**Answers**

**2.1**

3.

I chose to drop the 'date' and 'yr\_renovated' features. Both of them had close to 0 correlation with the price, and so I drop them for the simplicity of the algorithm.

I kept all the other features, because all the other ones had some correlation, and even so some of them had very small correlation, the optimal solution to the problem would give them small weights.

I designed three new features – 'grade15', 'view15', and 'waterfront15'.  
this three new features suppose to demonstrate the quality of the neighborhood- even though only the houses that are coplitly adjust to the waterfront – and to the view – its reasnalbe to assume that the whole area is more pricy due to the closing to the view and water-source. The 'grade15' suppose also to estimate the quality of the nighborhoods – nighborhoods with houses at good contioins tends to be more pricy, and also its reasonable to assume that If the majority of the houses in the area are in some shape, the other ones that close to this condtion.

Regards invalid or missing values – I tried two approaches, one was to correct the data (for example by changing 'yr\_renovated' to 0 when 'yr\_renovated' is later before 'yr\_built'), and the second one was to delete the unvalid rows.