**מטלה 3**

**Pandas**

מטרת המטלה – תרגול pandas מתקדם.

את כלל המטלה עליכם להגיש בקובץ זיפ ששמו יהיה תעודת הזהות של המגיש. קובץ הזיפ יכיל מחברת ג'ופיטר אחת בשם pandas ומחברת נוספת בשם visualization בהתאם לשני חלקי המטלה , אין צורך להגיש את קבצי excel\csv המצורפים למטלה זו.

**כל סעיף בתא נפרד כאשר בין השאלות מפריד תא ריק 1 בלבד, בכל תא דוקומנטציה מסודרת והסבר על איזה שאלה אתם עונים (בצורת הערה)**

שאלה 1

נתון קובץ drinks.csv המרכז את צריכת האלכוהול במדינות שונות בעולם.

העלו את הקובץ והשתמשו בפונקציות של pandas בשביל לענות על השאלות הבאות:

1. מהם ממוצעי שתית הבירה בכל יבשת?
2. הציגו את הסטטיסטיקות של צריכת יין בכל יבשת (כמות, ממוצע, סטיית תקן, כמות מינימלי, כמות מקסימלית)
3. הציגו את ממוצע צריכת האלכוהול מכל סוג בכל יבשת
4. הציגו את החציון של צריכת אלכוהול מכל סוג בכל יבשת
5. הציגו את הממוצע וערכי המינימום והמקסימום עבור צריכת הכוהל

שאלה 2

נתונים הקבצים car1.csv, car2.csv הנותנים נתוני רכבים.

העלו את הקבצים והשתמשו בפונקציות של pandas בשביל לענות על השאלות הבאות:

1. הורידו את העמודות המיותרות מ-car1.csv
2. כמה תכונות (פיצ'רים) וכמה רשומות יש בכל קובץ נתונים.
3. צרפו את 2 קבצי הנתונים ל-DataFrame אחד.
4. הוסיפו עמודה owners לטבלה והכניסו בה מספרים רנדומליים בטווח של 15000-73000

שאלה 3

1. צרו 3 DataFrames המבוססים על הנתונים הבאים-

raw\_data\_1 = {

'subject\_id': ['1', '2', '3', '4', '5'],

'first\_name': ['Alex', 'Amy', 'Allen', 'Alice', 'Ayoung'],

'last\_name': ['Anderson', 'Ackerman', 'Ali', 'Aoni', 'Atiches']}

raw\_data\_2 = {

'subject\_id': ['4', '5', '6', '7', '8'],

'first\_name': ['Billy', 'Brian', 'Bran', 'Bryce', 'Betty'],

'last\_name': ['Bonder', 'Black', 'Balwner', 'Brice', 'Btisan']}

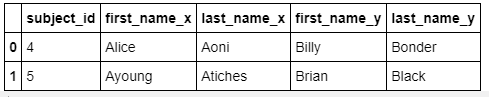
raw\_data\_3 = {

'subject\_id': ['1', '2', '3', '4', '5', '7', '8', '9', '10', '11'],

'test\_id': [51, 15, 15, 61, 16, 14, 15, 1, 61, 16]}

1. צרפו את 2 הטבלאות הראשונות לאורך (לחבר רשומות) והכניסו לטבלה חדשה.
2. מזגו את הטבלה החדשה עם הטבלה השלישית (subject\_id יחבר בינהם)
3. מזגו את הטבלאות הנ"ל, רק כאשר את ה-subject\_id ב-2 הטבלאות

זו התוצאה שאתם צריכים לקבל-



1. מזגו את הטבלאות הנ"ל, כאשר רשמות תואמות משני הצדדים

זו התוצאה שאתם צריכים לקבל-



שאלה 4 שימוש בgroupby

1. נתון df בגודל 4\*13 כאשר שמות הפיצ'רים הם : . animal, age , weight ,length צרו את הdf

עבור animal מלאו את df בחיות הבאות ולפי הסדר:

Dog (באינדקס 0)

Cat

Snake

Snake

Dog

Hamster

Cat

Alligator

Cat

Cat

Dog

Hamster

Alligator (באינדקס 13)

השלימו את שאר הdf כראות עיניכם אך בצורה הגיונית (למשל חתול שוקל בערך 7 ק"ג) שימו לב שכל ערך בטבלה חייב להיות ייחודי (אין שני חתולים למשל ששוקלים אותו הדבר).

1. סדרו מחדש את הערכים הייחודים מהפיצ'ר animal לתוך דאטהפריים שנקרא animal\_groups
2. הציגו את ממוצע המשקל של החתולים
3. הציגו את הגיל המקסימלי של הכלבים.
4. הציגו עבור כל סוג של בע"ח כמה פרטים ממנו יש ב animal\_group
5. הציגו את החציון של האורך של בע"ח.

בהצלחה.