Karatina University

BIT 316: Big Data Analytics (Take Away CAT)

DATE: 03/12/2025 **TIME: 0800 hrs – 1100 hrs**

Consider the *Education_attainment.csv* data-set. Column descriptions for this data are as follows:

- **Year:** The year this row represents. Note there may be more than one row for the same year to show the percent breakdowns by sex.
- **Sex:** The sex of the students this row pertains to, one of "F" for female, "M" for male, or "A" for all students.
- **Min degree:** The degree this row pertains to. One of "high school", "associate's", "bachelor's", or "master's".
- **Total:** The total percent of students of the specified gender to reach at least the minimum level of educational attainment in this year.
- **White / Black / Hispanic / Asian / Pacific Islander / American Indian or Alaska Native / Two or more races:** The percent of students of this race and the specified gender to reach at least the minimum level of educational attainment in this year.

a. Read this data into Python Pandas.
b. Determine the percentages for women vs. men having earned a Bachelor's Degree in 1980.
c. Analyze the data for the two most commonly awarded levels of educational attainment awarded between 2000-2010 (inclusive).
d. Compare the difference between total percent of bachelor's degrees received in 2000 to those in 2010.
e. Derive descriptive statistics for this data-set.
f. Using the *Seaborn* library **:**
    i. Plot the total percentages of all people of bachelor's degree as minimal completion with a line chart over years.
    ii. Plot the total percentages of women, men, and total people with a minimum education of high school degrees in the year 2009. Label the x-axis "Sex", the y-axis "Percentage", and title the plot "Percentage Completed High School by Sex".