**Housing Prices in California Project Checklist:**

1. Frame the Problem and Look at the Big Picture
2. The objective in business terms.
   1. To build a model to predict district’s median housing price.
3. How will the solution be used?
   1. The solution will be used to determine whether it is worth investing in a given area or not.
4. What are the current solutions/workarounds (if any)?
   1. The district housing prices are currently estimated manually by experts: a team gathers up-to-date information about a district, and when they cannot get the median housing price, they estimate it using complex rules. This is costly and time-consuming.
5. How should you frame this problem (supervised/unsupervised, online/offline, etc.)?
   1. It is clearly a typical supervised learning task since we are given labeled training examples. Moreover, it is also a typical regression task, since you are asked to predict a value. So, it is a **multivariate regression problem**.
   2. Finally, there is no continuous flow of data coming in the system, there is no particular need to adjust to changing data rapidly, and the data is small enough to fit in memory, so plain **batch learning** should do just fine.
6. How should performance be measured?
   1. We will use Root Mean Squared Error (**RMSE**). The higher the value means large errors the system is typically making in its predictions
7. Is the performance measure aligned with the business objective?
   1. The RMSE is more sensitive to outliers. But when outliers are exponentially rare, the RMSE performs very well and is generally preferred
8. What would be the minimum performance needed to reach the business objective? < 10%
9. List the assumptions you (or others) have made so far.
   1. We need actual prices and not categories