values which could be considered as outliers, Similary in the price section some prices doesnt match with the Purpose.

Based on the given data I converted the string price and areas columns to numeric and dropped the unnecessary columns like title, description. I removed outliers based on IQR plot.

Insights

The provided data shows that Lahore and Gujranwala has highest prices for houses in 'For Sale' Category and Karachi and Islamabad has highest prices for houses in 'For Rent' Category.

The correlation between property type and prices across cities shows that in almost every city and almost every house type has positive correlation with price. Farmhouses in Faisalabad and Gharo showed an inverse trend, which is suspicious as the farmhouse in Lahore shows a very strong correlation between size and prices.

For three major cities, Islamabad, Karachi and Rawalpindi, in the house category. The price increase as number of bedrooms increases and then it starts decreasing as number of bedrooms increases past 9 rooms. In Karachi it doesn't upper portion and lower portion property types follow the same trend. They keep on increasing to 5 bedrooms and then price decreases.