## 數值型特徵-補缺失值與標準化

* 在處理NA值的時候需要根據不同的features所擁有的資料意義以及資料的類型來決定要使用何種方式來對資料進行填補，主要的方式有兩種:

1. 填入”統計值”
   1. Mean:資料分布沒有特別集中時
   2. Median:資料分部有明顯的”偏態”(也就是靠左集中或靠右集中)
   3. Mode:類別型的欄位
2. 填補”指定值”:要填補指定值就必須要對該feature的資料意義有一定以上的了解，而且要每個NA值都填入一樣的指定值便於分辨那些原來是NA(?)

* 接下來要處理資料分布的問題希望可以平衡features的影響力，就是之前說過的標準化或最小最大化，他們的差異就在於使用時，最大最小化會受到"outlier"的影響，所以如果想要使用最好先整理過outlier。還有標準化會對非樹狀ML model的預測會有影響。

