

תרגיל הגשה 2 בקורס

Big Data

תשפ"ד

דגשים מחייבים להגשה - לכל המשימות

1. ההגשה במייל של כל המשימות (בהתאם לפרוט כאן ובסילבוס) בפורמט טקסט בלבד!.
2. חובה לשלוח את כל קבצי קוד, ואת כל הפלטים של ההרצות (קבצי פלט לפי הדרישות) בפורמט טקסט בלבד! – הערה: תרגיל זה כולל סה"כ 6 קבצי פלט שונים.
3. לא יתקבלו הגשות בפורמט אחר!!!
4. ההגשה של תרגיל 2 הינה ללא הגנה
5. הניקוד לכל תרגיל/סעיף – מופיע בתרגיל.
6. יש לשלוח במייל ההגשה את פרטי המגיש.ה המלאים לטובת הערכה.
7. הגשה עד התאריך (כולל) ה - **23:59 29/6/2024**.
8. חובה להשתמש ב 3 קבצי נתונים בהתאם להוראות
9. קובץ שנשלח אחרי המועד = כאילו ההגשה לא בוצעה וניתן ציון "0" (אפס) לתרגיל זה.
10. לא יתקבלו עדכונים, קבצים, תיקונים וכו' אחרי מועד ההגשה.
11. לא יתקבלו קבצי קוד על בסיס חומר שלא נלמד בקורס (כגון: שפה לא רלוונטית, פקודות שלא נלמדו וכו')
12. במקרה כזה (סעיף 11), כאילו העבודה לא בוצע וניתן ציון "0" (אפס) לתרגיל זה.
13. בהצלחה.

הוראות על שימוש בקבצי הנתונים

- יש (חובה!) להשתמש ב 3 קבצי נתונים בהתאם לכל דרישות התרגיל.
- קבצי הנתונים ממוקמים באתר הקורס במודל - תחת "תרגיל מספר 1 להגשה תשפד", הינם הקבצים לפי הסיפורה במיקום ה 2, 4, 7 בת"ז.
- במידה והסיפורה חוזרת מספר פעמים, יש להשתמש באותו הקובץ בהתאם לכמות החזרות.
- במידה והספורה הינה "0" יש להשתמש בקובץ data000.txt
- לדוגמה: נניח וה ת.ז. לדוגמה בלבד, הינה 12345678
- יש לקחת (בהתאם לדרישות) את קובצים: "data007.txt" + "data004.txt" + "data002.txt"

מבנה הקובץ

name	ID	buy1	buy2	buy3	buy4	buy5	buy6	buy7	buy8	buy9	buy10	text1	text2	text3	text4	text5
chararray	INT	INT	INT	INT	INT	INT	INT	INT	INT	INT	INT	chararray	chararray	chararray	chararray	chararray

- תשפ"ד תרגיל להגשה

משימה 1 – 40 נק'

- יש לבנות קובץ חדש [ראשון] (בכלים ובקוד שנלמד בקורס) שמכיל את השדות הבאים בלבד (20 נק') :

1. שם לקוח.ה

2. ID

3. סכום כל הקניות

4. ממוצע כל הקניות

– אם ממוצע הקניות גדולות מ T (יש לבחור מספר), יש לרשום – High average

– אם ממוצע הקניות קטן מ T (יש לבחור מספר) וגדול מ R (יש לבחור מספר), יש לרשום – Mid average

– אם ממוצע הקניות קטן מ R (יש לבחור מספר), יש לרשום – Low average

- הקובץ צריך להיות ממין מהממוצע הגבוה לנמוך

- יש לבנות קובץ חדש [שני] (בכלים ובקוד שנלמד בקורס) שמכיל את רק את הלקוח.ה עם הממוצע הקניה הגובה ביותר, עם הערכים הבאים בלבד: (10 נק')

- הבאים בלבד:

– מספר לקוח.ה

– ID

– ממוצע הקניה הגובה ביותר

- יש לבנות קובץ חדש [שלישי] (בכלים ובקוד שנלמד בקורס) שמכיל את רק את הלקוח.ה עם הממוצע הקניה הנמוך ביותר, עם הערכים הבאים בלבד: (10 נק')

– ID

– שם לקוח.ה

– ממוצע הקניה הנמוך ביותר

דגש:

3 פלטים (3 קבצים) במשימה 1

- תשפ"ד תרגיל להגשה

משימה 2 – 60 נק'

- יש לבנות קובץ חדש [רביעי] (בכלים ובקוד שנלמד בקורס) שמכיל את השדות הבאים בלבד (20 נ'):
 1. שם לקוח.
 2. ID
 3. כמות התווים בכל הטקסטים text1 עד text5 (מספר אחד בודד!)
 4. הצגת של text2 באותיות קטנות
 5. הטקסט ללא המילה " and " ב text1 עד text5
 6. הטקסט ללא המילה " the " ב text1 עד text5
- יש לבנות קובץ חדש [חמישי] (בכלים ובקוד שנלמד בקורס) שמכיל את השדות הבאים בלבד (20 נ'):
 1. שם לקוח.
 2. ID
 3. מספר המילים ב text1 עד text5
 4. מספר התווים שהורדו ב text1 עד text5 לאחר הסרה של המילים " and " ב text1 עד text5 ו " the " ב text1 עד text5
 5. מספר ההופעות של המילים " and " או " the " ב text1 עד text5 (מספר אחד כולל).
- יש לבנות קובץ חדש [שישי] (בכלים ובקוד שנלמד בקורס בלבד) שמכיל את השדות הבאים בלבד (20 נ'):
 1. ID
 2. שם לקוח.
 3. כמות המילים ב text1 עד text5
 4. במידה וכמות המילים גדולה מערך X (יש לבחור X מתאים) יש לרשום "High"
 5. במידה וכמות המילים קטנה מערך X (יש לבחור X מתאים) יש לרשום "Low"
 - יש למיין את הקובץ לפי כמות המילים (ראשי) ומספר לקוח (משני) מהערך הגדול לקטן.