Correzioni Report DM 1

* “The data needs to be transformed for a better analysis”:
  + alcune variabili sono state trasformate, elencarle
  + aggiungere questa frase in variable trasformations
* “This approach is preferred because the boxplots would be overly compressed and less informative.”:
  + Per gli outlier risentire audio prof per vedere se vanno bene
  + Specificare quali e quali variabili
* “Variable transformations”:
  + Le reviews sono state aggregate tra loro
  + Sono state rimosse le varaibili inutili come best e worst rating
  + Dire che la log trasform è stata fatta anche per migliorare la correlazione, PCA e clustering generalemente e dire perché
* “Pairwise correlations and eventual elimination of variables”:
  + Parlare dei valori delle correlazioni e dei risultati ottenuti
* ACP:
  + Da aggiungere
* Clustering:
  + Specificare motivo di scelta di queste 3 variabili (centroidi lontani)
* Analysis by centroid-based methods:
  + Chiedere Prof se dopo la log, nel nostro caso è meglio la z score o la min max
  + Specificare perche si fa la standardizzazione dei dati prima del clustering

Domande prof:

* Meglio mettere grafico bad frequenze variabili oppure le frequenze medie del rating e titletype
* Chiedere se va bene la gestione dei valori nulli
* Dobbiamo fare il Boxplot e iqr per outlier oppure basta la scattermatrix
* È giusto togliere le righe che hanno totalCredits = 0? Per noi è wrong data
* Chiedere Prof se dopo la log, nel nostro caso è meglio la z score o la min max
* Quando ci da tabella con con tutti i rating per la regression?
* Far vedere risultati k-means
* Dbscan eps e minpoints. È meglio fare una grid search?
* Opinioni sulle selezioni di alcuni variabili per il k-means