

פרויקט – Data Analysis עם Python בסיסי

עליכם לייבא מידע ממאגר נתונים כלשהו לתוך Pandas (DataFrame להלן נקרא לו DF), ולנתח את הנתונים הקיימים בו.

שאלות מנחות:

- לתאר את ה-DF שייבאתם, ולהסביר מה ברצונכם לנתח (כל פקודה שהנכם מוסיפים לפרויקט הוסיפו שורת תיאור(מטרה) עבור הפקודה)
- אם יש צורך לסנן או לשנות נתונים ריקים אשר משפיעים על המדדים שלכם, הגדירו פקודות שמבצעות זאת.
- הציגו מידע על ה-DF
- הציגו סטטיסטיקות כלליות על ה-DF
- הציגו ערכים גבוהים, נמוכים
- מיינו את ה-DF עפ"י נתונים שונים
- קבצו נתונים לפי קטגוריות שונות וחשבו מדדים: ממוצע, סכום, מינימום מקסימום, חציון, סטיות תקן ועוד..
- המירו תאריכים במידת הצורך
- קבצו נתונים על פי נתוני זמן (שנה, חודש, רבעון, יום בשבוע ועוד..)
- צרו ויזואליזציות הממחישות את הניתוחים שלכם כגון: line, histogram, scatter, barh, bar, Pie, heatmap, catplot - השתמשו לפחות בארבעה מסוגים השונים של גרפים ה"ל
- הוסיפו לתרשימים שלכם במידת הצורך מקרא, כותרות ועיצובים.

ניתן להיעזר בדוגמה הר"מ

מבוא

התוכנית "יהלום ממבט ראשון" רוצה להבין את Datan יהלומים שלהם בצורה יותר טובה, והיא צריכה את העזרה שלכם בלהפגיש בין אדם ליהלום שלו בצורה חכמה יותר. עד היום שום יהלום ומתמודד לא הצליחו להישאר גם אחרי התוכנית ביחד, האם תוכלו לעזור ליהלום ממבט ראשון לשפר את סיכויי ההצלחה?

עליכם להוריד את הקובץ [הנ"ל](#) למחשב שלכם, ולשים אותו באותה תיקייה שבה אתם עובדים (או להשתמש בPath אבסולוטי).

קריאת Datan

פתחו את הקובץ Data, ונסו להבין איך כדאי לכם לקרוא אותו. הוא מכיל מידע שאנשי Datan (הפחות מוצלחים) של התוכנית הצליחו לגבש.

קראו את השאלות שעליכם לענות עליהן ונסו להבין מה תצטרכו לבצע. רמז: הפונקציות `split` ו `readlines` יכולות לעזור.

שאלות

אנא בצעו את השליפות הבאות על קובץ הנתונים:

1. מה מחיר היהלום הגבוהה ביותר?

2. מה המחיר הממוצע של יהלום?
3. כמה יהלומים מסוג Ideal קיימים?
4. כמה צבעים שונים יש ליהלומים? מהם?
5. מה החציון קאראט של יהלומים מסוג Premium?
6. צרו ממוצע carat לכל סוג cut.
7. צרו ממוצע מחיר לכל סוג צבע.