**Opgave1 :**

1. Vurder datasættet fra rainforrest, beslut om du vil fjerne kolonner eller impute værdier, eller en blanding. Lad os se, hvad der giver det bedste resultat.
2. Del sættet op i samples og labels, så du har et datasæt, som kun består af talværdier for de enkelte træsorter.
3. Udfør en clustering med KMeans
   1. Hvor mange clusters skal vi inddele den i?
   2. Hvordan finder vi et godt antal?
4. Sammenhold wood mod branch i et scatterplot
5. Udfør en cross tabulation
   1. Hvor gode er dine resultater?
6. Brug en StandardScaler og en Pipeline til at få bedre resultater
   1. Test igen med cross tabulation
7. Check for intrinsic dimensions
   1. Lav en dimensionality reduction der svarer til bedste bud
   2. Blev dine resultater bedre?

**Opgave2:**

Classify song genres from audio data (fra datacamp projects)