**תרגיל בית – הדסים 5**

**קראי בעיון את ההוראות, ממשי באמצעות שפה לבחירתך, הקפידי על קוד קריא ויעיל.**

**מועמדת יקרה,**

נמצאת מתאימה להמשך תהליך המיון לתוכנית הדסים 5.0.

בימים הקרובים תשובצי לראיון מקצועי.

כתנאי לקיום הריאיון – עלייך להגיש מבחן בית כמפורט בהמשך.

הנחיות לראיון:

הריאיון **יתקיים** ב-Microsoft teams בפני מראיינים מקצועיים.

לינק אישי להתחברות ישלח אליך מראש.

**שימי לב** – עלייך להתקין מראש את תוכנת Microsoft teams ולהכין שם משתמש.

במהלך הריאיון נשוחח גם על תרגיל הבית ותדרשי להציג את דרך פתרון התרגיל.   
ייתכן שנבקש לבצע תרגיל נוסף במהלך הריאיון.

**משך הריאיון:** כ-60 דקות.

נבקשך להתחבר 10 דקות מראש ולהמתין בחדר ההמתנה הווירטואלי שלנו.

חובה להתחבר ממחשב נייד או נייח, עם חיבור יציב לרשת אינטרנט, מיקרופון, מצלמה פתוחה, בסביבה שקטה המאפשרת קיום ראיון. לא נוכל לאפשר ראיון מטלפון סלולארי.

עליך לוודא מראש כי העמדה ממנה תתחברי תקינה, מאפשרת שיחת זום/ Microsoft teams והצגת תרגיל הבית שביצעת.

חובה להציג תעודה מזהה (תעודת זהות/דרכון/רישיון נהיגה).

בריאיון תתבקשי להציג גיליון ציונים ממוסד לימודיך-אנא הכיני מראש.

הגשת תרגיל הבית:

**מועד ההגשה: יום שלישי י' בניסן התשפ"ה, 08/04/2025 עד השעה 23:59.**

איחור/אי הגשה/ אי הגשה בהתאם להנחיות - יגרור ביטול הראיון המקצועי שנקבע לך והפסקת תהליך המיון.

אופן ההגשה- שליחת הפתרון:

עם סיום העבודה, בצעי Tag בשם "FINAL\_VERSION" בענף בו נמצאת הגרסה הסופית של הקוד.

את הקישור ל- Repository יש לשלוח בגוגל פורמס שקיבלת, יש לציין בכותרת "הגשת מבחן בית-הדסים 5.0" בציון שמך המלא ותעודת זהות בכותרת המייל.

במידה ולא מתאפשר להעלות את הקוד ל-Repository Git ,יש לצרף את הקוד כקובץ ZIP (קובץ דחוס). אופן הגשה אחר -לא יתקבל.

המבחן לביצוע אישי שלך בלבד. חשד להעתקה ממועמדת אחרת או ממקורות חיצוניים יוביל להפסקת תהליך המיון.

**בהצלחה רבה!**

**חלק א'**

**סעיף א'**

מצורף קובץ לוג גדול (logs.txt) המכיל מיליוני שורות. כל שורה מייצגת רישום לוג ומכילה קוד שגיאה. עלייך למצוא את N קודי השגיאה השכיחים ביותר.

**משימות:**

1. כתבי קטע קוד לפיצול קובץ ה-logs.txt לחלקים קטנים יותר.
2. עבור כל חלק, ספרי את השכיחות של כל קוד שגיאה.
3. חברי את ספירות השכיחות מכל החלקים.
4. מצאי את N קודי השגיאה השכיחים ביותר מהספירות הממוזגות.
5. נתחי את סיבוכיות הזמן והמקום של הפתרון שלך.

**קלט:** קובץ טקסט גדול (logs.txt) ומספר שלם N.

**פלט:** רשימה של N קודי השגיאה השכיחים ביותר וספירותיהם.

בחרתי לכתוב בשפת פיתון כי היא מכילה טכנולוגיות שמתאימות לעבודה הזו, כג'

* ספרית PANDAS המותאמת לעבודה עם קבצים גדולים בפורמט טבלאי.
* ספרית NUMPY המכילה אופציות מהירות למיונים, חישובים וסכימות. יותר יעיל מלולאות.
* קוד נקי, ברור וקומפקטי.
* סטנדרטים מקובלים למשימות כאלו.
* מאפשר עיבוד מקבילי בקלות לאחר שינויים קלים

X JAVA – קוד ארוך, מורחב ולא יעיל

X C# - עלות פיתוח גבוהה, המון קוד

אחרות לא מתאימות לתנאים דינמיים

**סעיף ב**'

1. מצורף קובץ נתונים גדול של סדרת זמן (time\_series.csv) המכיל שתי עמודות:

* חותמות זמן – timestamp
* ערכים מספריים – value

קלט לדוגמא:

|  |  |
| --- | --- |
| Timestamp | value |
| 2025-06-10 6:10:00 | 15.3 |
| 2025-06-10 7:49:00 | 3.2 |
| 2025-06-10 7:34:00 | 12.6 |
| 2025-06-10 6:55:00 | 5.3 |

1. כתבי קטע קוד המבצע בדיקות לפני עיבוד הנתונים כמו בדיקה של פורמט התאריך או כפילויות, חשבי על בדיקה נוספת.
2. כתבי קטע קוד המחשב את הערך הממוצע עבור כל שעה.

|  |  |
| --- | --- |
| זמן התחלה | ממוצע |
| 2025-06-10 6:00:00 | 10.3 |
| 2025-06-10 7:00:00 | 7.9 |

פלט רצוי:

כמו בשאלה הקודמת וגם מאפשר **שדרוג קל לעיבוד מבוזר או בזמן אמת**

כאן כתבתי בטעות MEDIAN במקום MEAN וצריך לשנות את זה

1. בעולם בו כמות הדאטה היא עצומה – נהוג לחלק את הנתונים לחלקים קטנים ולפתור את הבעיה עבור כל חלק בנפרד,

לאחר מכן מחשבים את התוצאה הסופית באמצעות איחוד כל התוצאות.

ממשי פתרון לבעיה מהסעיף הקודם באמצעות השיטה המתוארת לעיל.

כתבי קוד המחלק את קובץ ה-time\_series.csv לחלקים קטנים יותר המבוססים על זמן (למשל, חלקים יומיים).

חשבי את הממוצעים השעתיים עבור כל קובץ בנפרד,

ולבסוף אחדי את התוצאות וצרי באמצעותן קובץ סופי כמו בדוגמת הפלט בסעיף ב נקודה 1.

1. אם הנתונים מגיעים בזרימה (stream) במקום מקובץ, כיצד תתכנני את הפתרון כדי לעדכן את הממוצעים השעתיים בזמן אמת?
2. לתרגיל מצורף קובץ נוסף בפורמט שונה המכיל את אותו מידע, התאימי את החלקים הרלוונטיים בקוד על מנת לתמוך בפורמט החדש, וכתבי תיעוד המתאר את היתרונות באחסון המידע בפורמט הנתון.

**קלט:**

* קובץ CSV (time\_series.csv) עם עמודות חותמת זמן וערך.
* קובץ PARQUET (time\_series.parquet) עם עמודות חותמת זמן וערך.

**פלט:** קובץ CSV או פלט מובנה המכיל ממוצעים שעתיים (חותמת זמן, ערך ממוצע).

**חלק ב'**

**ט​בלאות (כתובי‌ם בהתאמה)**

Ρerso​n\_Id | Рersona​l\_Νame ​| Fam‌ily\_Name | Gen‌der | Fathеr\_​Id | Mother\_Id | Spo​uѕe\_Id‌

ID​, שם פרטי, שם משפחה , מ‌גדר, ID של א‌ב, ID של​ א‌ם, ID ש​ל בן/בת זוג

**תרגיל‌ 1 הקמת עץ ​משפחה**

לבצע טבלה ‌שיש בה כמה רשומות​ לכל בן אדם בהן יופי‌עו כל האנ‌שים בקרב‌ה ראשונה אליו.

הטב​לה צריכה להיות​ בנויה כ‌ך:

Person\_‌Id | Relativ​e\_Id | Co‌nnection‌\_Type

Connection\_Typ‌e ゠ יכול להכיל את ​הערכים הבאים:‌ א​ב, אם, אח, אחות, בן​, בת,​ בן זוג ובת זוג.

**תרגי‌ל 2 השלמת​ בני / ב‌נות ז‌וג**

השלמ​ה ש​ל בני ז​וג. לקחת בחשב‌ון שיש נתו​נים חלקיים‌. ‌כלומר, לאדם עם‌ ID 111 ‌מוגדר בת זוג עם ID 2​22. אבל, לבת זו‌ג עם ID 222 לא מוג​דר שיש בן זוג בכלל​. צריך לבצע ​השלמה שבטבל​ה שיצרנו גם‌ לבנאדם עם ​ID 222 יוגדר בן ז‌וג.

**חלק ג'**

1. דרגי​ מ 1-6 (1- ל​א מע‌וניינת, ​6- מאוד ‌מעוניינת​), האם ​הינך מעוניינת לעסוק​ בתפקי‌די חומרה/אלקט‌רוניקה (ל​דוג'- ​פ‌יתוח מעגלים​ חש‌מליים, אינטגרציית מ‌ע​רכת, בדיקו​ת חשמליות וכד')?
2. א​יך ע​ובד שלט של מזגן? יש‌ להתייחס לסעיפ​ים הבאים (שימי לב,​ ניתן ל‌הציע מגוון אפשרויו‌ת‌ למימ​וש אי​ן תשובה 'נכונה'​):​
   1. מה​ו אופן השידור ‌בין השלט למ​זגן? ‌
   2. אילו רכיבים צריכ‌ים ל‌היות בצד השלט ואי​לו ב​צד המזגן?​
   3. איך המזג‌ן 'יודע' על א‌יז‌ה לח‌צן לוחצים ‌בשלט? (במילים א​חרות- איך ניתן לייצ​ג מקשים ש​ונים בשידור מ​השלט למזגן?), יש לה‌ציע ​מספר אפשרו​יות.

**חלק ד'**

**מערכת לניהול מכלות:**

בעל המכולת השכונתית ביקש ממך לפתח עבורו מערכת שתסייע לו בניהול החנות.  
המערכת תכיל צד שרת, אליו יוכל להתחבר בעל המכולת, וצד לקוח עבור הספקים שלו.

**שלב א'**

עלייך לפתח את המערכת בהתאם לדרישות הבאות:

### צד לקוח – כניסת ספק

* רישום משתמש חדש למערכת / כניסה למשתמש קיים
* צפיה בהזמנות שביצע בעל המכולת- getOrderByProviderId getOrderProductsByIdלשניהם
* אישור הזמנה – הזמנה שאושרה תעבור לסטטוס "בתהליך". פוט של עסקה- לשניהם

בעת הרישום כל ספק ימלא את הפרטים הבאים:

* שם חברה
* מספר טלפון
* שם נציג
* רשימת הסחורות שהספק מציע (במבנה נתונים לבחירתך)

### צד שרת – כניסת בעל המכולת

* הזמנת סחורה מספק. Post order
* צפייה בסטטוס הזמנות קיימות getall orders
* אישור קבלת הזמנה – תעבור לסטטוס "הושלמה" ותעדכן את הספק שההזמנה התקבלה. פוט לשניהם
* מאגר של כל ההזמנות (כולל כאלו שהושלמו) getall orders

רשומה של סחורה צריכה להכיל את הנתונים הבאים:

1. שם המוצר
2. מחיר לפריט
3. כמות מינימלית לרכישה

עצבי את פורמט ביצוע הזמנה באופן הנוח ביותר.

### דגשים

1. ניתן לבחור כל סוג מאגר מידע איתו את מרגישה בנוח לעבוד. לדוגמא: Mongodb, SQL.
2. את צד השרת ניתן לממש בכל טכנולוגיה איתה את מרגישה בנוח לעבוד. לדוגמא:  
    Node.js, python, asp.net, java, Csharp
3. יש לממש API אליהם יוכל לפנות צד הלקוח עם routeים של **שליפה והכנסה בלבד** (אין צורך לממש עריכה ומחיקת רשומות).

**סעיף בונוס – ניהול סחורות והזמנה אוטומטית**

בעל המכולת התלהב מהמערכת, ורוצה לנהל בעזרתה את כל הסחורות בחנות.  
המערכת תכיל מאגר של כלל הסחורות במכולת, ותתחבר לקופה דרך API.  
לכל סחורה בעל המכולת יגדיר מהי הכמות המינימלית שלה שהוא מעוניין שתהיה בחנות בכל זמן נתון. ברגע שכמות הסחורה תרד מתחת לערך הזה, המערכת תבצע הזמנה אוטומטית מהספק שמציע את המחיר הטוב ביותר עבור אותה סחורה. אם אין אף ספק שמספק את הסחורה המדוברת, היא תתריע על כך.

**API של הקופה**

בכל פעם שתתבצע קנייה השרת יקבל קובץ JSON הבנוי באופן הבא:

{

"merchandise\_1": quantity,

"merchandise\_2": quantity,

…  
}

לדוגמה, קנייה של 2 קרטוני חלב וכיכר לחם תראה כך:

{

"bread": 1,

"milk\_carton": 2

}