Homework 1

1-1.   Read Chapter 1 of Python Machine Learning book by Sebastian Raschka.

1-2.   In this problem, we will look at the Singapore population statistics collected by the Singapore government. Download Singapore population data from 1960 to 2018 at https://www.singstat.gov.sg/
   a. Graph the total population vs year.
   b. Use linear regression to build an estimator the population of Singapore in the future. Use the data for years 2013 and earlier as training data.
   c. What are the slope and y-intercept of the best fit line? What is the mean squared error of the estimator on the training data?
   d. Use years > 2013 as test data and predict the data for those five years. What is the mean squared error of the estimator on the test data?
   e. What is your estimate of Singapore's population in 2030 and 2050? Do you think these estimates are reasonable?
   f. How could you improve your estimates of the future population of Singapore?
   g. Submit your ipynb notebook and pdf file in LumiNUS. Please name your files as Name_Matric_HW1.ipynb and Name_Matric_HW1.pdf.