**3.5 Descriptive statistics**

In this section, we will examine the more apparent patterns of our collected data. This is done to familiarize ourselves with the data, before we commence on the analysis.

**3.5.1 Key Statistic**

Some key statistical characteristic is presented in table 1:



Initially, it should be addressed that some of the properties has 0 rooms and consists of 0 . This is due to the fact that we have also included land on which housing has not yet been build. Examining the 25%-quintile and the 75%-quintile of the valuation prices, it becomes apparent that there are some substantial outliers both to the cheaper and to expensive side. This is also the case with rooms and , where the highest values are sizably higher than the 75%-quintile.   
It is worth noting, that there is quite a big difference between the lowest average municipal income of DKK 257,776. and the highest of DKK 583,331. This difference becomes further noteworthy when assessing the 75%-quintile. The difference from the 75%-quintile to the highest average municipal income is more than 4 times the difference of the 75%-quintile and the lowest income. The income distribution is visualised in the following figure:

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

The boxplot illustrates that the distribution is heavily left-skewed, where some municipalities has a sizably higher average income than the rest of the Danish municipalities.

**3.5.2 Prices in Municipalities**

We plot the average square meter valuation price in the different municipalities:

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

It becomes apparent that the distribution is quite left-skewed and that there are a few municipalities who’s price per square meter is much higher than the rest of the municipalities. Recalling the geoplot in figure 1, these expensive outliers are heavily consentrated around- and north of Copenhagen.

The distribution looks similar to that of figure XX (The previous boxplot), though we cannot declare any correlation between average municipal income and average municipal price per square meter. Nonetheless, it becomes apparent that the valuation price of property is highly discriminated by municipal factors. The scope of this assignment is exactly to examine these factors and attempt to use these to evaluate an unseen, out-of-sample property.

**3.5.3 Property Type**

Another worthwhile consideration is that we have included all types of properties. It would be reasonable to assume that there is an average difference in valuation pricing depending on the type of property. Figure XX displays the median valuation price for each type:

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

It is interesting to note that the most expensive properties are apartments as opposed to houses. This is not especially surprising though, as it is a well-established trend that real estate prices in major cities are skyrocketing. Refraining from delving deeper into to a discussion of global urbanization, we retain the fact that property type does have an apparently significant effect on valuation pricings on average. We will include the ‘*type’* feature in our impending model training to control for this effect.