Day 5 資料清理數據前處理 EDA之資料分布

EDA - 統計量化的方式?

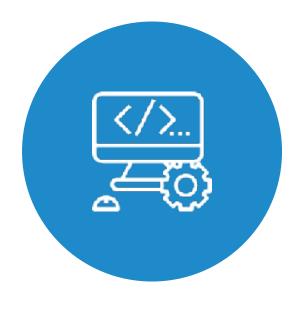


以單變量分析來說,量化的分析方式可包含

- 計算集中趨勢
- ●計算資料分散程度
- ●平均值 Mean
- ●最小值 Min
- ●中位數 Median
- •最大值 Max

● 眾數 Mode

- 範圍 Range
- •四分位差 Quartiles
- •變異數 Variance
- •標準差 Standard deviation



基本上使用上述統計特徵就可以讓我們初步了解資料的樣子,並且觀察是否有異樣

EDA視覺化的方式?

有句話「一畫勝千言」,除了數字,視覺化的方式也是一種很好觀察資料分佈的方式,可參考 python 中常用的視覺化套件

畫圖沒靈感的時候可以到這兩個套件的範例網頁逛逛!

matplotlib

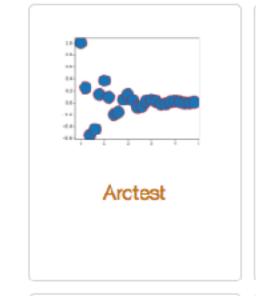
seaborn

Gallery

This gallery contains examples of the many things you can do with Matplotlib. Click on any image to see the full image and source code.

For longer tutorials, see our tutorials page. You can also find external resources and a FAQ in our user guide

Lines, bars and markers

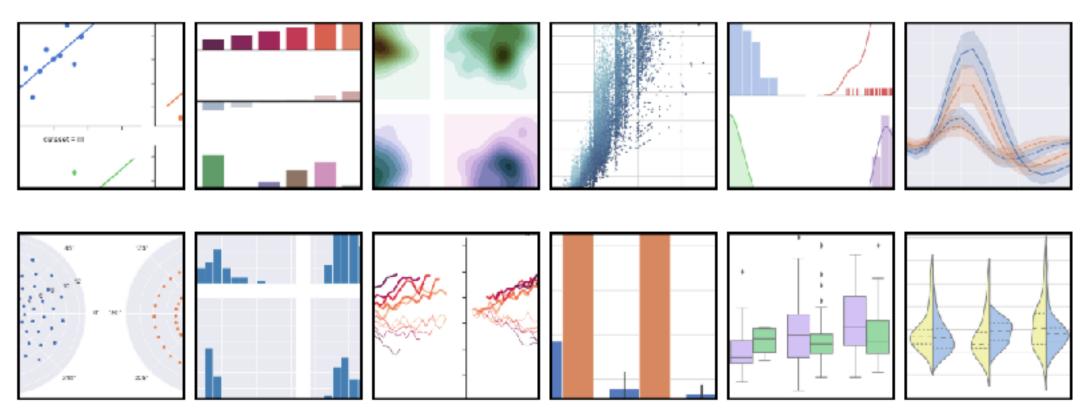








Example gallery





請跳出PDF至官網Sample Code&作業開始解題

