**תרגיל ויזואליזציה 2**

בתרגיל זה נשתמש במאגר הנתונים [Boston Housing dataset](https://www.kaggle.com/datasets/arunjangir245/boston-housing-dataset)

במאגר זה כל שורה מייצגת נתונים על שכונה בבוסטון, ומכילה מדדים שמאפיינים את האזור ואת מחירי הדיור בו. המאגר מבוסס על מחקר אמיתי משנות ה-70, שמטרתו הייתה לחקור את הקשר בין גורמים דמוגרפיים, כלכליים וגיאוגרפיים למחירי הנדל"ן.

**תיאור העמודות (משתנים):**

1. **CRIM** - שיעור הפשיעה למספר תושבים בשכונה.
2. **ZN** - אחוז הקרקע באזור המיועד לבנייה למגורים על שטח של לפחות 25,000 רגל רבוע.
3. **INDUS** - אחוז שטח העסקות שאינו קמעונאי בשכונה.
4. **CHAS** - האם השכונה גובלת בנהר צ'ארלס (1 = כן, 0 = לא).
5. **NOX** - ריכוז תחמוצת החנקן (זיהום אוויר, במיליגרם לליטר).
6. **RM** - ממוצע מספר החדרים למגורים ליחידה.
7. **AGE** - אחוז יחידות המגורים שנבנו לפני 1940.
8. **DIS** - מרחק ממוקדי תעסוקה מרכזיים בבוסטון.
9. **RAD** - מדד של נגישות לכבישים מרכזיים. ערכים בין 1 ל 24. ערכים נמוכים, גישה מועטה או רחוקה לצירי תחבורה ראשיים, ערכים גבוהים, גישה ישרה ונוחה
10. **TAX** - שיעור מס רכוש ליחידת ערך (פר 10,000 דולר).
11. **PTRATIO** - יחס תלמידים למורים בבתי הספר בשכונה.
12. **B** - מדד חברתי-כלכלי הקשור ליחס האוכלוסייה האפרו-אמריקאית.
13. **LSTAT** - אחוז תושבים מהמעמד הכלכלי-חברתי הנמוך.
14. **MEDV** - המחיר החציוני של בתים בשכונה (באלפי דולרים).

ניתן למצוא את ה data set גם בחומרי הקורס

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

**שאלה 1: שני משתנים רציפים**

1. אילו משתנים במאגר הינם רציפים?
2. בחרו 3 זוגות של משתנים. המחישו אותם באמצעות כל הגרפים הבאים: scatterplot, lmplot ו jointplot kind=’hex’))
3. אילו מסקנות אתם מסיקים מגרפים אלו?

**שאלה 2: 3 משתנים – שימוש בפרמטר hue**

1. בחרו 2 משתנים נוספים
2. צרו שוב את הגרפים שיצרתם בסעיף הקודם עם המשתנים הנוספים.
3. אילו מסקנות אתם מסיקים מגרפים אלו?

**שאלה 3: 4 משתנים – שימוש בפרמטר size**

1. בחרו משתנה נוסף
2. צרו שוב את הגרפים שיצרתם בסעיף הקודם עם המשתנה הנוסף.
3. אילו מסקנות אתם מסיקים מגרפים אלו?

**שאלה 4: מטריצת קורלציות ומפת חום**

1. בחר מספר משתנים שאתה מעריך שיש להם קורלציה למשתנה medv
2. צור מטריצת קורלציות
3. צור מפת חום
4. באיזה משתנים אתה מוצא קורלציה גבוהה ל medv?