**Feedback EDA Rienk Heins**

**Onderzoeksvraag:** *Using machine learning, can the malignancy of a breast clump be predicted by the cell and nuclei attributes and which are the best predictors?*

**Voorstel:** *Can the malignancy of a breast clump be predicted by the cell and nuclei attributes and which are the best predictors?*

Het is niet nodig om de te gebruiken techniek te noemen in de onderzoeksvraag.. Dit wordt vaak genoeg verteld in het verslag. Dus om het iets overzichtelijker te houden zou ik dit weglaten.

De data wordt verder uitstekend omschreven en het doel van dit project is duidelijk aangegeven.

**Tabellen:**

Verder zou ik de tabellen knitten met iets als Kable, zodat het net wat mooier en overzichtelijker wordt. Het is nu heel moeilijk te volgen wat elke kolom zegt. Ook mist er een codebook in het bestand. Deze tabellen zou je ook beter een caption kunnen geven, zodat je later bij results kan terugverwijzen naar deze tabllen.

**Plots:**

Alle plots zijn goed weergegeven met juiste captions en duidelijke omschrijvingen. De gegeven results zoals: *## `geom\_smooth()` using formula 'y ~ x'* kan weggelaten worden door in de chunk (of in het algemeen bij chunksettings) results = false in te stellen. Dit zorgt ook weer voor een iets nettere pdf.

Prima formulering verder en reproduceerbaarheid is prima. Bronnen verwijzing in de tekst ziet er goed uit.