Dissertation outline

(1) Introduction
   (a) Reasons for performing clustering (with examples of applications in the real world)
   (b) Taxonomy of clustering algorithms and theoretical introduction to the fundamental ones (kmeans, dbscan, hierarchical)
   (c) Talk about the goal of the dissertation which is to do a fair comparison (from a theoretical and computational point of view) of the traditional algorithms (ones mentioned before + spectral clustering, BIRCH, meanshift, optics, minibatch kmeans, GMM, k medoids, CLIQUE/BANG (example of grid based clustering), BSAS/MBSSAS/TSSAS (example of sequential based clustering) versus more recent and promising algorithms (hdbscan, quickshift++, classix, genieclust and possibly more when found on the way)
   (d) Talk about how some traditional implementations happen to be suboptimal and their runtime/number of calculations can be cut with the usage of mathematical tricks
   (e) Why fair clustering benchmarking is important and why in many cases it is conducted in the wrong way (https://publikationen.bibliothek.kit.edu/1000073378)
   (f) Touch on different evaluation measures
(2) Literature review
   (a) Bring up different articles that contain instructions for fair benchmarking (https://arxiv.org/pdf/1809.10496v2.pdf) and the main points that we implemented from them (for example careful choice of datasets with justification for the intended scope of generalisation)
   (b) Discuss examples of some other clustering benchmarking/comparison papers (for example https://link.springer.com/chapter/10.1007/978-3-030-60104-1_20 https://upcommons.upc.edu/bitstream/handle/2117/23414/R13-8.pdf https://doi.org/10.1016/j.ejor.2005.03.039) comment on approaches taken, evaluation measures used, and point out aspects of fair benchmarking overlooked where appropriate
   (c) Comment on current progress on the acceleration of some traditional algorithms. For kmeans reference algorithms such as Elkan's, Hamerly's, Annular, Exponion and Shallot. For dbscan, mention https://doi.org/10.1016/j.patcog.2016.03.008 (2016) and possibly more recent one if found. For hierarchical probably this one https://www.jstatsoft.org/article/view/v053i09
(3) Methodology
   (a) In each separate subsection, go through the theory of every aforementioned algorithm, and based on that theory comment on the caveats, characteristics of data where a certain algorithm is expected to perform well and pay attention to the use of linear algebra.
   (b) Go through the theory of the acceleration algorithms similarly to above
   (c) Introduce our benchmarking design: what environment was used, implementations used, datasets (synthetic and real world) and explain the evaluation metrics used
(4) Results
   (a) Present the results in the form of tables/plots of clustering/heatmaps/line charts and comment on each result if it converges with our theoretical expectations and if no then what might be the reason
(5) Discussions

    (a) Reiterate the reasons for certain approach choices (datasets, algorithms, and evaluation metrics) and suggest further work
(6) Conclusions
    (a) Reiterate findings from the results section