



## תרגיל 9 – Clustering

### תרגיל בונוס (רשות)

הוראות הגשה:

1. בתרגיל הבא יש לענות על השאלות באמצעות שימוש בקוד פייתון ושימוש ב-Scikit-Learn.
  2. יש להגיש את העבודה בזוגות בלבד.
  3. התרגיל יוגש כמחברת colab, כאשר לתיבת הגשה יש להגיש קובץ המכיל לינק למחברת עם הרשאות קריאה למייל [tamarin@post.bgu.ac.il](mailto:tamarin@post.bgu.ac.il) + [yaelhoc@post.bgu.ac.il](mailto:yaelhoc@post.bgu.ac.il)
  4. שם הקובץ יהיה מספרי הזהות של המגישים בצורה הבאה: זהות1\_זהות2  
במחברת הפתרון, יש לציין את מספר השאלה עליה עניתם עבור כל חלק בפתרון
- ענו על השאלות הבאות באמצעות הנתונים על [Credit Card for Clustering](#) ושימוש באלגוריתם KMeans:
1. כתבו פונקציה שמקבלת כפרמטר מספר K ומחלקת את המידע לK קבוצות
  2. כתבו פונקציה שמקבלת כפרמטר מספר K ומחלקת את המידע לK קבוצות. רק שהפעם, החלוקה לקבוצות תעשה לאחר נרמול כל העמודות לערכים שבין 0 ל-1
  3. כתבו פונקציה שמקבלת כפרמטר מספר K ומחלקת את המידע ל-K קבוצות ובנוסף, מחזירה את המרחק האוקלידי הממוצע ואת סכום המרחקים של כל הרשומות למרכזים שאליהם הנקודות שייכות לפי אלגוריתם KMeans.
  4. כתבו פונקציה שמקבלת כפרמטר מספר K ומחלקת את המידע לK קבוצות ובנוסף, מחזירה את המרחק מנהטן הממוצע ואת סכום המרחקים של כל הרשומות למרכזים שאליהם הנקודות שייכות לפי אלגוריתם KMeans.
  5. באמצעות הפונקציות שכתבתם בשאלות 3 ו-4, ועבור ערכי K שונים, ציירו שני תרשימים (כאשר ישנו תרשים עבור כל מטריקה), שבהם ציר ה-X הוא ערכי ה-K וציר ה-Y הוא ממוצע המרחקים של הרשומות למרכזים