

# פרויקט: מנתה לוגים של תעבורת רשת

שלב 1 - טעינה וסינון בסיסי

## רקע

קובץ לוג (Log File) הוא קובץ שמתעד פעילות המערכת. בפרויקט שלנו נעבד עם לוג של תעבורת רשת - כל שורה מייצגת חבילה מידע שעברה ברשת. המטרה: לבנות מערכת שמזהה תעבורת חשודה ומודוחת עליה.

## מבנה הנתונים

קובץ הלוג הוא קובץ CSV (Comma Separated Values) - קובץ טקסט שבו כל שורה מכילה ערכים מופרדים בפסיקים.

כל שורה בקובץ בנוייה כך:

```
timestamp,source_ip,dest_ip,port,protocol,size
```

דוגמה	משמעות	שדה
08:23:45 2024-01-15	תאריך ושעה	timestamp
192.168.1.100	כתובת IP מקור (מי שלח)	source_ip
10.0.0.5	כתובת IPיעד (למי נשלח)	dest_ip

443	פורט - מספר שמצוין סוג שירות	port
HTTPS	פרוטוקול התקשורת	protocol
1024	גודל החבילה בבייטים	size

דוגמה לשורה:

2024-01-15 08:23:45, 192.168.1.100, 10.0.0.5, 443, HTTPS, 1024

## הادرת תעבורת חשודה

במערכת שלנו נגדר 4 סוגי חשודות:

הדרה	סוג חשד	קוד
כתובת מקור שלא מתחילה ב- 192.168.1.100 או 10.	IP חיצוני	EXTERNAL_IP
포רטים: 22 (SSH), 23 (Telnet), 3389 ,(RDP)	פורט רג'יש	SENSITIVE_PORT
חבילה מעל 5000 בייט	גודל חריג	LARGE_PACKET
פעילות בין 00:00 ל-06:00	שעות לילה	NIGHT_ACTIVITY

**שיםו לב:** המערכת צריכה להיות דינמית - כלומר כל להוסיף סוג חדש  
נוספים בעתיד בלי לשנות הרבה קוד.

דרישות שלב 1

.**List Comprehension** בשלב זה נتمكن בטעינה הנתונים וסינון בסיסי באמצעות

לכל דרישת יש לכתוב פונקציה נפרדת.

טעינת הקובץ 1

כתבו פונקציה שמקבלת נתיב לקובץ CSV ומחזירה רשימה של רשימות - כל שורה כרשימה של שדות.

2 חילוץ כתובות ISO חיצונית

כתבו פונקציה שמקבלת את הנתונים ומחזירה רשימה של כתובות IP מקור  
חיצונית בלבד.

**למה חיצונית?** כתובות שמתחלות ב- 192.168.10. הן כתובות פנימיות (פרטיות) - הן שייכות לרשף המקומית שלנו. כל כתובה אחרת היא חיצונית - מוגעה מהאינטרנט, ולכן עשויה להיות מקור לאיום.

סינון לפי פורט רגיש 3

כתבו פונקציה שמקבלת את הנתונים ומחזירה רשימה של כל השורות עם פורט רגיש (22, 23, או 3389).

## למה פורטיטם אלה רגישיים?

- **פורט 22 (SSH)** - מאפשר שליטה מרוחק במחשב דרך שורת פקודה
  - **פורט 23 (Telnet)** - פרוטוקול ישן לשיליטה מרוחק, לא מוצפן ולכן מסוכן
  - **פורט 3389 (RDP)** - מאפשר שליטה מלאה במחשב מרוחק Windows

גישה לפורטים אלה מ-IPו חיצוני עלולה להעיד על ניסיון פריצה.

#### 4 סינון לפי גודל

כתבו פונקציה שמקבלת את הנתונים ומחזירה רשימה של כל השורות עם חבילות מעל 5000 ביט.

#### 5 תיאוג תעבורות

כתבו פונקציה שמקבלת את הנתונים ומחזירה רשימה שבה כל שורה מתויגת: "NORMAL" אם הגודל מעלה 5000, אחרת "LARGE".

## דוגמה לפט הסופי (לשלבים הבאים)

כך יראה הפלט בסוף הפרויקט:

```
{ "45.33.32.156": [ "EXTERNAL_IP", "LARGE_PACKET" ],  
"87.120.5.22": [ "EXTERNAL_IP", "NIGHT_ACTIVITY",  
"SENSITIVE_PORT" ], "203.0.113.50": [ "EXTERNAL_IP" ] }
```

כל IP חדש ממופה לרשימת סוג החשדות שלו.