Q1.1

Using the a* to search a tree of states of the game where the children generated from a state node are the possible moves with a heuristic that leads to getting closer to the finished state node.

Q1.2

1. The A* search algorithm is a shortest path finder within a graph (subset tree) from A to B, with additional heuristics that provide a gauge for progress and a better knowledge of which node to traverse to next as it will be closer to the goal.

NOTE:
Format for input of a* implementation
D or M for choice of admissible heuristic used

Example input for grid

1 2 0
3 4 5
6 7 8

2. Number of tiles out of position, is admissible as a heuristic as it shows every tile should be moved at least once.

   Sum of the manhattan distances, also known as the sum of the distance of the tile distances from its corresponding goal stat, will show minimum distance for each tile to move closer to their end states and the sum is an accurate and admissible heuristic since in every move, one tile can only move closer to its goal by one.

3. Two optional implementations for seperate heuristics
   initial_state = [['0' ,'2', '4'] ,['5', '7', '6'], ['8', '3', '1']]
   goal_state = [['0' ,'4', '6'] ,['2', '7', '1'], ['5', '8', '3']]
   Takes 8 moves

4. Both give the same best result but the manhattan is just more complicated to calculate but appears to have no noticeable performance issues. The manhattan distance is overall more accurate which may have an advantage on a variation of the game on a larger scale.

Q2.1

K means clustering with recognition of hand writing works by taking in the pixels on a 2d plain which represents a letter and it will be fitting to a letter it is closest to, with k in this case being a fixed 10 (0-9).

Q2.2

I presented the efficiency in terms of time and quality of the result with the built in accuracy and inertia value, which is trying to be minimised.
The graphs I plot show that the average of the centers of the clusters somewhat depict all the digits.
However the confusion matrix shows and supports what was seen in the centers that the prediction will get confused between 8 and 1.

Q2.3

Limitations of k-means clustering are that the user has to specify k and that it can only take numerical data.
So even in data where everything is categorical it could not be categorized unless quantitatively so.
If the k is not specified k-means clustering would not work and additional calculations or human intervention would be needed.