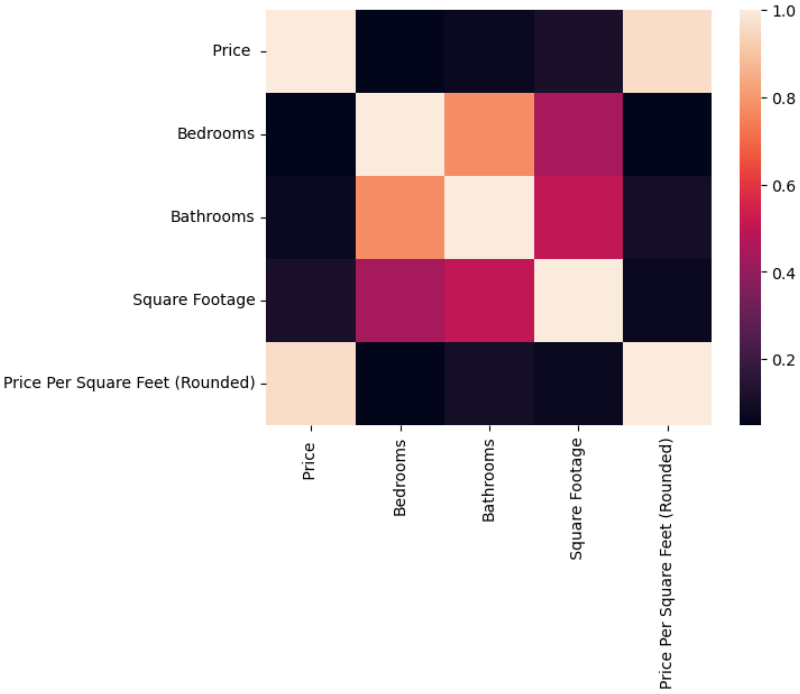
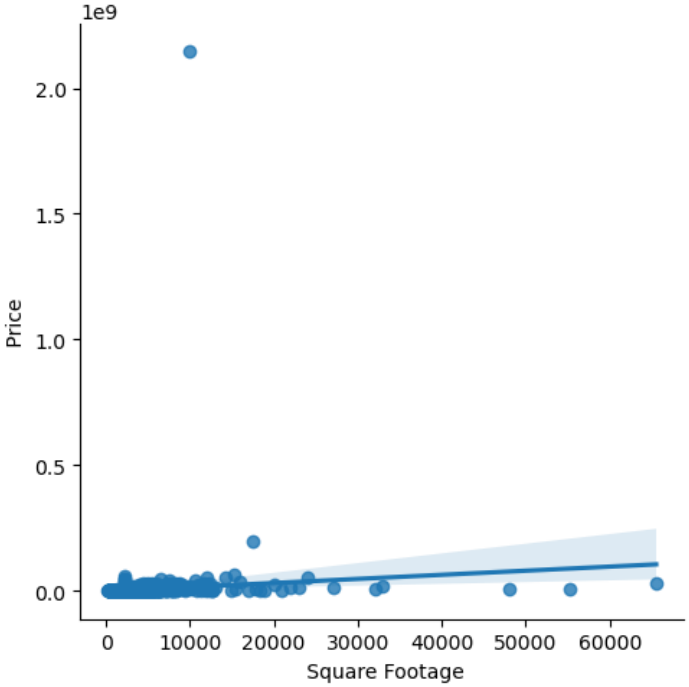
Task 6.2 – After viewing visualizations

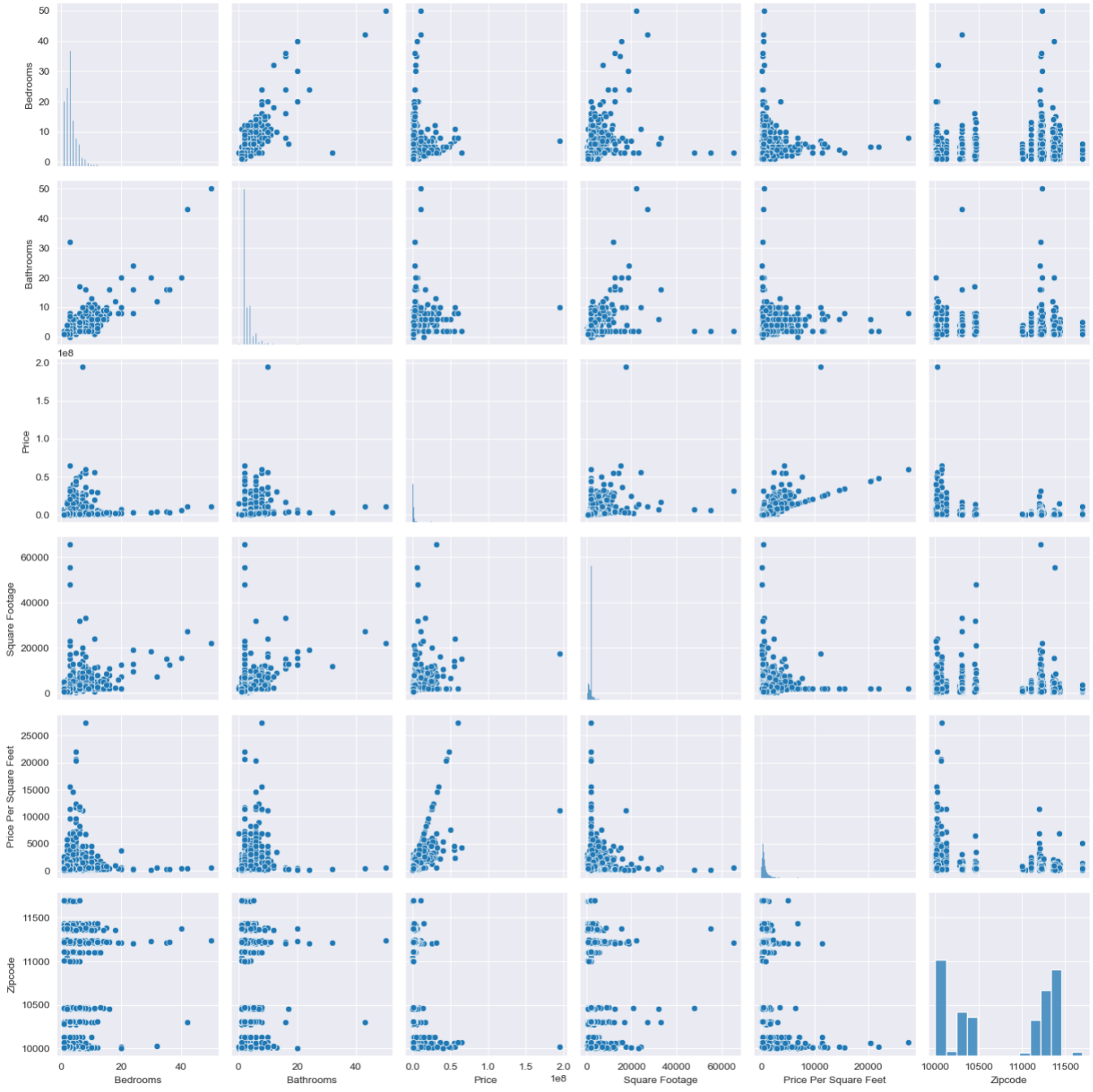
Visualization #1: Heatmap



Visualization #2: Scatterplot:



Visualization #3: Pair Plots



Step 7/8: Revisit the questions you generated in the previous task and write answers to those you can based on the exploration you've conducted so far. Add any new questions that may have arisen based on the early findings in your visual exploration. Create hypothesis.

From the heatmap's results we learn a surprising relationship, that Price of the property has minimal correlation with bedrooms, bathrooms, and square footage. One may simply think big house, more expensive. Small house, less expensive. However, within our current data set and circumstances that is not the case. My hypothesis is that this is because Location & Price has the highest correlation. We will need to setup parameters to compare different locations, perhaps each specific zip code and average pricing within that zip code. For example look up each and every zipcode within the data set that we have. Then from having the average price of each zip code, going further and finding the average price of "insert property type" within each zip code. EX: "Average price of condos within 10090", "Average price of single family houses within 10090".

From our scatterplot we identified there is a small positive correlation between square Footage and price. I expected a greater correlation. Nonetheless we should look further into price per square feet to answer questions such as "Is this property in the Low-Medium-High category of price per square feet". For example if a property's price per square feet is below $400 it can be considered Low, from $400-$700 medium, $800+ being High. For easy filtering of price per square feet. Also combining the factor above of location, we should answer questions like: "What is the average price per square feet for properties in 11211?". This may require data from another data set.

From our pair plot is that there are relationships between price per square feet and price as well as bathrooms and bedrooms. Neither of these are surprising and completely make sense. However to be thorough we should briefly look into this. What sort of bathroom to bedroom ratio should be expected? When is it too low or too high? For example there are properties with 5 bedrooms and 1 bathrooms, which seems to be not practical & should be flagged as something to look out for – for the potential buyer. Categorically separating "Viable bedroom to bathroom ratio" to a "yes" or "No" is a possible flag we can include in the data.

To summarize what the next questions to answer are:

1. Identify the average prices for each zipcode.
2. Identify the average prices of "Insert property type" for each zipcode
3. Is this property in the Low, Medium, or High category of price per square feet?
4. What is the average price per square feet for properties in "Insert zipcode"?
5. What is a viable bedroom to bathroom ratio? What properties do NOT have a sufficient bathroom to bedroom ratio?