ויזואליזציה

UNIT 4

Histograms– מציג התפלגות של משתנה רציף בין טווח ספציפי.

Boxplot **-** מציג פיזור של משתנים הנותן אפשרות להשוואה עמוקה יותר.

|  |  |  |
| --- | --- | --- |
| שם הפונקציה | תצוגה | הסבר |
| .plot() |  | נשתמש נרצה לראות את ההתפלגות בין טווחים.  הקוד:  Data.plot() |
| .hist() |  | מציג התפלגות של משתנה רציף בין טווח ספציפי.  "התפלגות x לפי כמות"  הקוד:  ym.hist('height\_f') |
| gym.plot.hist(bins=20, alpha=0.7) |  | הקוד:  gym.plot.hist(bins=20, alpha=0.7) |
| .boxplot |  | מציג פיזור של משתנים הנותן אפשרות להשוואה עמוקה יותר .  בבוקס נראה: ערך מינ'+מקס', רבעון ראשון ורבעון 3, חציון.  הקוד:  boxplot = df.boxplot(column=['C1', 'C2', 'C3'])  או grouped\_df.boxplot('people\_fully\_vaccinated\_per\_hundred') |
| sns.boxplot |  | הקוד:  plt.figure(figsize=(4,4))  sns.boxplot(x='Dependents', y='TotalMnt', data=mrkt\_df); |

UNIT 5

Correlation– מתאים כאשר רוצים למצוא קשר בין 2 משתנים- קורלציה. כאשר אני רוצה להראות חריגים ולהבין את הדאטה.

|  |  |  |
| --- | --- | --- |
| שם הפונקציה | תצוגה | הסבר |
| .plot.scatter |  | מראה פיזור ולא מדגיש  הקוד:  house\_df.plot.scatter(x = 'median\_house\_value', y = 'median\_income') |
| .corr(method='') |  | יוצר טבלה המציגה את הקורלציות  המתודות:  method='pearson' – בערכים רציפים בעלי התפלגות נורמלית  method='spearman' – ערכים מדורגים או לא מפולגים נורמלית  method='kendall' – עדיף כשיש מדגם קטן של ערכים  הקוד:  house\_df[['median\_income', 'median\_house\_value']].corr(method='pearson') |
| sns.scatterplot |  | הקוד:  sns.scatterplot(data=house\_df, x='total\_bedrooms', y='households') |
| sns.regplot |  | מוסיף קו לניארי  הקוד:  sns.regplot(data=house\_df, x='total\_bedrooms', y='households') |
| sns.pairplot |  | יוצר גרפים לפי פיצ'רים.  הקוד:  features = ['median\_house\_value', 'housing\_median\_age',  'median\_income']  sns.pairplot(house\_df[features], height = 2.5) |
| sns.heatmap |  | נגדיר את המשתנים שנרצה לעשות עליהם קורלציה. ניצור את הטבלה ולאחר מכן נפעיל את המפת חום.  הקוד:  features = ['median\_house\_value', 'housing\_median\_age','median\_income','total\_bedrooms','population']  correlation\_matrix = house\_df[features].corr().round(2)  sns.heatmap(data=correlation\_matrix,cmap='Greens', annot=True) |

Unit 6

|  |  |  |
| --- | --- | --- |
| שם הפונקציה | תצוגה | הסבר |
| sns.regplot |  | הקוד:  sns.regplot(x='Income', y='TotalMnt', data=mrkt\_df); |
| הצגת שתי קורלציות במקביל |  | חשוב לשים את האימפורט  הקוד: from matplotlib import pyplot as plt  fig, xx =plt.subplots(1,2, figsize=(10,5))  sns.scatterplot(ax=xx[0],data=YouTube, x='likes', y='views')  sns.scatterplot(ax=xx[1],data=YouTube, x='likes', y='comment\_count') |
| הצגת קורלציה בין שתי משתנים לפי עמודה מסוימת  מתוך יוניט 9 |  | למשל קורלציה בין אורך למקור לעומק מקור לפי סוג החיה.  קוד:  ax = sns.scatterplot(data=penguins, x="bill\_length\_mm", y="bill\_depth\_mm", hue = 'species', palette = 'colorblind')  קוד 2:  ax = sns.scatterplot(data=penguins, x="bill\_length\_mm", y="bill\_depth\_mm", hue = 'species', palette = 'colorblind', legend=False)  style = dict(size=12, color='black')  ax.text(35, 15, "Adelie", \*\*style)  ax.text(55, 20, "Chinstrap", \*\*style)  ax.text(52, 14, "Gentoo", \*\*style)  plt.show() |
| .plot.pie |  | הקוד:  plt.figure(1, figsize=(5,5))  plt.title("Distribution of Species")  iris\_df['class'].value\_counts().plot.pie(autopct="%1.1f%%")  plt.show() |

Unit 9

|  |  |  |
| --- | --- | --- |
| שם הפונקציה | תצוגה | הסבר |
| pivot\_table | כאשר מבקשים להציג כמות של משהו ביחס לשני פרמטרים. כמו מכירות ביחס לשנה ולסוג, כמות צפיות לפי קטגוריה בכל שנה. | הברירת מחדל – ממוצע  ניתן לשנות את הברירת מחדל כך: aggfunc={'views':sum}  הקוד:  data.pivot\_table(['math score','reading score','writing score' ],'gender') |
| sns.countplot |  | הקוד:  sns.countplot(x="parental level of education", data=data) |
| sns.barplot |  | הקוד:  sns.barplot(x="sex", y="survived", hue="class", data=titanic) |
| גרף שמבקשים עם לוגים – כמו מטלה 5. |  | הקוד:  data['total\_sold'].plot.hist('total\_sold',bins=8)  plt.yscale('log')  plt.show() |