

Python Libraries - Moed A Exam

חלק א'

הסבר במילים שלך-

1. מה היתרונות ומה החסרונות של שימוש במידע גולמי מהלקוח ללא עיבוד או ניקוי?
2. אילו סוגי טעויות עלולות להופיע במידע גולמי שמגיע מהלקוח, וכיצד ניתן לזהות ולתקן אותן?
3. קיבלת תשובות משוב מסקר שערכת לחברת מים מינרלים המכיל: מס' לקוח, תאריך המשוב, מקום מגורים וציון מספרי של איכות השירות (בין 1-10).
בחלק מהרשומות חסר תאריך, הצע דרך פתרון להשלמת המידע.
בחלק מהרשומות חסר מקום מגורים, הצע דרך פתרון להשלמת המידע.
בחלק מהרשומות מופיע פעמיים או יותר רשומות עם אותו מס' לקוח, הצע דרך פתרון לבעיה.
בחלק מהמקומות חסר הציון המספרי- הצע 3 רעיונות (לפחות) להשלמת הציון

חלק ב'

פייטון-

4. לפניך רשימת סקר מדגם מסעיף 3:

גם בלינק

<https://github.com/pythonai170624/exam/blob/main/seker.py>

```
data = [  
{"cust_id": 1, "timestamp": "2024-01-01", "residence": "Tel Aviv", "grade": 8},  
{"cust_id": 2, "timestamp": "2024-01-02", "residence": "Jerusalem", "grade": None},  
{"cust_id": 1, "timestamp": "2024-01-03", "residence": "Tel Aviv", "grade": 8},  
{"cust_id": 3, "timestamp": "2024-01-04", "residence": None, "grade": 7},  
{"cust_id": 2, "timestamp": "2024-01-05", "residence": "Jerusalem", "grade": None}  
]
```

- כתוב קוד (ללא שימוש בספריית PD) אשר מוחק כפילויות עבור אותו הלקוח, ומשאיר רק את המופע הראשון
- כתוב קוד (ללא שימוש בספריית PD) אשר מוחק כל שורה שיש בתוכה את הערך None
- כתוב קוד (ללא שימוש בספריית PD) אשר מחשב את ממוצע הציונים ומחליף את ציוני ה- None בממוצע
- כעת צור DF עם המידע, ובצע את 3 המשימות בפקודות של PD
- השתמש בפקודת apply ו- groupby והחליף את ציוני ה- None בממוצע של אותו העיר (לדוגמא ברשומה שחסר בה ציון והתושב גר בתל-אביב, יוחלף הציון None בממוצע של תל אביב)

חלק ג'

הכנת מידע

לפניך קבצי המידע הבאים:

Customers.csv, Items.csv, order_item.csv, orders.csv

5. חקור את המידע

- הדפס לכל טבלה (בפעולה אחת פר טבלה) את פרטי שמות העמודות, הטיפוס, וכמות העמודות שאינן Null (ועוד)
- הדפס לכל טבלה (בפעולה אחת פר טבלה) את שמות העמודות, הערכים הממוצעים, ערך MIN, ערך MAX חציון (ועוד)
- הדפס לכל טבלה את 5 הערכים הראשונים

6. מחיקת כפילויות

- בטבלת items קיימים מוצרים שמופיעים מספר פעמים (מוצר עם אותו השם) והוחלט לשמור במוצרים בעלי אותו שם, את המוצר עם הכי הרבה פרטים not-null, ואת היתר למחוק.
רמז: (עקוב אחרי הדוגמא שעשינו בשיעור ויישם אותה כאן),
דרך פעולה-
- הוסף עמודה חדשה הסופרת כמה ערכי not-null יש בכל שורה
 - מיין לפי שם המוצר + העמודה החדשה שייצרת, מהגדול לקטן
 - הסר את הכפילויות לפי שם המוצר (שמור רק את המופע הראשון)

7. *אתגר: עדכון טבלת order_item:

- החלף את ה-id של המוצרים עם הכפילויות ל-id של המוצר עם הכי הרבה פרטים not-null. לדוגמא:

48	Camera	Photography Equipment	2039.08	471	Panasonic	TRUE
28	Camera		695.71		Sony	FALSE

כאן, מכיוון שתוסר המצלמה עם id 28 יש לשנות בטבלת order_item את כל המכירות של מוצר המצלמה עם id 28 ל- 48

8. מחק מטבלת customers את כל הלקוחות אשר חסר להם 3 או יותר שדות
9. *אתגר: מחק מטבלת order_item את כל המכירות של הלקוחות שהסרת בסעיף הקודם

חלק ד'

ניתוח מידע

השתמש בקבצי המידע הבאים:

Customers.csv, Items.csv, order_item.csv, orders.csv

1. מה מחיר ממוצע של פריט
2. מי הלקוח שרכש הכי הרבה מוצרים
3. הוסף עמודה total_price לטבלת order_item, המחשבת את עלות הקנייה הכוללת (מחיר מוצר * כמות)
4. הצג כל קנייה ואת המחיר הכולל שלה
5. מצא מה היא הקנייה הכי יקרה שבוצעה? הכי זולה? ממוצע?
6. מטבלת customers, צור טבלת PIVOT ובה הצג כל לאום כאינדקס, ואת ה-gender כעמודות. הצג כמה לקוחות משתייכים לאותו המגדר באותו הלאום. לדוגמא (פלט חלקי של הטבלה...)-

	gender	Agender	Bigender	Female	Genderfluid	Genderqueer
nationality						
Argentina		1	0	10	1	0
Brazil		0	0	6	0	0

חקור והצג בגרף את הפרטים הבאים:

המלצה: הסר שורות עם ערכים ריקים

- (1) פילוג לקוחות לפי מגדר, gender
- (2) פילוג כמות לקוחות לפי לאום
- (3) הסטוגרמת פילוג לקוחות לפי גיל
- (4) גרף המציג כמות לקוחות חדשים שהצטרפו, לפי שנים
- (5) גרף המציג כמות מכירות לפי חודשים
- (6) הסטוגרמת פילוג כמות קניות לפי גיל (יש להחשיב כל קניה בספירה אחת. אין צורך לספור כמות מוצרים וכו') – האם יש מגמה ליותר/פחות קניה בטווחי גיל מסויימים?
- (7) *אתגר: עלות קניות לפי לאום. כלומר- נסכום את סך עלות הקניות שנעשו לכל המוצרים עבור כל לקוח, נחבר ביחד את כל קניית הלקוחות מאותו לאום – ואז נראה כמה עלות יש פר לאום

בהצלחה