K-means

Group 04:

1. Kummer Lorenz, student ID – 00947188
2. Sidak Kevin, student ID – 01249373
3. Higgs Sam, student ID – 11844757
4. Setezhev Grigory, student ID – 11808684

# Introduction

K-means algorithm is used in the solution of the clustering problem. It aims to partition *n* given data points (or observations) into *k* different clusters. This algorithm dates back to 1957 by Stuart Lloyd and 1965 by E. W. Forgy.

# Programming setup

Python (version 3) was used as a programming language for this assignment. It was chosen for it’s ease of use and comprehensive list of available packages which could simplify the development of the program. Package-management system **pip** allowed us to use the following packages:

1. NumPy – library which supports multi-dimensional arrays and matrices and also provides the functions to operate on them. We also used it as a tool to import data from TXT and CSV files.
2. Scikit-learn – machine learning library which gave us the ability to calculate the Normalized Mutual Information score.
3. Matplotlib – plotting library which was used by us to visualize the clustering result.

# Project structure