



הטכניון
מכון טכנולוגי
לישראל

Visualization

סוגי משתנים

- משתנה בדיד: יכול לקבל ערכים בודדים בתוך טווח.
 - יום בשבוע
 - מספר ילדים במשפחה
- משתנה רציף: יכול לקבל כל ערך בין שני ערכים נתונים.
 - משקל
 - כמות משקעים

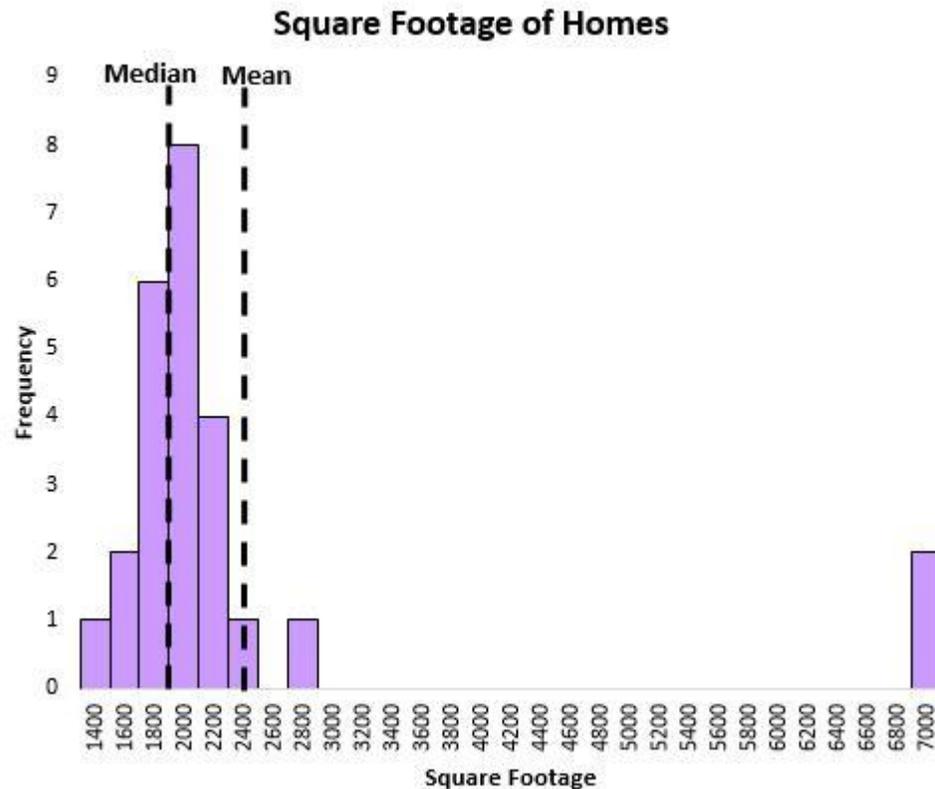
סולמות מדידה

- שמי - הערכים הם שמות של קטגוריות
 - דוגמה: מגדר, עיר
- סדר - יש משמעות לסדר של המספרים, אך אין משמעות למרווחים
 - דוגמה: דרגות בצבא, השכלה
- רווח - יש משמעות למרווחים, אך אין משמעות ליחסים
 - דוגמה: טמפרטורה במעלות Celsius
- מנה - יש משמעות ליחסים
 - דוגמה: משקל, טמפרטורה במעלות Kelvin

פירוט ודוגמאות נוספות [כאן](#)

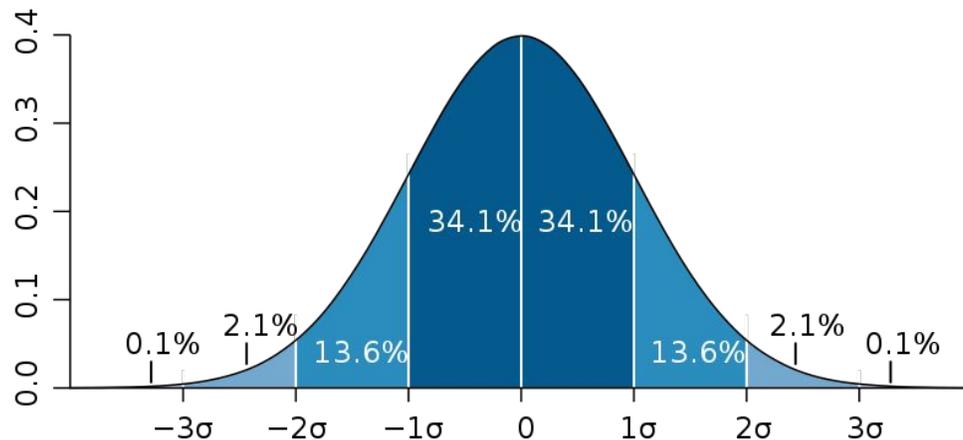
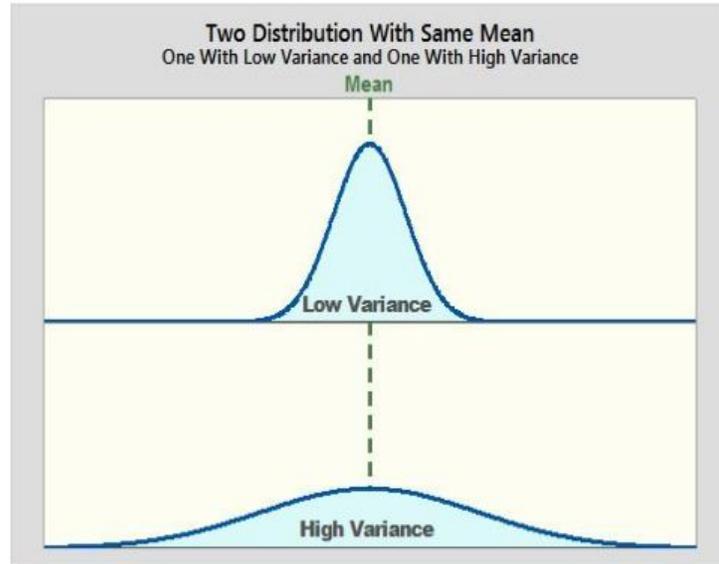
מדדי מרכז

- ממוצע (mean)
- חציון (median)
- שכיח (mode)



התפלגות נורמלית ומדדי פיזור

- שונות (variance)
- סטיית תקן (standard deviation)





הטכניון
מכון טכנולוגי
לישראל

ויזואליזציה עם ספריות Seaborn

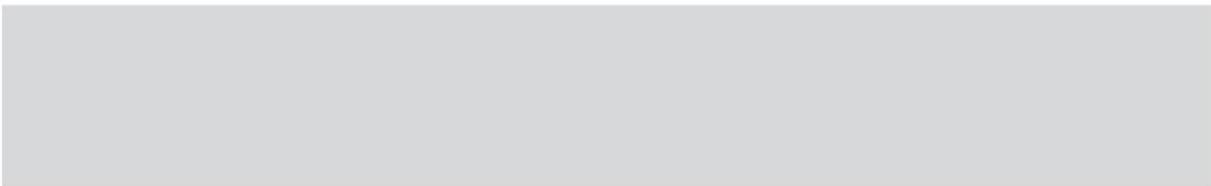
בשיעור זה נלמד:

- כיצד להמחיש נתונים על ידי גרפים
- כיצד להציג גרפים באמצעות הספרייה seaborn
- כיצד לנתח גרפים ולהסיק מהם מסקנות

תזכורת – תהליך העבודה במדעי הנתונים



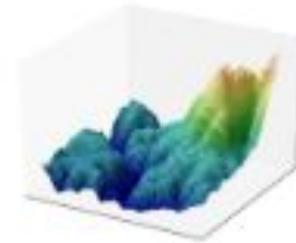
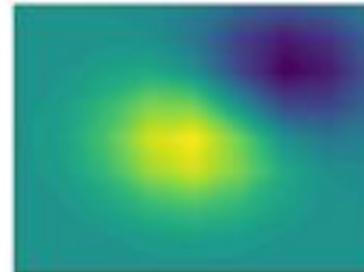
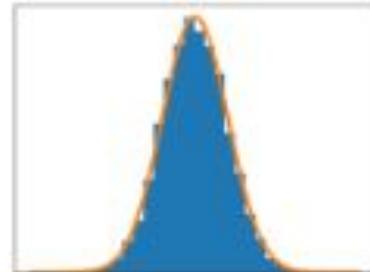
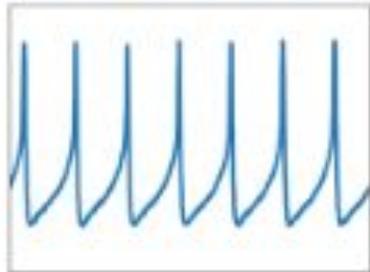
ספריית Seaborn
ויזואליזציה



ספריית matplotlib

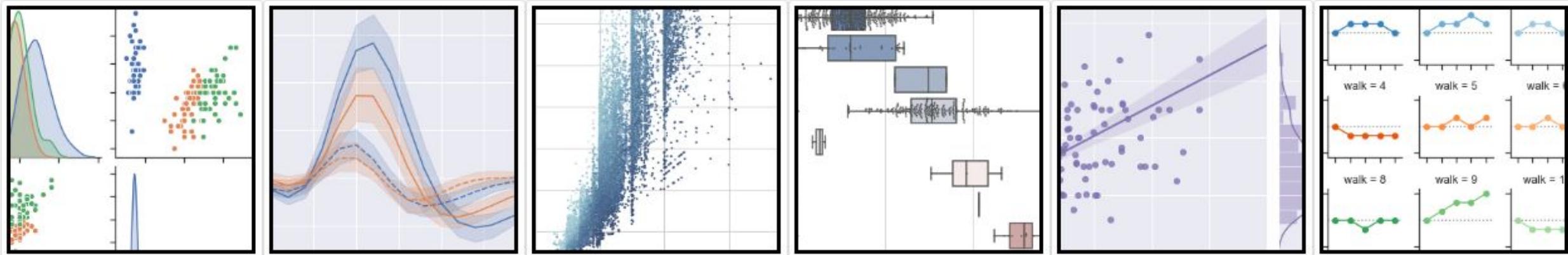
- ספרייה גרפית לשפת Python

matplotlib
Version 3.1.2



ספריית seaborn

- ספרייה גרפית ממוקדת בסטטיסטיקה
- מבוססת על matplotlib



ייבוא הספריות

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

סוגי הגרפים שנלמד

משתנים רציפים

- Pairplot

הערה: ניתן להוסיף לגרפים משתנים בדידים נוספים.
למעט Pie plot.

זוג משתנים

- רציף ובדיד

- Bar plot

- Hist plot

- KDE plot

- Box plot

- Violin plot

- שני רציפים

- Scatter plot

משתנה יחיד

- רציף

- Bar plot

- Hist plot

- KDE plot

- Box plot

- Violin plot

- בדיד

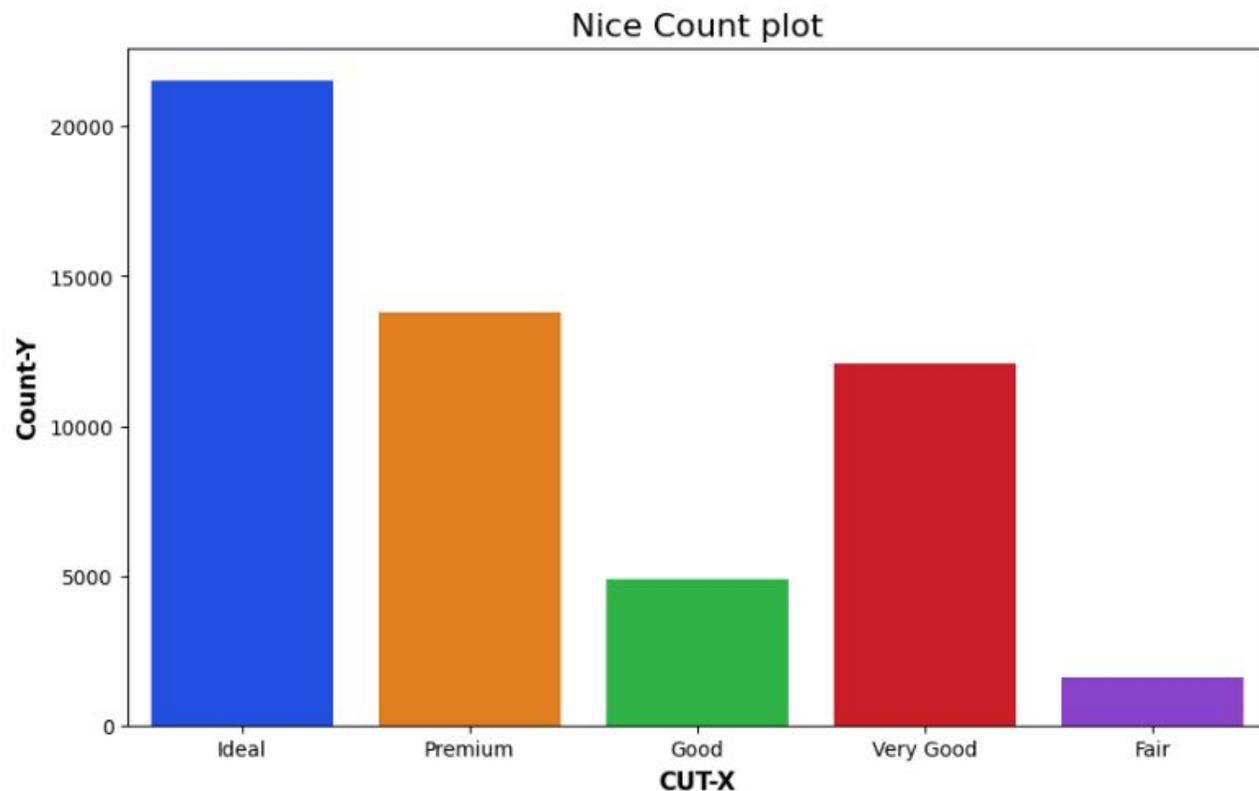
- Count plot

- Pie plot



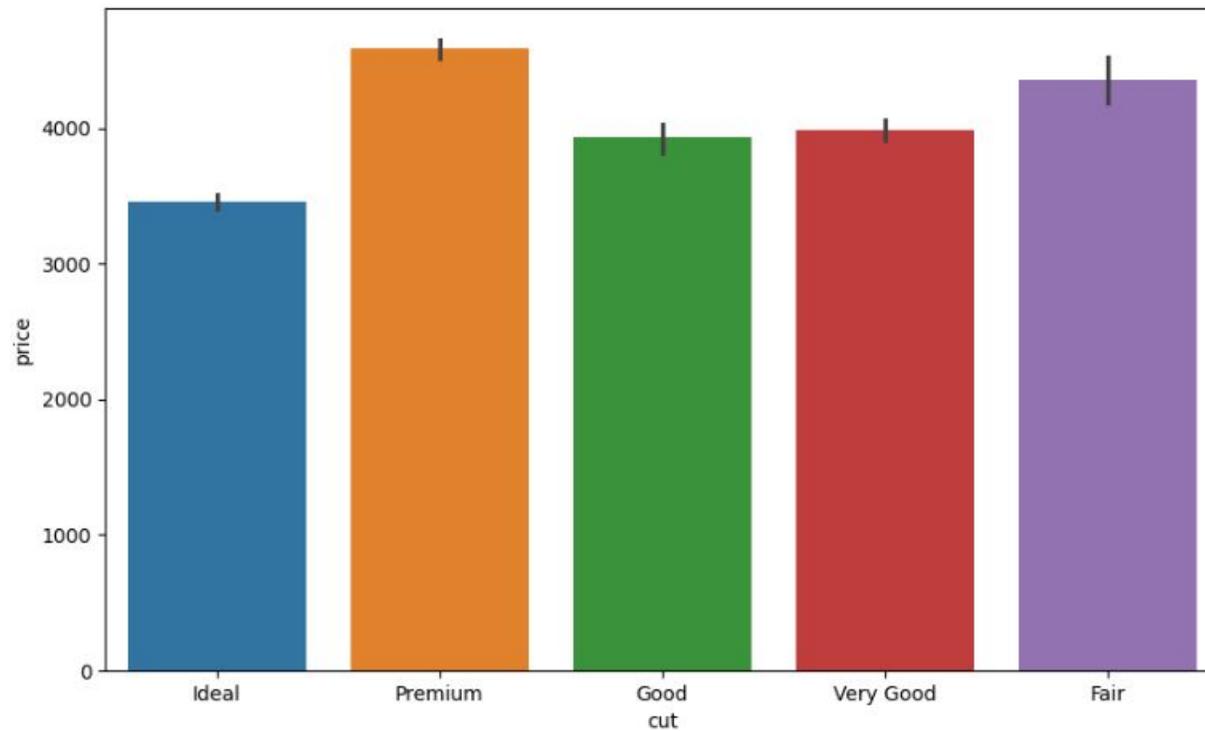
משתנה יחיד – בדיד - Count plot

```
plt.figure(figsize = (10,6))
sns.countplot(x='cut',hue='cut',data=df, palette='bright');
plt.xlabel("CUT-X", fontsize = 12, fontweight='bold')
plt.ylabel("Count-Y", fontsize = 12,fontweight='bold')
plt.title("Nice Count plot",fontsize = 16);
```



שני משתנים רציף + בדיד – Bar plot

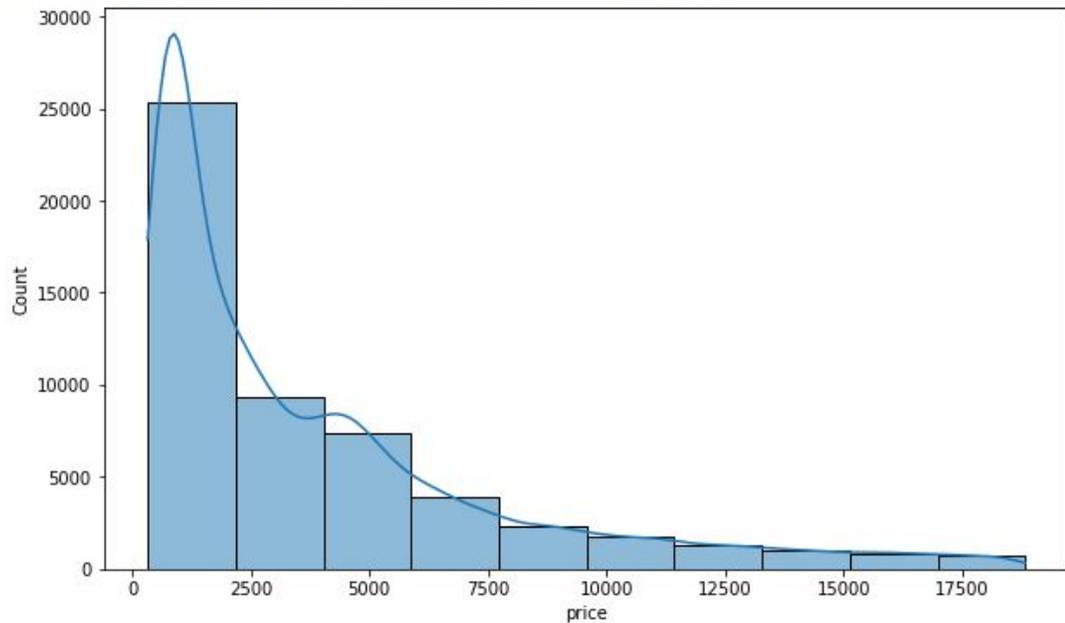
```
plt.figure(figsize = (10,6))  
sns.barplot(x='cut', hue='cut', y='price', data=df,palette='tab10');
```



משתנה יחיד – רציף – Hist plot

```
[7] plt.figure(figsize = (10,6))  
sns.histplot(data=df, x='price', bins=10,kde=True)
```

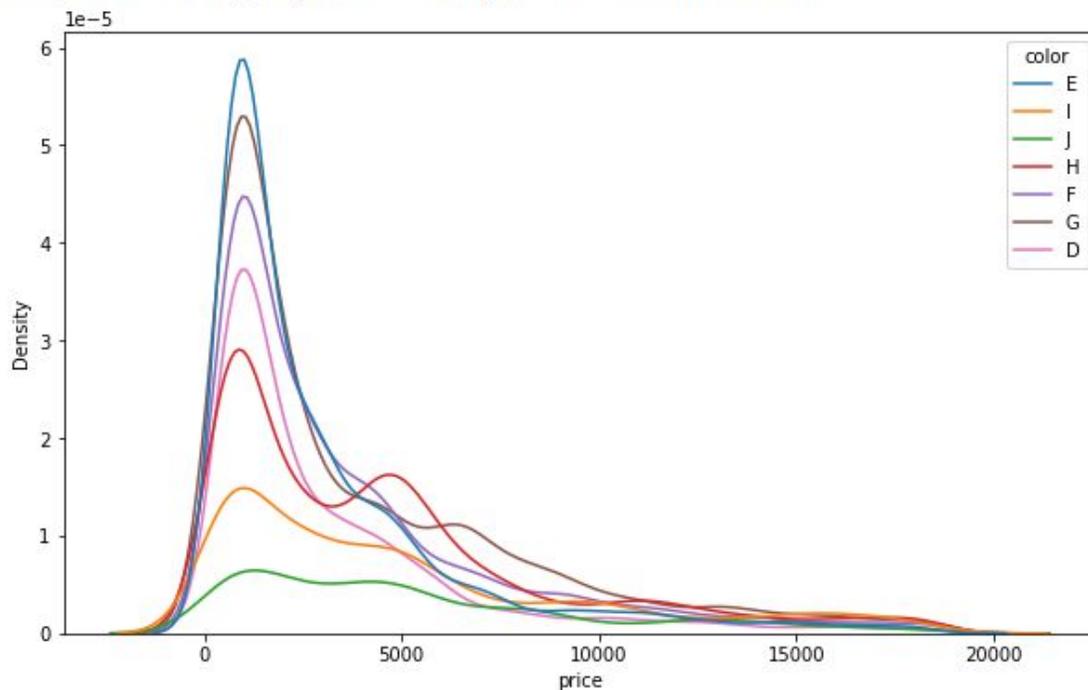
<matplotlib.axes._subplots.AxesSubplot at 0x7fad8e9578d0>



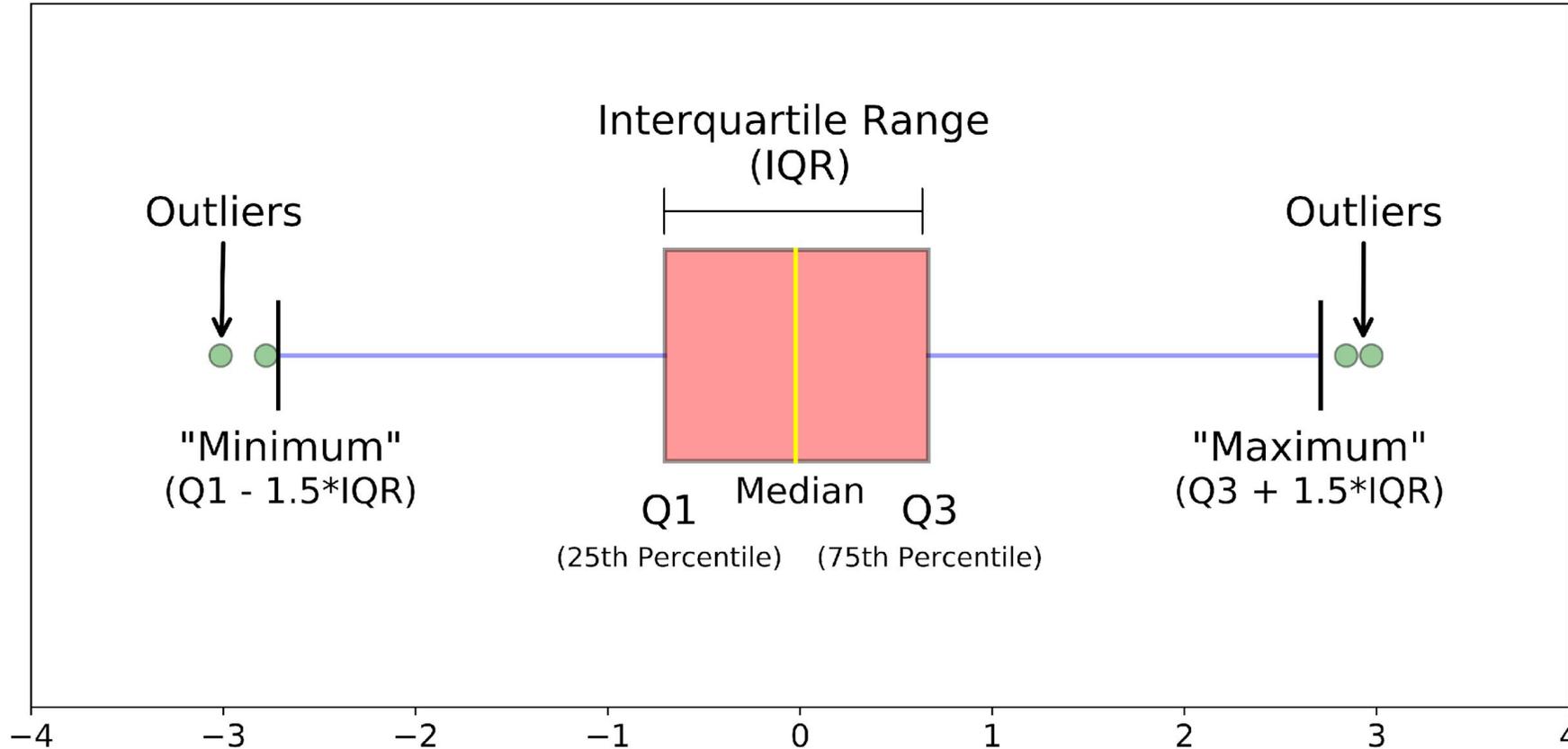
שני משתנים רציף + בדיד – KDE plot

```
[16] plt.figure(figsize = (10,6))  
sns.kdeplot(data=df, x='price',hue='color')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fad7e0fbbd0>

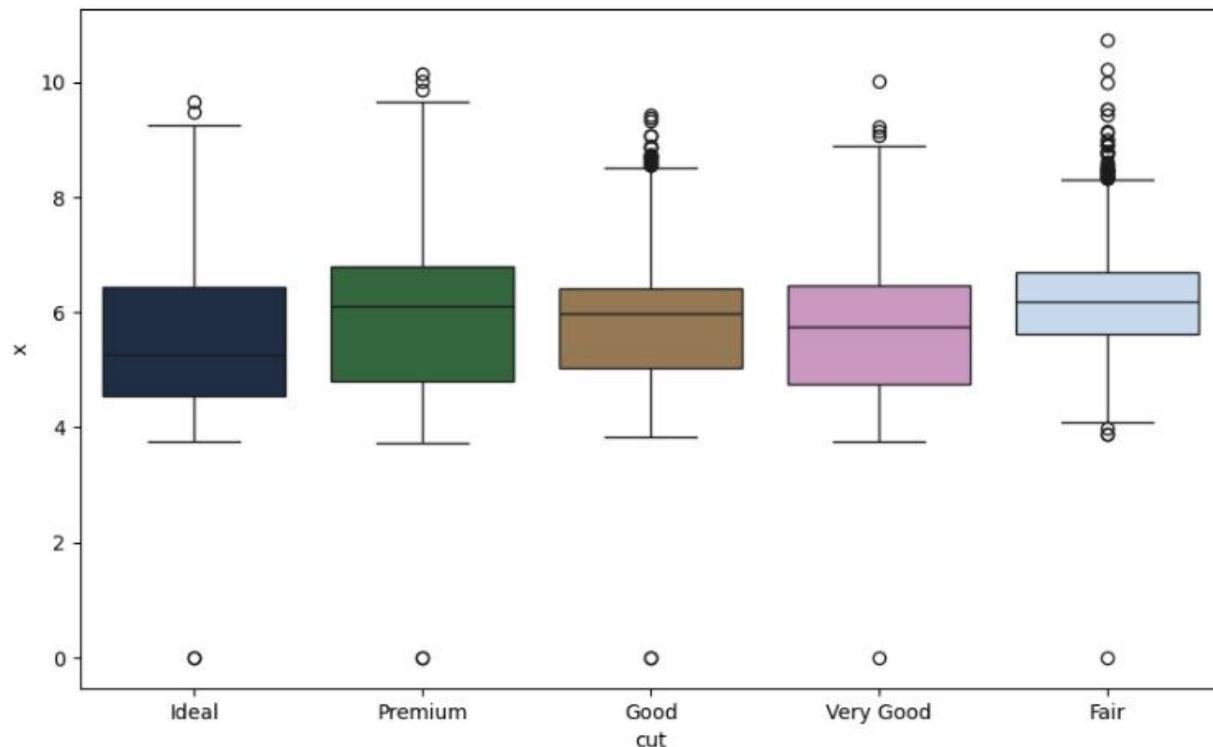


Understanding Box plot



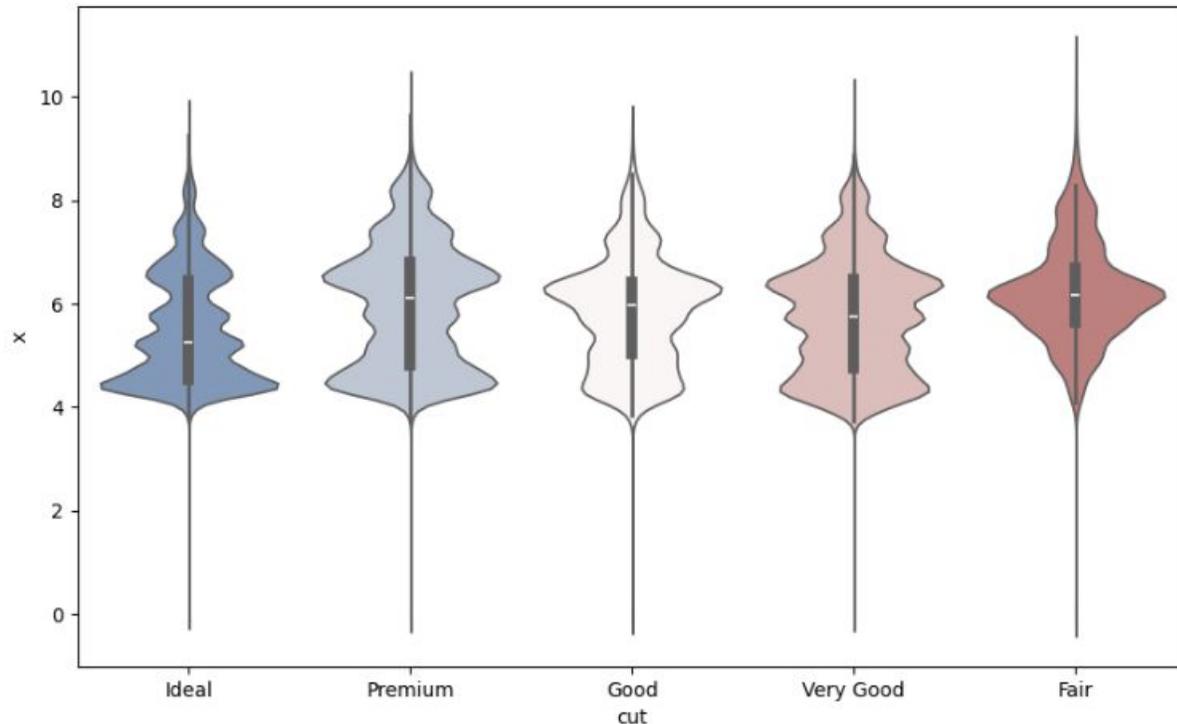
שני משתנים רציף + בדיד – Box plot

```
plt.figure(figsize = (10,6))  
sns.boxplot(x='cut',hue='cut', y='x',data=df,palette='cubehelix');
```



שני משתנים רציף + בדיד – Violin plot

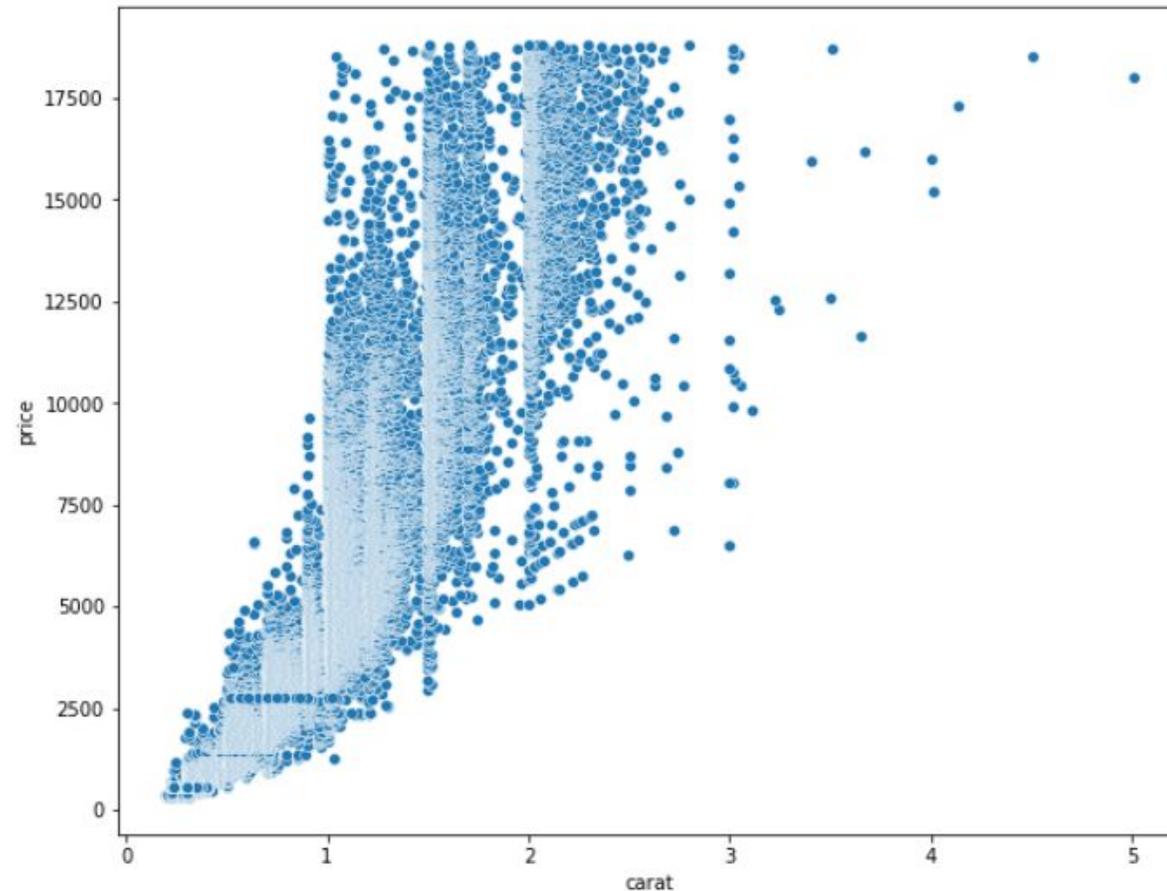
```
plt.figure(figsize = (10,6))  
sns.violinplot(x='cut',y='x',data=df,hue='cut',palette='vlag');
```



**Violinplot =
boxplot + kdeplot**

שני משתנים רציפים - Scatter plot

```
✓ 1s ▶ plt.figure(figsize = (10,8))  
sns.scatterplot(x = "carat", y = "price", data = df);
```

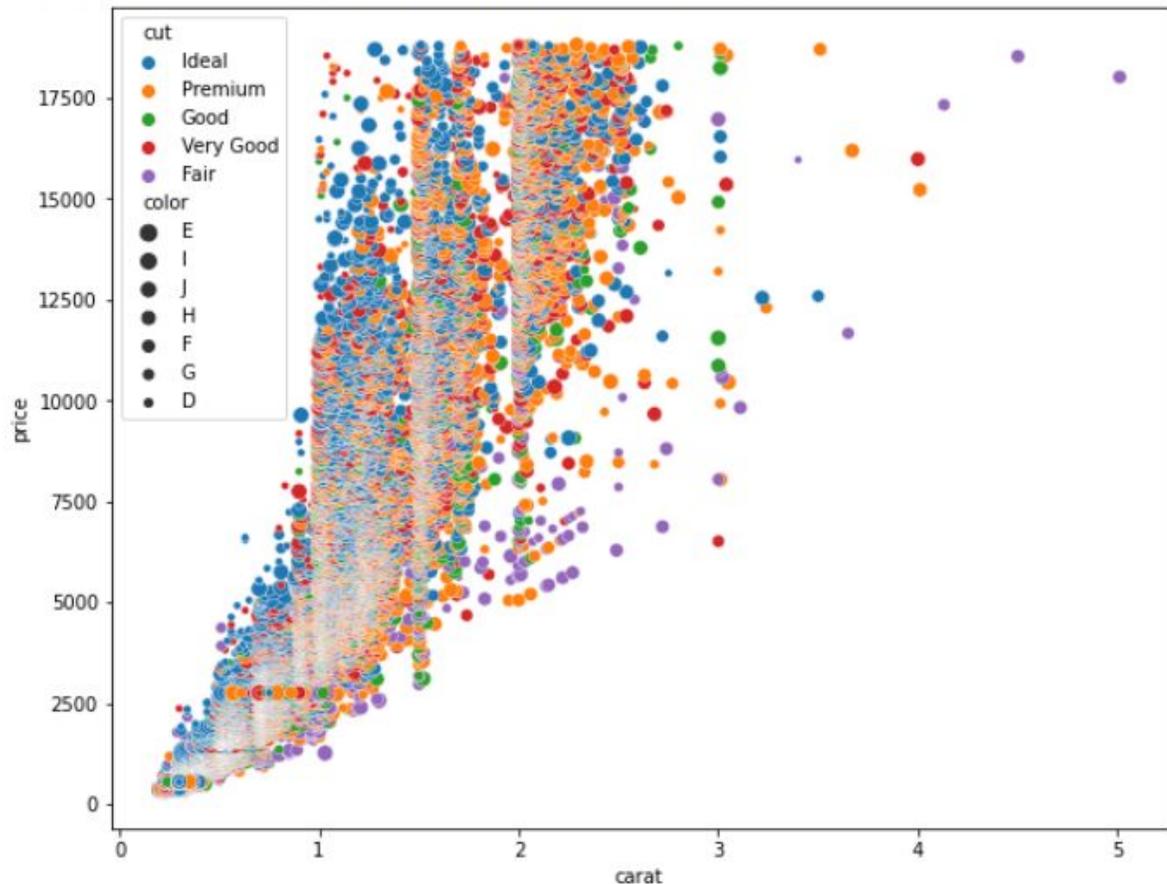


שני משתנים רציפים ושני משתנים בדידים

Scatter plot -

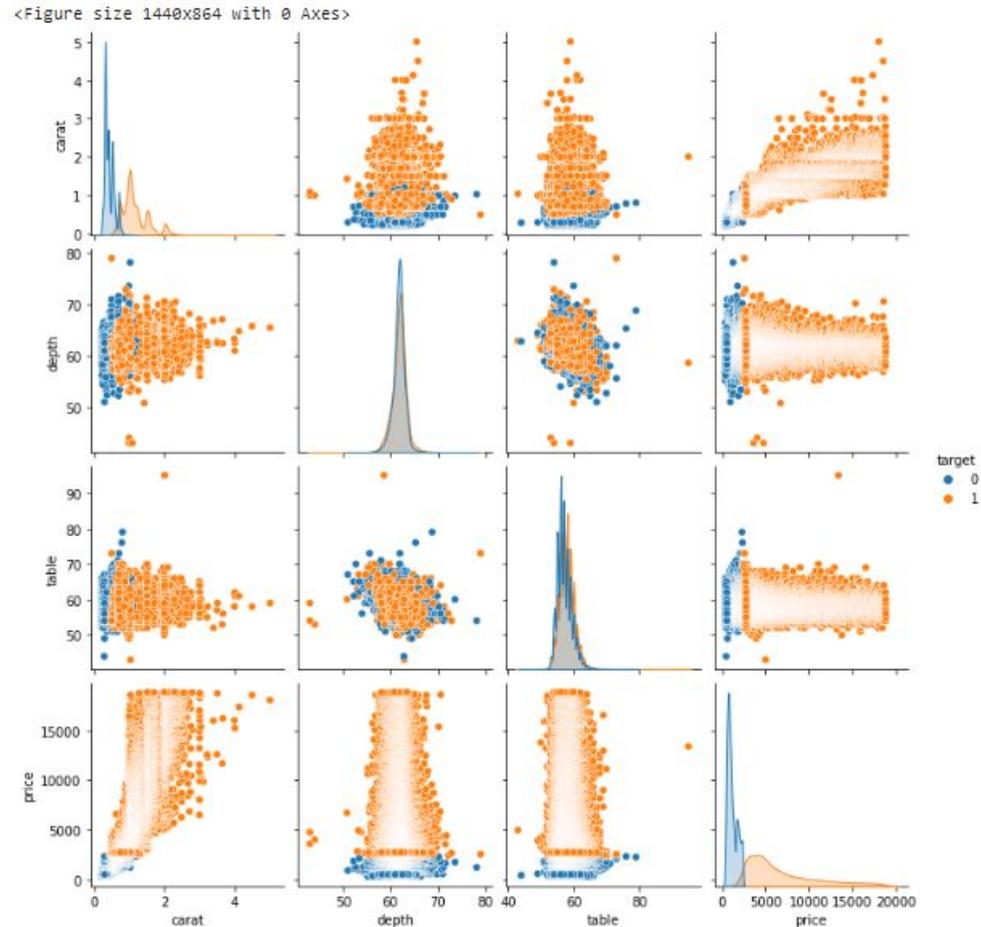
```
plt.figure(figsize = (10,8))  
sns.scatterplot(x = "carat", y = "price", data = df, hue = "cut",size='color')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fb02bf2f390>



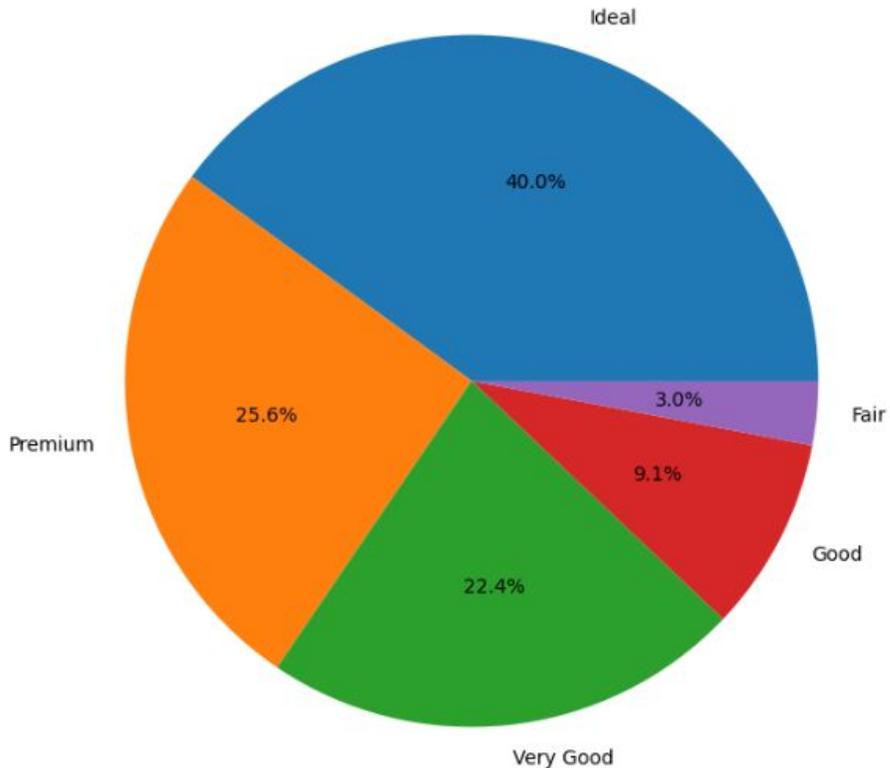
Scatter של זוגות משתנים - pairplot

```
[20] plt.figure(figsize = (20,12))  
sns.pairplot(data=df[['carat', 'depth', 'table', 'price', 'target']], hue='target');
```



משתנה בדיד – Pie plot

```
plt.figure(figsize = (8,8))
cut_counts=df['cut'].value_counts()
labels=cut_counts.index
sizes=cut_counts.values
plt.pie(sizes,labels=labels, autopct='%1.1f%%',startangle=0);
```



בחירת Palletes

- קישור Seaborn palletes כאן
- קישור Matplotlib palletes כאן

תרגיל 3.1

- טענו מחברת טיטאניק
- הוסיפו 5 גרפים למחברת
- שנו מידות של שני גרפים
- שנו את שמות הצירים והכותרת של שני הגרפים
- שנו את סוג ה Pallette של שני גרפים

הדגמות והעמקות על מחברת Colab

- קישור למחברת Diamonds Visualization [כאן](#)
- קישור Seaborn API reference [כאן](#)

תרגיל 3.2 בקבוצות

• נמשיך את המחברת של תרגיל 3.1

את ההערות והמסקנות יש לכתוב בתא טקסט.

כדי לרשום בתא טקסט בעברית מומלץ להוסיף בתוך תא הטקסט: `div dir =>`
`<rtl`

כדי ליישר את הכיתוב לימין.

• רשמו לכל גרף את סוגי המשתנים שמופיעים בו (בדיד, רציף)

• רשמו מסקנה אחת לכל גרף

3.3 תרגיל ויזואליזציה כולל ניתוח מקיף

טענו את [Penguin Dataset](#)

את ההערות והמסקנות יש לכתוב בתא טקסט.
כדי לרשום בתא טקסט בעברית מומלץ להוסיף בתוך תא הטקסט: `<div dir = rtl>`
כדי ליישר את הכיתוב לימין.

- הוסיפו פקודת Pandas Info ורשמו מסקנה מהפקודה.
- אופציונלי: מחקו רשומות חסרות עם `df.dropna(inplace=True)` לאחר טעינת הנתונים. זה יכול לפתור בעיות מסוימות עם גרפים.
- הוסיפו פקודת Pandas Describe ורשמו מסקנה מהפקודה.
- הוסיפו 3 פקודות Pandas Groupby שונות ורשמו מסקנה אחת מכל פקודה.
- הוסיפו 5 גרפים שונים למחברת
- רשמו לכל גרף את שמות וסוגי המשתנים שמופיעים בו (בדיד, רציף)
- רשמו מסקנה אחת לכל גרף



הטכניון
מכון טכנולוגי
לישראל

תודה על ההשתתפות