**PCA**

<https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60>

**טעינת מידע ויצירת משתמש**

יצירת יוזר בגיטהאב (לזכור את הסיסמא!):

<https://github.com>

יצירת ריפוזיטורי חדש בגיטהאב-

Graphical user interface, text, application

Description automatically generated

לתת שם לריפוזיטורי, להגדיר כציבורי וליצור (אין צורך בהגדרות נוספות).

העלאת קובץ המידע שמצורף בזיפ (Mall\_Customers):

Graphical user interface, text, application, email

Description automatically generated

להעתיק את כתובת הקובץ לאחר התצוגה שלו בתור raw:

Graphical user interface, text, application

Description automatically generated

לאחר הלחיצה להעתיק את הכתובת. כתובת לדוגמא תיראה כך:

<https://raw.githubusercontent.com/EyalHadad/Files/main/Mall_Customers.csv>

(מתחיל במילה raw)

יש 3 דרכים לטעון קובץ–

<https://towardsdatascience.com/3-ways-to-load-csv-files-into-colab-7c14fcbdcb92>

אנחנו נשתמש באפשרות הראשונה.

**תרגיל מעבדה**

להפעיל את קובץ המחברת שמצורף לקובץ זיפ באמצעות colab.

להשלים את המלבנים החסרים

הערות על המחברת:

ב**data exploration** יש להחליף את נתיב הקובץ בנתיב שלכם מהגיטהב.

**בdata visualization** -> histogram –דוגמא להיסטגרמות של שלושת המשתנים.

**בdata visualization** -> count plot of gender– יש להציג תצוגה כלשהי של מין הלקוח\ה

**בdata visualization** -> Relation between features – דוגמא לקורלציות בין משתנים

**בdata visualization** -> heatmap – יש להציג קורלציות באמצעות heatmap

**בdata visualization** -> scatter plot – דוגמא להצגת נתוני גיל והכנסה תוך הפרדה לפי מין הלקוח\ה

**בClustering**-> דוגמא לבדיקת מספר הקלאסטרים לפי מדד בשם Inertia. בתרגיל זה נבחר 5 קלאסטרים.

יש להריץ אלגוריתם kmeans עבור 5 קלאסטרים ולהציג את התוצאות.

מומלץ להיעזר בדוקומנטציה של Sklearn וניתן להשתמש בדוגמאות כמו:

<https://towardsdatascience.com/k-means-clustering-with-scikit-learn-6b47a369a83c>

(אם צריך להתחבר\לשלם אפשר להפעיל במצב אנונימי)