**I. Pen-and-paper**

1. You can use OpenDocumentText, Word or Latex to write your report.

We suggest Cambria Math font with 11pt for main text, using single line spacing and 6pt paragraph.

You can use 10pt in tables and formulas, e.g.

You can paste images of a handwritten solution, yet guarantee their high-quality resolution and contrast.

1. Answer 2
2. Answer 3
3. Answer 4

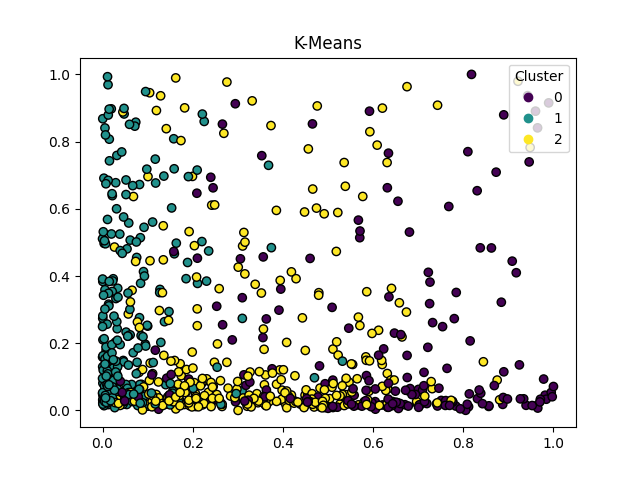
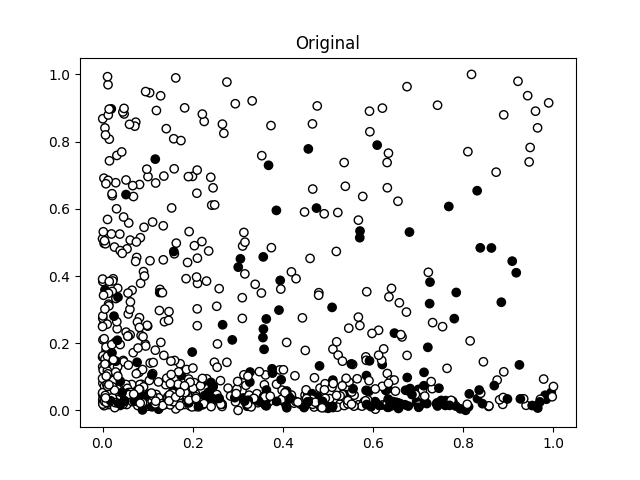
**II. Programming and critical analysis**

1. For state = 0 K-means purity score is 0.7671957671957672 and silhouette score is 0.11362027575179426

For state = 1 K-means purity score is 0.7632275132275133 and silhouette score is 0.11403554201377068

For state = 2 K-means purity score is 0.7671957671957672 and silhouette score is 0.11362027575179426

1. The K-means algorithm is non-deterministic since the initial k number of centroids start in a random position, then the objects are reassigned to the closest centroids and recalculates these centroids in relation to the new data points. These steps are repeated enough times until a consistent result is obtained or the maximum number of iterations is reached.



1. 31 principal components were needed to reach a variability of 0.8006422402169662.

**III. APPENDIX**

