a. The 2 most important properties the activation functions have is that they have to be differentiable and non-linear. The latter is to be able to learn non-linear relationships, which makes the model much more applicable. The need for differentiability is to be able to apply the backpropagation algorithm. The backpropagation algorithm relies on the chain rule of derivatives to update weights. If an activation function was non-differentiable, weights prior to that layer wouldn't be able to get updated.

b. The advantage of using ReLU over sigmoid is due to the fact that for outliers the derivative of sigmoid is nearly 0 which causes some weights not to update at all. ReLU handles this problem well for positive values, however it still has the same issue for negative inputs. Using Leaky ReLU, can somewhat handle the issue ReLU has with negative inputs while preserving the ReLU's good performance for positive inputs.