Analysis of Property Data Set

Justin Gatonby

2024-11-26



# Analysis of Property Prices by Type

**Price Trends**

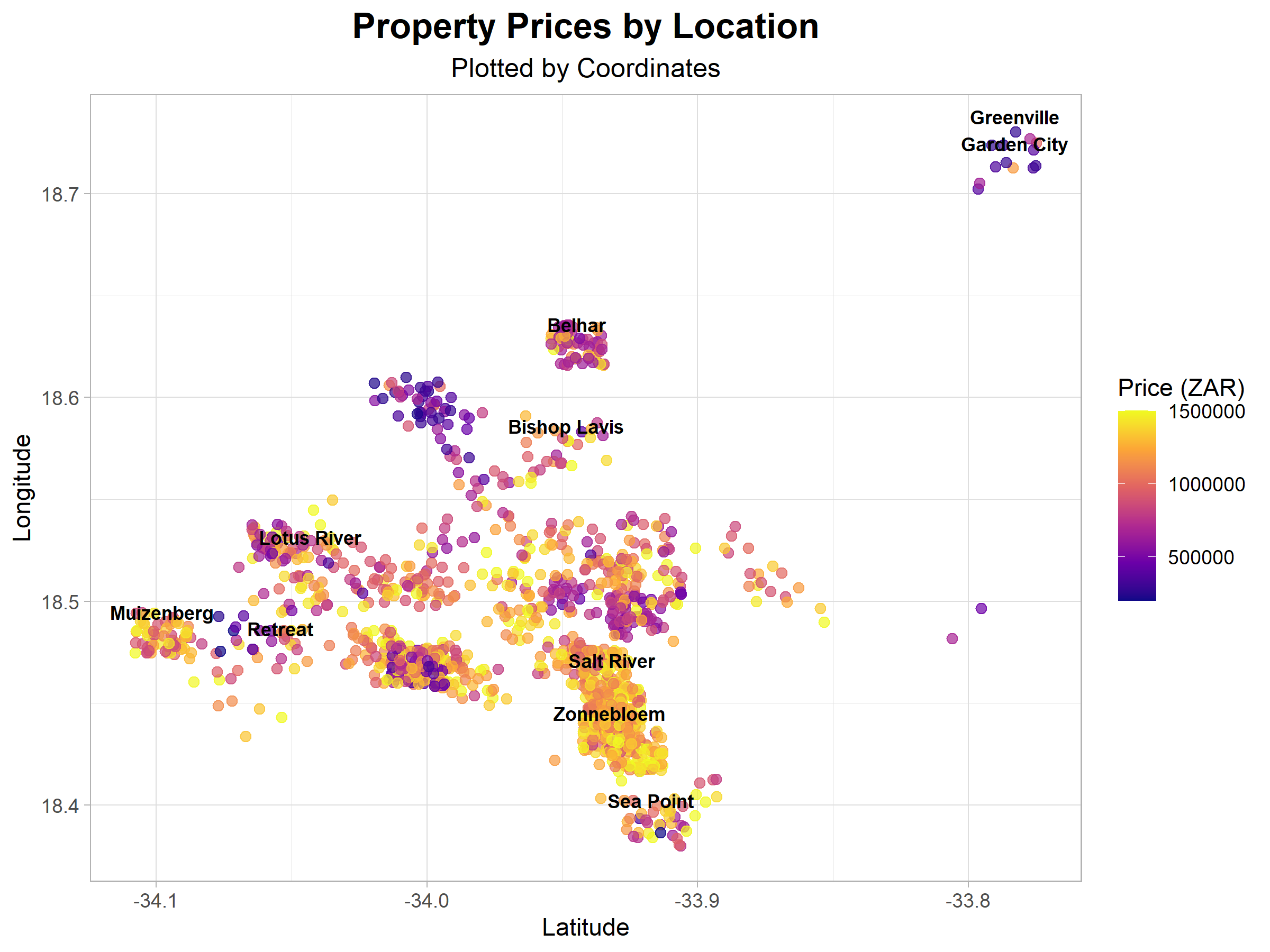
* Indicates properties with greater amenities, development, and complexity command higher prices.
* Suggests apartments and townhouses are typically more expensive than vacant land and houses.

**Distribution**

* Vacant land has a symmetrical distribution, with a large range relative to its interquartile range, suggesting that there are outliers on both ends (overvalued and undervalued).
* Houses have the largest interquartile range, suggesting that other variables will be needed to determine their price.
* Apartments have a long tail extending towards lower prices, displaying an opportunity to identify undervalued properties.
* Townhouses are negatively skewed, with fewer townhouses priced significantly lower than the median price and fewer opportunities to find undervalued properties.

**Composition**

* The bulk of the data set is composed of apartments (66,4%), followed by houses (28,2%).
* Land (2,5%) and townhouses (2,8%) make up a far smaller portion of the dataset.



# Analysis of Property Prices by Location

**Price Trends**

* Location is a strong predictor of property price.
* Areas around the coast tend to be more expensive, while those further away are on the lower end of the price range.

**Distribution and Composition**

* Most properties in this data set are clustered around the coast.
* As longitude increases, measuring similarity becomes more challenging, as property dispersion and price variability increases.



# Analysis of Property Prices by Bedrooms and Bathrooms

**Price Trends**

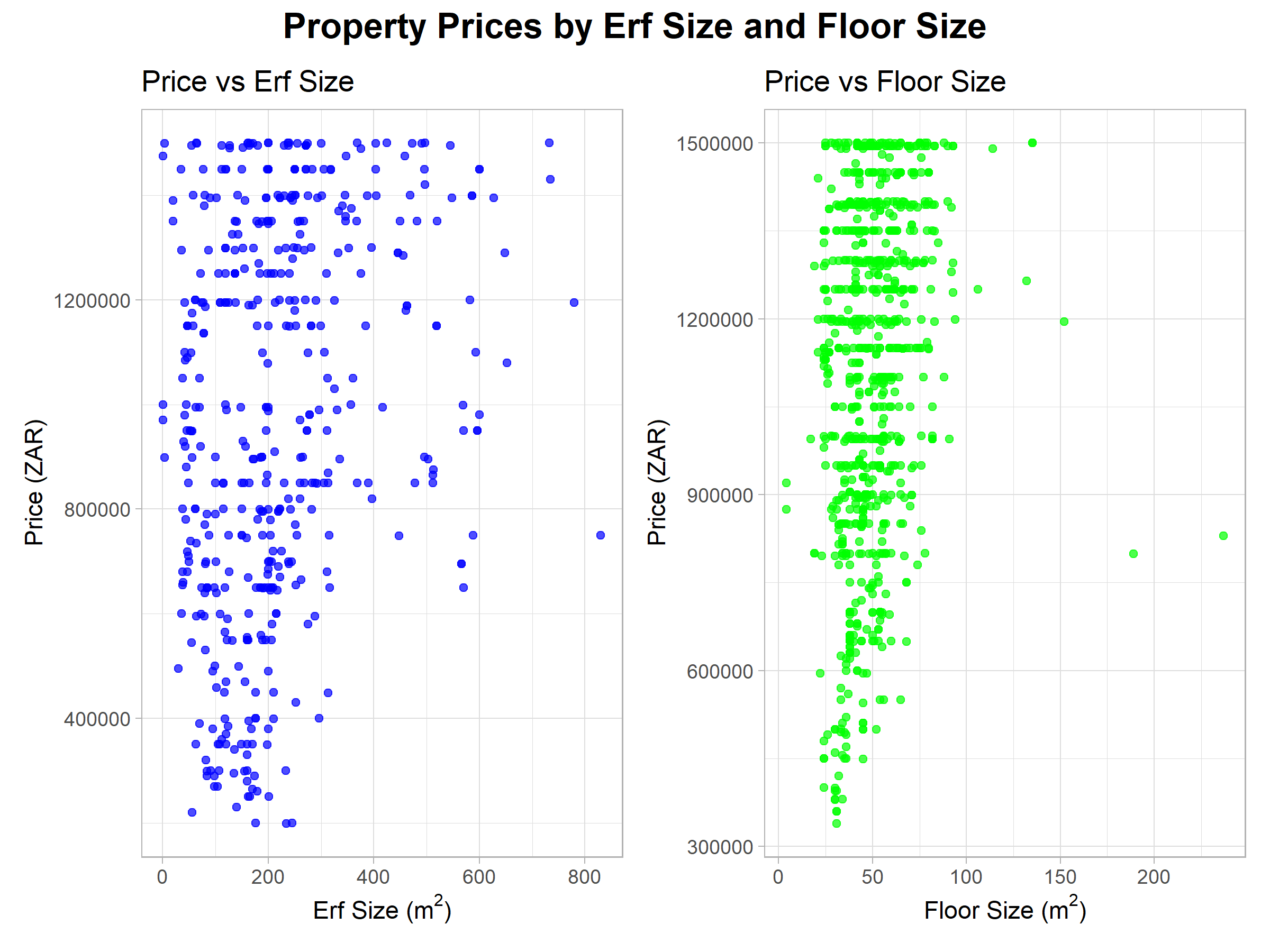
* Properties with more bedrooms and bathrooms tend to demand higher prices.
* However, more bedrooms will only correspond with higher prices, if the number of bathrooms increases as well.

**Distribution**

* Most points are clustered around combinations of 1-3 bedrooms and 1-2 bathrooms.

**Outliers**

* Extreme observations have been removed from the plot, that is, properties with more than 7 bedrooms or more than 4 bathrooms.
* Zero bedrooms and bathrooms is not an anomaly in the data, but rather an indicator of vacant land.



# Analysis of Property Prices by erf and floor size

**Correlation**

* Both erf size and floor size show a positive correlation with price, with larger properties commanding higher prices.
* The strength of the relationship between erf size and price (r=0.26) and floor size and price (r=0.24) is similar.

**Outliers**

* There are many outliers in both plots at both ends.
* Many small properties are priced at the top of the price range, while some of the largest properties are at the lower end.

**Reliability**

* Although there is some correlation between each variable and price, the weakness of the correlation and the number of outliers make size on its own a weak predictor of property price.

# Summary

* The coordinates of a property appear to be the best predictor of its price.
* Therefore, a k-nearest neighbor model should produce accurate predictions of property price.
* The model may use other informative variables such as number of bedrooms, number of bathrooms, erf size, floor size and number of parking spaces as well.
* Categorical variables may be excluded to avoid the “Curse of Dimensionality” and challenges converting them to numeric features.
* This should not affect model performance as the location is described by coordinates, while the number of bedrooms and bathrooms are indicative of property type.