**תרגיל ויזואליזציה**

בתרגיל זה נשתמש במאגר הנתונים [Student Alcohol Consumption](https://www.kaggle.com/datasets/uciml/student-alcohol-consumption)

ניתן למצוא את ה data set גם בחומרי הקורס

עליכם לפתור את התרגיל ולכתוב את התשובות לשאלות והקוד במחברת של קולאב

בדוק האם במאגר יש ערכי Nan. אם כן, טפל בהם

שאלה 1: משתנה יחיד קטגוריאלי

1. אילו משתנים במאגר הינם קטגוריאליים?
2. המחישו 5 מתוכם ע"י גרף countplot
3. צור 2 מאגרים חדשים שיסוננו לפי שם בית הספר (GP או MS). המחישו שוב את הגרפים עבור מאגרים אלו
4. צור 2 מאגרים חדשים שיסוננו לפי מגדר התלמיד. המחישו שוב את הגרפים עבור מאגרים אלו עם משתנים רלוונטיים להבדלים בין מגדרים.
5. רשמו את המסקנות שאתם מסיקים מגרפים אלו

שאלה 2: משתנה יחיד רציף

1. אילו משתנים במאגר הינם קטגוריאליים סידוריים?
2. המחישו 5 מהם ע"י הגרף boxplot
3. חזור על ההמחשה עבור שני המאגרים המסוננים לפי שם בית הספר
4. חזור על ההמחשה עבור שני המאגרים המסוננים לפי מגדר.
5. אילו מסקנות אתם מסיקים מגרפים אלו?

שאלה 3: 2 משתנים, קטגוריאלי ורציף

1. צרו 5 גרפים מסוג barplot ו 5 גרפים מסוג boxplot. על ציר ה y להיות משתנה סידורי
2. אילו מסקנות אתם מסיקים מגרפים אלו?

נושאים שונים מעניינים למחקר:

1. היחס בין משתנים השונים להרגלי השתייה. למשל איזה בית ספר, מגדר התלמיד, גיל התלמיד.
2. היחס בין משתנים שונים לציון הסופי. למשל: זמן הלימוד, קשרים רומנטיים, אינטרנט בבית, השכלת ההורים.
3. נתוני האם מול נתוני האב: השכלה, עיסוק