

K-means klustering

- Läs in diabetes.csv. Inspektera datan med `head()` och använd `sns.pairplot` för att visualisera distributionerna. Använd "Outcome" från dataframen som hue-parameter (label). Tolka resultaten.
- Skapa K-means kluster. Testa först med 3 kluster. Verkar den fungera bra? Kom ihåg att standardisera datan!
- Välj sedan flera kluster. Skapa K-means där ni börjar med 1 kluster och ökar med ett kluster i taget upp mot 10 kluster. Plotta SSE för att se om ni hittar "elbowen", dvs knäcken i ploten. Tips här är att använda en for-loop och spara `KMeans.inertia_` i en array inuti loopen.
- Beräkna även silhouette score för upp till 10 klusters. Det görs genom att använda metricen `silhouette_score` och jämföra datan med vilka kluster de har hamnat i (`labels_`). Tips är att spara alla silhouette scores i en array i for-loopen.
- Hur många klusters tycker ni är bästa valet?

Lösningsförslag kommer på slutet av dagen!