

תרגיל בית 5

הנחיות כלליות

- קראו **היטב** את השאלות והקפידו שהתכניות שלכם פועלות בהתאם לנדרש.
- את התרגיל יש לפתור לבד!
- הקפידו על כללי ההגשה המפורסמים באתר. בפרט, יש להגיש את כל השאלות יחד בקובץ `ex5_012345678.py` המצורף לתרגיל, לאחר החלפת הספרות 012345678 במספר ת.ז שלכם, כל 9 הספרות כולל ספרת הביקורת.
- מועד אחרון להגשה: כמפורסם באתר.
- בדיקה עצמית: כדי לוודא את נכונותן ואת עמידותן של התוכניות לקלטים שגויים, בכל שאלה, הריצו את תוכניתכם עם מגוון קלטים שונים, אלה שהופיעו כדוגמאות בתרגיל וקלטים נוספים עליהם חשבתם (וודאו כי הפלט נכון וכי התוכנית אינה קורסת).
- היות ובדיקת התרגילים עשויה להיות אוטומטית, **יש להקפיד על פלטים מדויקים על פי הדוגמאות (עד לרמת הרווח)**.
- אופן ביצוע התרגיל: בתרגיל זה עליכם להשלים את הקוד בקובץ המצורף.
- **אין לשנות את שמות המשתנים שכבר מופיעים בקובץ השלד של התרגיל.**
יש לעבוד עם המשתנים שמופיעים בשלד התרגיל. על הקוד של כל שאלה לעבוד ולספק את התוצאה הדרושה עבור קלט שיוזן במשתנים שמופיעים בשלד (המשתנים שלידם סימני שאלה ומחכים לקלט כפי שראינו בדוגמא מהתרגול). יחד עם זאת, אתם רשאים להוסיף משתנים נוספים כראותם עינכם.
- **אין למחוק את ההערות שמופיעות בשלד.**

הנחיה כללית לתרגיל :

בקובץ התרגיל מופיע חלק של בדיקות המסתמכות על קבצים המצורפים לתרגיל. על מנת להריץ את הבדיקות הללו, קבצי הקלט לדוגמא צריכים להימצא באותה התיקיה בה נמצא קובץ ה py של התרגיל. לחילופין, אתם יכולים למקם את קבצי הקלט במקום אחר, אבל אז תצטרכו לעדכן את הבדיקות עם הנתבי המלא של קבצי הקלט.

שאלה 1- סכימת מספרים המופיעים בקובץ

ממשו את הפונקציה `sum_num(file_name)` המקבלת מחרוזת המציינת שם של קובץ קלט `(file_name)`. הניחו שבתוך הקובץ מופיעה שורה בודדת הכוללת סדרת מספרים שלמים וחיוביים המופרדים על ידי רווח בודד. על הפונקציה לקרוא את הקובץ ולהחזיר את סכום המספרים המופיעים בו.

לתרגיל מצורף הקובץ `q1_input_1.txt` שמכיל את התוכן הבא :

4 55 3 67 10

עבור הקובץ `q1_input_1.txt` הפונקציה תחזיר את הערך 139.

הערה : בשאלה זו ניתן להניח שהקלט תקין ואין צורך לטפל בשגיאות.

שאלה 2- העתקת שורות שמכילות תת מחרוזת

ממשו את הפונקציה `copy_lines_with_str(in_file_name, out_file_name, target_str)` המקבלת שלוש מחרוזות כפרמטרים, כאשר שתי המחרוזות הראשונות מייצגות נתיבים של קבצים.

הפונקציה תעתיק לתוך `out_file_name` את כל השורות ב `in_file_name` שבהן מופיעה המחרוזת `target_str`. סדר הכתיבה ל `out_file_name` יהיה זהה לסדר הופעת השורות ב `in_file_name`.

במידה וקרתה שגיאת IO בעת הקריאה מהקובץ `in_file_name` או בעת הכתיבה לקובץ `out_file_name`, עליכם לתפוס אותה (ע"י `except`) ולסיים את ריצת התוכנית בצורה מסודרת ללא שגיאה. דאגו לסגור את כל המשאבים הפתוחים ע"י שימוש ב `finally`. במקרה של שגיאת IO – ניתן להשאיר קובץ פלט ריק, כתוב חלקית או כלל לא לייצר את קובץ הפלט. בחרו בעצמכם את אופן המימוש המועדף.

דוגמאות : עבור קובץ הקלט `q2_input_1.txt` שתוכנו הוא :

Now somewhere in the Black Mountain Hills of Dakota There lived a young boy named Rocky Raccoon
--

הרצת הפקודה

```
copy_lines_with_str("q2_input_1.txt", "q2_output_1_Rocky.txt", "Rocky")
```

תייצר את קובץ הפלט q2_output_1_Rocky.txt אשר יכיל רק את השורה השניה של קובץ הקלט (המחרוזת Rocky צבועה באדום בשורה השניה).

הרצת הפקודה

```
copy_lines_with_str("q2_input_1.txt", "q2_output_1_ere.txt", "ere")
```

תייצר את קובץ הפלט q2_output_1_ere.txt אשר יכיל רק את שתי השורות המופיעות בקובץ הקלט, שכן המחרוזת ere מופיעה בשתייהן (המחרוזות צבועות בכחול).

שני קובצי הפלט מצורפים לתרגיל.

לתרגיל מצורף קובץ קלט נוסף q2_input_2.txt וקבצים המכילים את הפלטים הנכונים עבור המחרוזות Rocky, boy, Nancy.

הנחיות:

- הניחו כי target_str אינה ריקה, והיא מכילה תוים שהם אותיות, ספרות וסימני פיסוק בלבד.
- במידה והמחרוזת target_str אינה מופיעה בשום מחרוזת בתוך in_file, הקובץ out_file ייוצר ריק.
- החיפוש הוא case sensitive, כלומר, מבחין בין אותיות גדולות לקטנות. לדוגמא, עבור קובץ הדוגמא לעיל, אם נגדיר את target_str להיות rocky, קובץ הפלט ייוצר ריק.

שאלה 3- תאומים ראשוניים

תאומים ראשוניים הם זוג מספרים ראשוניים שההפרש ביניהם הוא 2. (3,5) הוא זוג של תאומים ראשוניים, וכמוהו גם (11,13), (59,61) וכו'. קיימת השערה (שטרס הוכחה) שיש אינסוף זוגות של תאומים ראשוניים.

(לקריאה נוספת: https://en.wikipedia.org/wiki/Twin_prime)

ממשו את הפונקציה `write_twin_primes(num, out_file_name)` אשר כותבת לתוך הקובץ `out_file_name` את זוגות התאומים הראשוניים הראשונים. כל זוג יכתב בשורה נפרדת, כאשר שני המספרים יופרדו ביניהם על ידי פסיק (,).

לדוגמא, הקריאה

```
write_twin_primes(4, "q3_output_1.txt")
```

תייצר את הקובץ `q3_output_1.txt` אשר מכיל את התוכן הבא :

```
3,5
5,7
11,13
17,19
```

הנחיות :

- אם `num` קטן או שווה ל 0, עליכם לזרוק שגיאה מסוג `ValueError` עם ההודעה הבאה :

`Illegal value num=XXX`

כאשר `XXX` יוחלף בערכו של `num`.

העזרו בפקודה :

```
raise ValueError("Illegal value num={}".format(num))
```

בשביל לזרוק את השגיאה עם המלל הנכון (הפקודה מופיעה גם בגוף שלד התרגיל)

- אם נזרקת שגיאת IO בעת הכתיבה לקובץ `out_file_name` עליכם לתפוס (except) אותה ואז לזרוק (raise) שגיאת `ValueError` עם ההודעה הבאה :

`Cannot write to XXX`

כאשר `XXX` יוחלף בשם הקובץ.

העזרו בפקודה:
`raise ValueError("Cannot write to {}".format(out_file_name))`

בשביל לזרוק את השגיאה עם המלל הנכון (הפקודה מופיעה גם בגוף שלד התרגיל)

הערה: ניתן ואף מומלץ להיעזר במימוש שראיתם בכיתה לבדיקת ראשוניות של מספר (במצגת ההרצאה של שיעור 2). את המימוש לבדיקת ראשוניות של מספר כדאי (אך לא חובה) לממש בפונקציה נפרדת.

שאלה 4- ניתוח נתונים מקובץ

בשאלה זו עליכם לנתח נתונים של דירוגי שירים במצעד ה Billboard האמריקאי עבור שלוש להקות: ABBA, The Beatles, Radiohead. כל שורה בקובץ הנתונים תכיל את שם השיר, שם הלהקה (אחת מבין שלוש הלהקות) ודירוג מקסימלי במצעד, והנתונים בכל שורה יופרדו באמצעות פסיקים (,).

דוגמא לשורה מתוך קובץ קלט אפשרי:

The Winner Takes It All,ABBA,8

נרצה לבדוק מהו הדירוג הממוצע של שירי כל אחת מהלהקות.

ממשו את הפונקציה `calc_avg_position_per_band(in_file_name)`, המקבלת כקלט את הקובץ csv בפורמט המתואר, ומחזירה מילון בו מופיע הדירוג הממוצע עבור כל אחת מהלהקות. כל מפתח במילון יהיה שם הלהקה, וכל ערך יהיה הדירוג הממוצע.

הנחיות:

- במידה ולהקה כלשהי מבין שלושת הלהקות לא מופיעה בקובץ ה csv, תיזרק שגיאת ValueError עם המלל הבא:

At least one of the bands does not appear in the file XXX

כאשר XXX יוחלף בשם קובץ הקלט.

העזרו בפקודה:
`raise ValueError("At least one of the bands does not appear in the file {}".format(in_file_name))`

בשביל לזרוק את השגיאה עם המלל הנכון (הפקודה מופיעה גם בגוף שלד התרגיל)

- הניחו כי הקלט חוקי. בפרט, כל שורה מכילה שני פסיקים שמפרידים בין שם השיר (שאינו כולל פסיקים), בין שם הלהקה (גם כן, ללא פסיקים) ובין מספר שלם. כמו כן, ניתן להניח שבקובץ לא יופיעו להקות אחרות פרט לשלוש המדוברות.
- הדירוג הממוצע ישמר כמספר שלם. במידה והממוצע הוא עשרוני, עליכם לעגל אותו למספר השלם הקרוב ביותר. העזרו בפונקציה round שמעגלת כל מספר למספר השלם הקרוב ביותר.
- הניחו כי שם הקובץ הוא חוקי. אין צורך לטפל בשגיאות IO כתוצאה מהקריאה מהקובץ.

דוגמאות :

עבור הקובץ q4_input_1.txt אשר מכיל את התוכן הבא (השורות המתייחסות לכל להקה צבועות בצבע שונה, לנוחותכם) :

The Winner Takes It All,ABBA,8

While My Guitar Gently Weeps,The Beatles,63

Fake Plastic Trees,Radiohead,11

Something,The Beatles,1

Here Comes The Sun,The Beatles,4

Dancing Queen,ABBA,1

הפונקציה תחזיר את המילון :

{'The Beatles': 23, 'Radiohead': 11, 'ABBA': 4}

הקובץ q4_input_2.txt מכיל תוכן זהה לקובץ הראשון, אך ללא שורה עבור Radiohead. במקרה זה התוכנית תזרוק שגיאת ValueError.