

**מטלה 3**

מטרת המטלה – שיפור והבנה מעמיקה של numpy , pandas .

את כלל המטלה עליכם להגיש בקובץ זיפ ששמו יהיה תעודת הזהות של המגיש. קובץ הזיפ יכיל תיקייה בשם Ex, התיקייה תכיל שתי מחברות ג'ופיטר בלבד- האחת נקראת boring והשניה נקראת kickstarter , אין צורך להגיש את קבצי excel\csv המצורפים למטלה זו. מטלה זו תחושב עם בונוס על שאלת יצירתיות במחברת kickstarter , כמו כן, המחברות איתן תרגלנו מכינות בצורה מספקת למטלה, אך מומלץ ביותר להשתמש בדוקומנטציה של pandas ו numpy לענות בצורה המיטבית, כמו גם למצוא מחברות לדוגמה ברחבי האינטרנט. כולי תקווה שתהנו מהמטלה ובפרט מהמחברת קיקסטארטר. עבור הסעיפים המודגשים, הציון יהיה גבוה יותר מאלו שלא.

**מחברת boring:**

עליכם לבצע כל סעיף בתא נפרד-

1. יש לייבא את numpy .
2. צרו list בגודל 30 שיכיל את המספרים מ 1(!) עד 30 מסודרים מהקטן לגדול.
3. המירו ((cast את הרשימה למערך של numpy.
4. צרו מטריצה מהמערך מצורה 5\*6 כאשר 6 הוא מספר השורות, 5 מספר העמודות.
5. **חשבו את סכום האלכסון הראשי של המטריצה באמצעות פקודות פייתון (לולאות).**
6. **חשבו את סכום השורה השניה.**
7. צרו מטריצה נוספת מצורה 5\*6 שכל ערך במטריצה יהיה שווה ל-1.
8. חברו את שתי המטריצות הנ"ל.
9. צרו מטריצה 30\*30 עם ערכים רנדומליים מסוג integer .
10. **הדפיסו את האיבר המינימלי ואת האינדקס של האיבר המינימלי בכל אחת מ30 השורות.**
11. **הציגו את תת המטריצה בגודל 10\*10 המופיעה בדיוק באמצע המטריצה.**
12. **חשבו למטריצה שמסעיף 10 את הממוצע, החציון ואת סטיית התקן והדפיסו אותם על המסך.**
13. צרו את האינטרוול (0,20] וחלקו אותו למקטעים בגודל 1.6666666. הדפיסו את תתי הקטעים למסך.

**מחברת kickstarter:** (סטודנטים שאינם מכירים את הפלטפורמה קיקסטארטר מתבקשים לקרוא על החברה ועל המוצר אותו היא מציעה)

**שלחתי לכם בוואטספ את הדאטה פריים. אם ישנם סטודנטים שלא מסתדרים עם הקובץ משום מה, עליכם להכנס לאתר Kaggle.com ולחפש את הדאטה פריים של קיקסטארטר, להוריד אותו ולעבוד איתו. שם הקובץ הוא : ks-projects-201801.csv**

רקע לתרגיל: לתרגיל זה מצורף קובץ excel בפורמט csv. הטבלה המצורפת היא טבלה אמיתית שפלטפורמת קיקסטארטר הפיצה בכדי שמדעני נתונים כמונו ינסו לחזות איזה פרוייקט קיקסטארטר יצליח לגייס את הסכום המבוקש ואיזה לא, כמו גם שימצאו תובנות מעניינות על הדאטה.

עליכם לבצע כל סעיף בתא נפרד:

1. יש לייבא את pandas.
2. צרו DataFrame מהקובץ המצורף וקראו לו df.
3. הציגו את 10 השורות הראשונות ב df
4. הציגו את 5 השורות האחרונות בdf
5. **הציגו את 5 הערכים הראשונים של עמודת הcategory בלבד.**
6. **רשמו כהערה פירוט על כל אחד מהפיצ'רים בדאטה- . (תזכורת, פיצ'ר או שם עמודה, הוא בעצם התא המודגש המייצג את העמודות. רשמו מה מייצגת כל עמודה)**
7. **רשמו כהערה על כל אחד מהפיצ'רים איזה סוג דאטה הוא מייצג (אם ניתן כמובן) ( types of data) יש להקפיד על סדר בתא זה על אחת כמה וכמה.**
8. **הציגו את צורת הדאטה (מספר שורות ומספר עמודות) בצורת tuple**
9. בדקו כמה ערכי NaN מופיעים בדאטה.
10. **הציגו בסמוך לכל פיצ'ר כמה ערכי NaN יש בעמודה הספציפית.**
11. צרו series מהעמודה ID והציגו את חמשת הערכים הראשונים שבה.
12. מחקו את העמודה ID מ df לגמרי. (נראה שהפיצ'ר הזה לא תורם להבנת הדאטה בצורה טובה, אין טעם להשאיר אותו)
13. הריצו את הפקודה df.info() להתרשם קצת מיכולות pandas
14. הריצו את הפקודה df.describe()
15. אתגר : לאחר שאתם מבינים מה הוא הפיצ'ר state , הציגו כמה פרוייקטים של קיקסטארטר הצליחו וכמה נכשלו. הסבירו מדוע סכומם קטן מסך כל הרשומות (השורות)
16. **בונוס:** הוסיפו תאים כאוות נפשכם, למדו את df ככל שתוכלו, ונסו למצוא תובנות מעניינות רק באמצעות pandas. לדוגמה: " בשנת 2014 יותר פרוייקטים הצליחו מאשר ב2015" דוגמה נוספת: "סך כל הסכום שגויס בקטגוריית FOOD הוא 2,500,352$" וכן הלאה. תובנות אלו חייבות להיות מגובות בקוד. **ממליץ במיוחד לבצע את הסעיף הזה**.

**בהצלחה.**