

# פרויקט למידת מכונה

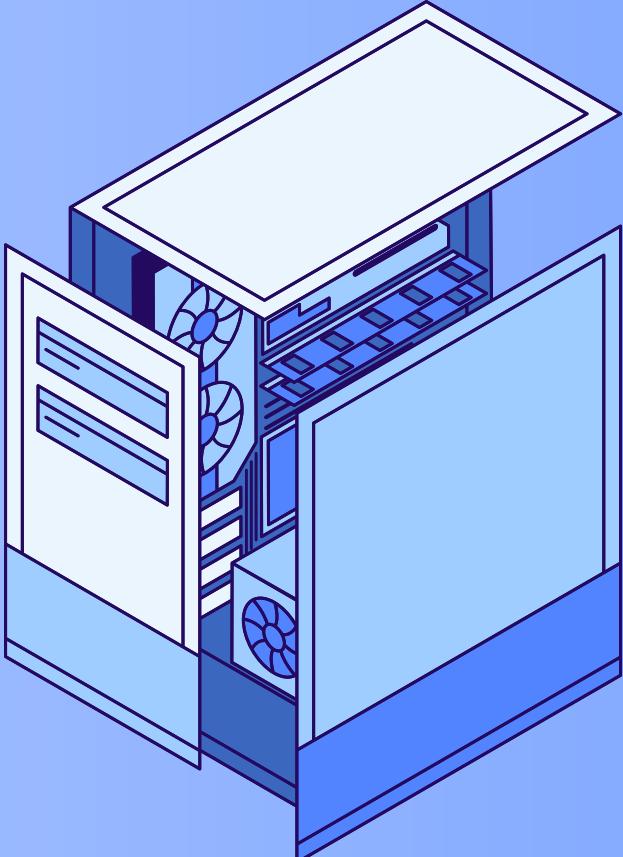
מרצה : חן חנגן'

gal plg, orhab zr

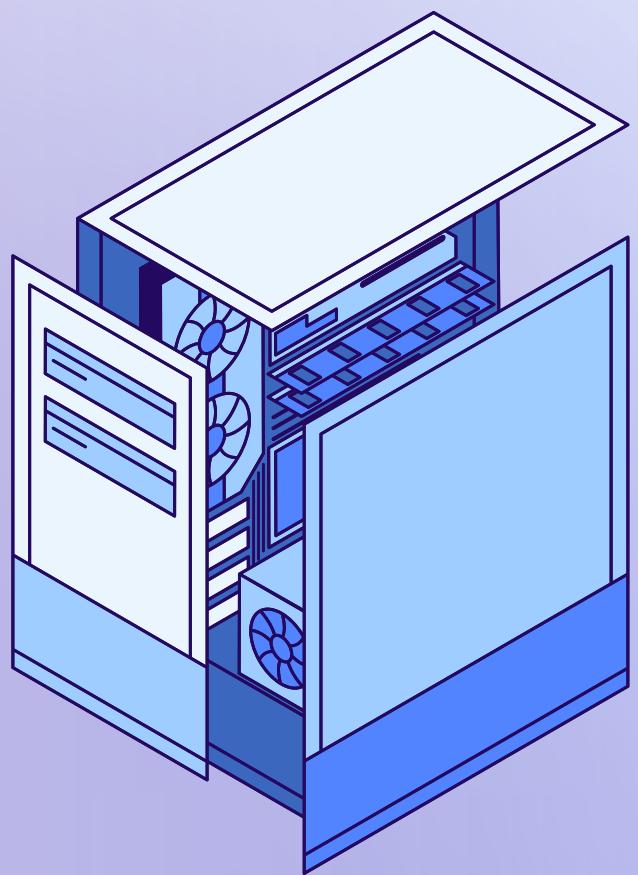


# הפרויקט שלנו

בפרויקט ניתחנו נתונים של משטרת אוסטין (טקסס) המתעד אירועי פשע سنאנה בין השנים 2017–2025. פשע שנאה אלו עבירות פליליות שהמניע שלהן זהה דעות קודומות כלפי קבוצות מסוימות באוכלוסייה, כגון גזע, דת או נטייה מינית.



נושא זה בעל חשיבות רבה - ניתוח הנתונים והבנתם יכול לעזור לחשוט דפוסי התנהגות של הפשעים, ובכך לסייע לגורמי אכיפת החוק ולקהילה להבין היכן,מתי וכליפי מי מתרחשים האירועים ובפיתוח אסטרטגיות מניעה יעילות ובהגברת תחושת הביטחון הציבורי.



המטרה המרכזית של הפרויקט היא לבחון מודלי חיזוי ולבדוק האם ניתן לחזות את יעד התקיפה הבא, כלומר את המניע העומד מאחוריו פשע השנה – האם מדובר בפשע על רקע גזעני, דתי או נטייה מינית בהתבסס על מאפייני האירוע כפי שהם מדוחים נתונים

# גישה אחורית בה לא השתחשנו

## בפרויקט: רשת ניירוניים

**מודל זה מסוגל ללמד קשרים לא-lienאריים מורכבים בין מאפייני האירוע לבין המנייע לפשע, יוכל לשפר ביצועים כאשר קיימים דפוסים מורכבים בדאטה.**

הדאטה שלנו הוא יחסית קטן (300 שורות) ורשתות ניירוניים דורשות לרוב כמות גדולה יותר של נתונים וכוונון פרמטרים מדויק. בנוסף רשתות ניירוניים נחבות לפחות ניתנות לפרשנות (לא כמו לדוגמה מודל הרגרסיה הלינארית שנตอนן קשר ישיר בין פרמטרים) ולכן בהקשר של ניתוח פשיעה, זה מהו זהירותן, משומש שגורמי אכיפת החוק זוקקים לא רק לתוצאות מדויקות, אלא גם להבנה של הגורמים שהובילו אליהן.

# DATA FEATURES

Month	Incident Number	Date of Incident	Day of Week	Number of Victims under 18	Number of Victims over 18	Number of Offenders under 18	Number of Offenders over 18	Race/Ethnicity of Offenders	Offense(s)	Offense Location	Bias	Zip Code	APD Sector
0	Jan 2017-241137	01/01/2017 12:00:00 AM	Sun	0	1	0	1	White/Not Hispanic	Aggravated Assault	Park/Playground	Anti-Black or African American	78704.0	Henry
1	Feb 2017-580344	02/01/2017 12:00:00 AM	Wed	0	1	0	1	Black or African American/Not Hispanic	Aggravated Assault	Highway/Road/Alley/Street/Sidewalk	Anti-White	78702.0	Charlie
2	Mar 2017-800291	03/21/2017 12:00:00 AM	Tue	0	0	0	0	Unknown	Destruction	Highway/Road/Alley/Street/Sidewalk	Anti-Jewish	78757.0	Ida
3	Apr 2017-1021534	04/12/2017 12:00:00 AM	Wed	0	0	0	0	White/Unknown	Simple Assault	Air/Bus/Train Terminal	Anti-Jewish	78723.0	Ida

Feature Name	מה הפיצ'ר מייצג	הסבר
Total Victims	מספר הקורבנות באירוע	חיבור קורבנות מתחת ומעלה גיל 18
Total Offenders	מספר התוקפים באירוע	חיבור תוקפים מתחת ומעלה גיל 18
Victim-Offender Ratio	יחס קורבנות-תוקפים	חישוביחס עם smoothing למניעת חלופהAPS
Day of Week	היום בשבוע של האירוע	חילוץ מהתאריך (0-6)
Weekend Indicator	האם האירוע התרחש בסוף"ש	בינאי: 1 = שבת/ראשון, 0 = יום חול
Season	עונה השנה של האירוע	Winter / Spring / Summer / Fall החודש
Zip Code Frequency	שכיחות אירועים באזורי	ספרת מופעים לכל Zip Code
Bias	מניע פשע השנה	אחד לארבע קבוצות-על: מוצא אתני, נטייה מינית, דת, אחר.
Offense Location Group	סוג מקום האירוע	מקומות > 5 מופעים אחדו Other_Location

```
array(['Anti-Black or African American', 'Anti-White', 'Anti-Jewish',
       'Anti-Gay (Male)', 'Anti-Hispanic or Latino', 'Anti-Transgender',
       'Anti-Islamic (Muslim)', 'Anti-Black', 'Anti-Lesbian', 'Anti-Gay',
       'Anti-Hispanic', 'Anti-Arab', 'Anti-Disability',
       'Anti-Hispanic/Latino', 'Anti-Islamic(Muslim)',
       'Anti-Gay (Male); Anti-Transgender',
       'Anti-Gay (Male); Anti-Jewish', 'Anti-Buddhist',
       'Anti-Lebian/Gay/Transgender', 'Anti-Lebian (Female)',
       'Anti-Other Race/Ethnicity/Ancestry',
       'Anti-Lebian/Gay/Bisexual/Transgender (Mixed Group)',
       'Anti-Bisexual', 'Anti-Protestant', 'Anti-Asian',
       'Anti-Physical Disability',
       'Anti-Black or African American; Anti-Gay (Male)',
       'Anti-American Indian/Alaskan Native', 'Anti-Female',
       'Anti-Multi Races, Group', 'Anti-Multi-Races Group; Anti-Islamic',
       'Anti-Religion (Other)', 'Anti-Other Religion',
       'Anti-Gay (Male); Anti-Black or African American',
       'Anti-Lebian/Bisexual/Transgender (Mixed Group)',
       'Anti-White/Anti-Hispanic or Latino',
       'Anti-Physical Disability; Anti-Black or African American',
       'Anti-Black or African American; Anti-Hispanic or Latino; Anti-Lebian/Gay/Bisexual/Transgender (Mixed Group)',
       'Anti-Other Christian'], dtype=object)
```

הדאטה המקורי כל 41 קטגוריות שונות של פשעי שנאה.  
על מנת לשפר את יכולת החיזוי של המודל ולהפוך את  
הניתוח לעיל יותר איחדנו את כל הקטגוריות הללו ל-4  
**קטגוריות**

# METHODOLOGY

## Unsupervised Techniques

### Clustering

זהה קבוצות של אירועים דומים וחושף דפוסים הקשורים למניע, מיקום או מאפייני קורבן ותוקף, ומספק תובנות נוספות למידה מונחית.

### Anomaly / Outlier Detection

מאתר אירועים חריגים, כגון מבניינים נדירים או מספר>Kיצוני של קורבנות, עשויים להעיד על מקרים בעלי סיכון גבוה או חריגות מיוחדת.

### Supervised Models

#### Logistic Regression

מודל פשוט המאפשר להעריך את ההשפעה של כל פיצ'ר על סוג המנייע

#### Decision Tree

löcid קשרים לא-לינאריים בין פיצ'רים כגון מיקום, יום וסוג העבירה, ויוצר חוקים המפרידים בין סוגי מניעים שונים.

#### Random Forest

מודל של עצים רבים שטפחית התאמת יתר ומשפר את דיוק הסיווג.

#### Gradient Boosting

זהה דפוסים עדינים בנתונים המבדילים בין סוגי מניעים, ומתאים במיוחד למצבים שבהם ההבדלים קטנים בין הקטגוריות

#### Support Vector Machine (SVM)

יעיל בהפרדת סוגים מניעים כאשר קיימות אינטראקציות מורכבות בין פיצ'רים, באמצעות יצרת גבולות החלטה אופטימליים במרחב רב-ממדי.

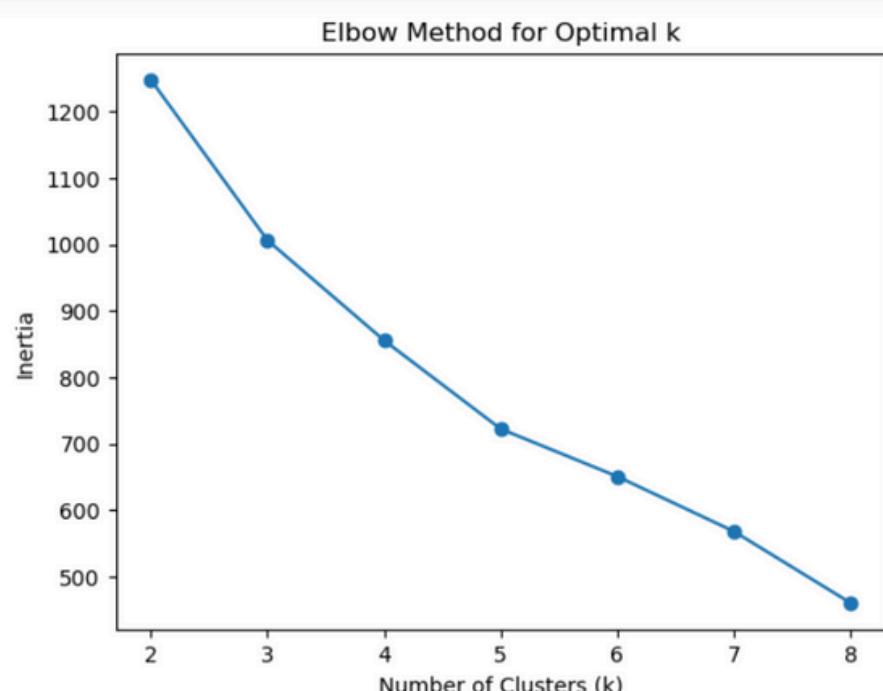


# ניסויים והערכת ביצועים



## למידה לא-מנוחית מערך הניסוי

- נבחרו פיצ'רים המציגים מאפיינים קורבן/תיקף, זמן ומקום Total Victims, Total Offenders, Victim-Offender Ratio, Zip Code Frequency, Is (Weekend, Is Offender Known)
- ערכים חסרים הושלמו ב-0 וונעשה שימוש ב-StandardScaler עלקב רגישות - K-Means לגדיל הפיצ'רים.
- מספר האשכולות נבחר באמצעות שיטת המרפק (Elbow Method for Optimal k), ונמצא כי  $k=5$  מספק איזון בין פירוט לאינטראטיביות. יושם Clustering K-Means, ובוצעה הפחיתת ממדים באמצעות PCA לצורך ויזואлизציה.



## מערך הניסוי

הדאטה חולק לסט אימון וסט בדיקה. אומנו מספר מודלים מונחים לצורכי סיווג מניע פשע השנאה (Bias Super-Class). כל המודלים הוערכו תחת אותו מערך ניסויים לצורכי השוואת הוגנת.

## טכניות הערכה

- Accuracy – מדידת דיוק כולל.
  - Precision, Recall, F1-score
- הערכת איכות הסיווג ברמת המחלקות.

## שיטות לא-מנוחיות

- Clustering – ניתוח מבנה ודפוסים בתונונים.
- Outlier Detection – זיהוי אירועים חריגים וקיצוניים.

# RESULTS & CONCLUSION

## Supervised Models

דיק המודל המפוקח הגבואה ביותר שהושג היה 0.63 למרות שהתוצאה הזאת לא הכי גבוהה מדובר בתוצאה משמעותית בהתחשב בפייזור הרחב והמשתנה של הנתונים, ולא באיכות ירודה של המודל. גם לאחר מיזוג הנתונים ל-4 קטגוריות מרכזיות אומנם החיזוי עלה אך גודל הדאטה נשאר קטן יחסית ביחס לדפוסים המורכבים. ולכן השגת רמת חיזוי זו בתנאים אלו משמעותית ומשתפרת באופן משמעותי פניו סיכויי אקראי, ומוכיחה שהמודל הצליח לזהות את האותות הבסיסיים בנתונים.

1]:

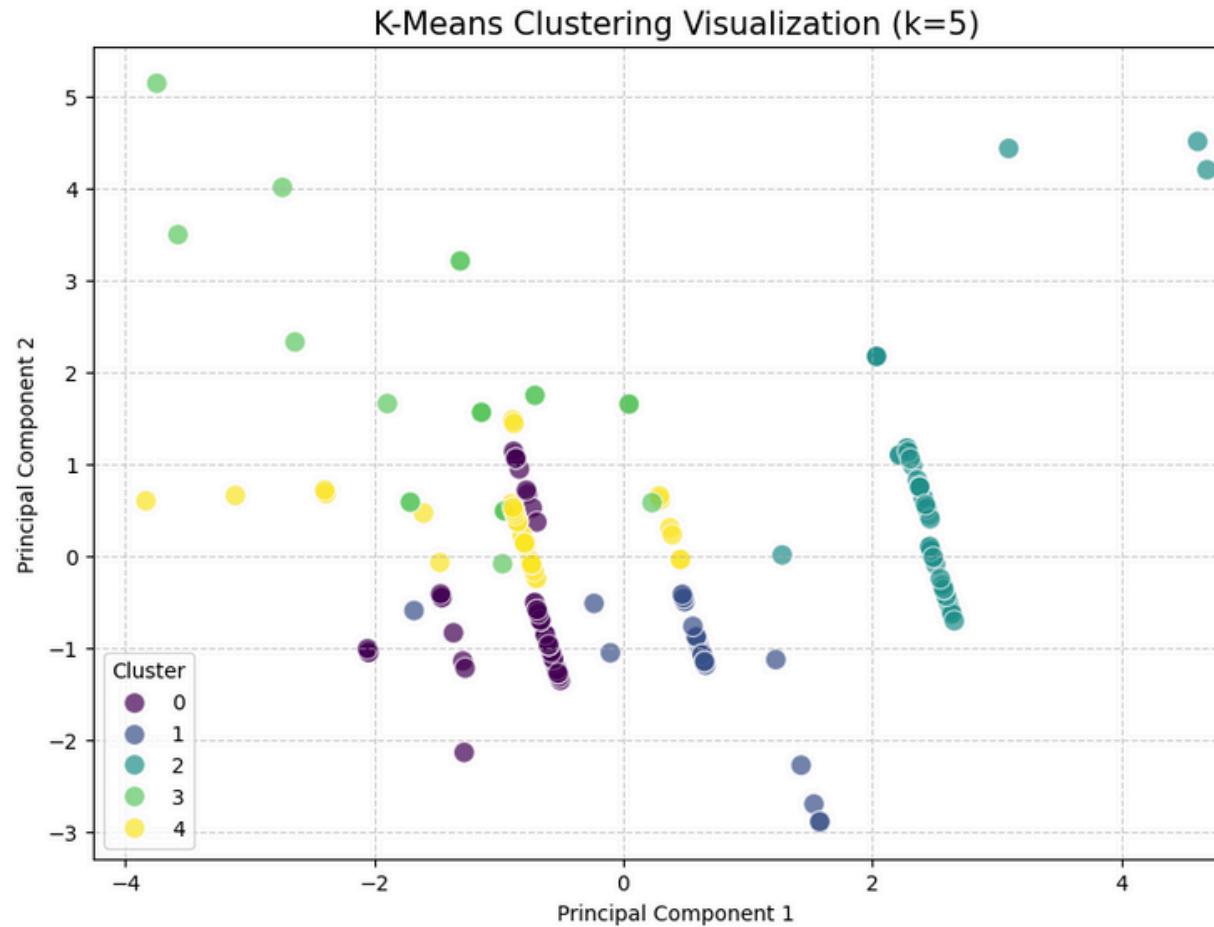
	Model	Accuracy	Weighted F1-score
0	Logistic Regression	0.48	0.49
1	Decision Tree	0.57	0.57
2	Random Forest (Tuned)	0.63	0.63
3	Gradient Boosting	0.57	0.57
4	SVM	0.52	0.52

# RESULTS

## & CONCLUSION

Unsupervised models:

### Clustering



המאפיינים שנבחנו מאפשרים זיהוי דפוסים נסתרים  
ומשלימים את המודלים המפוקחים להבנת מגמות  
הפשעים.

Cluster	Total_Victims	Total_Offenders	Victim_Offender_Ratio	Zip_Code_Frequency	Is_Weekend	Is_Offender_Known
0	1.075472	1.122642	1.028302e+00	13.838095	0.000000	1.000000
1	0.833333	1.000000	7.750000e-01	11.400000	0.033333	0.000000
2	1.078125	0.000000	1.078125e+06	17.062500	0.343750	0.015625
3	1.162791	1.348837	9.399225e-01	47.372093	0.581395	0.860465
4	1.035088	1.228070	9.669591e-01	12.719298	1.000000	0.894737

**Cluster 0 - פשע يومي ידוע:** מעשים יומיומיים ביום השבוע, לרוב בין אנשים שמכירים זה את זה, עם קורבן אחד ופושע ידוע.

**Cluster 1 - אירועים אוניבריים שגרתיים ביום השבוע:** אירועים יומיים ביום השבוע ללא פורע ידוע, לרוב מקרי עבירה קטנים או דיווחים בלי חשוד.

**Cluster 2 - אוניבריים/ונדליזם:** עבירות לא-אישיות או דיווחות (כגון ונדליזם), ללא מעורבות ישירה של פורע וקורבן.

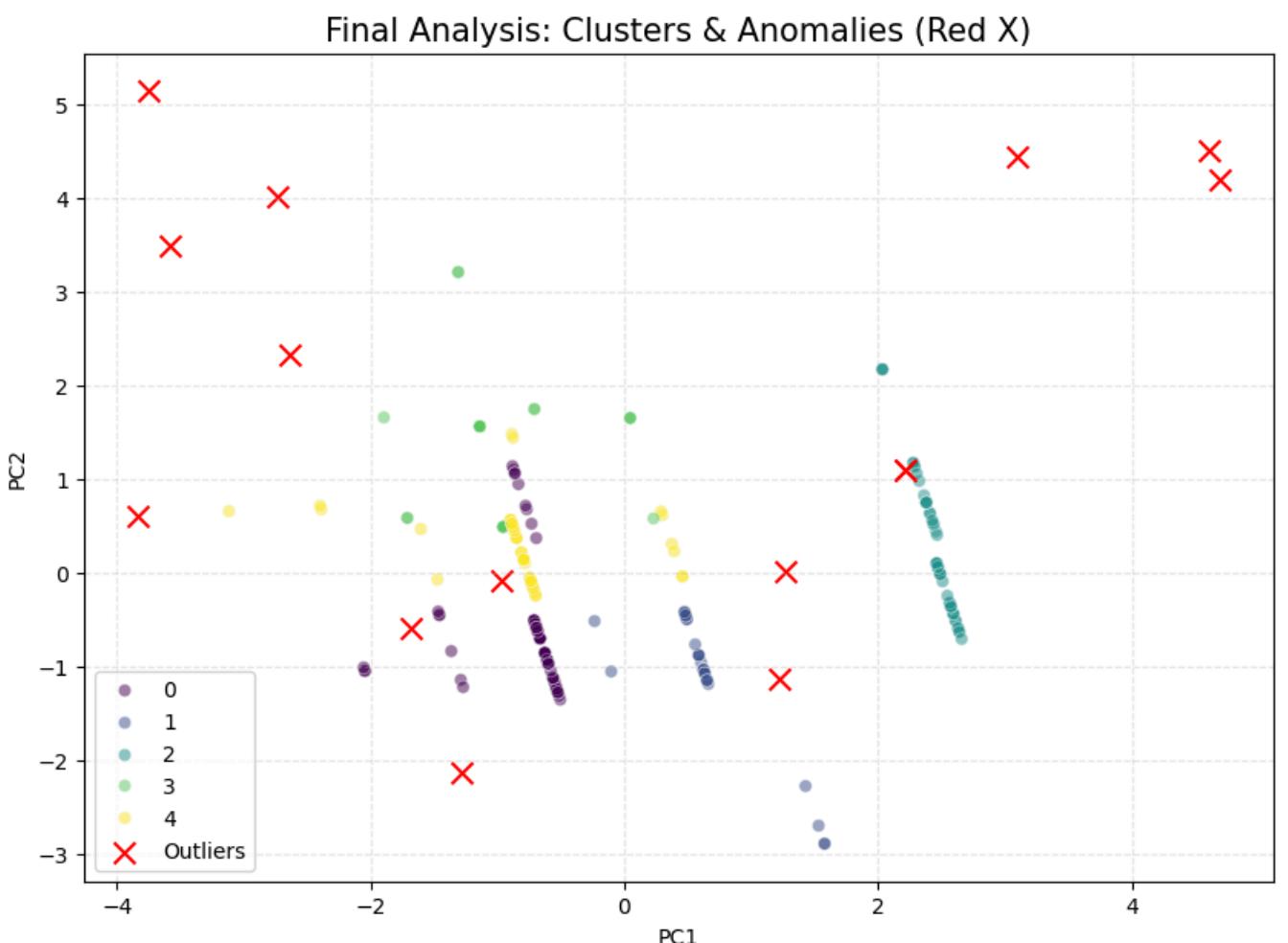
**Cluster 3 - מוקדי פשע צפופים עם נתיחה לסופי שבוע:** מוקדי פשע צפופים, בעיקר בסופי שבוע, עם פורעים מזוהים לעיתים קרובות.

**Cluster 4 - אירועי עימות בסופי שבוע:** אירועים תוקפניים בסופי שבוע, עם מעורבות ישירה של פורעים וזהויות גבוהות שלהם.

# RESULTS & CONCLUSION

Unsupervised modrls:

## Outlier Detection



Total_Victims	Total_Offenders	Victim_Offender_Ratio	Super_Bias_Class
21	1	4	RACE_ETHNICITY
22	0	2	RACE_ETHNICITY
37	2	4	SEXUAL_GENDER
80	3	0	3000000.00 RACE_ETHNICITY
159	2	0	2000000.00 SEXUAL_GENDER

נקודות רגילהות: 283  
חריגות: 15

- חלק מהאירועים כוללים מספר חריג של עבריינים (למשל 4), דבר נדיר בפשע שנאה.
- אירועים עם 0 עבריינים אך קורבות גביה יוצרים יחסים קורבן-ערביין קיצוני, לרוב מקרים אוניבימיים כמו ונדייז או גרפייטי.
- מקרה אחד ללא קורבות מצביע על פשע שנאה ללא מגע ישיר, גם הוא כנראה גרפייטי או נזק לרכוש.

# CONCLUSION & FUTURE DIRECTIONS

**פשי שנה אינטלקטואליים: קיימת דפוסיות לפי זמן (שבוע/סוף שבוע), מקום (אזורים בערים מול רגעים) וסוג המקרה (אנונימי/ונדלים מול אירוע קבוצתי)**

**הגדלת אוסף הנתונים מאפשר למודלים ללמוד מגוון רחב יותר של מקרים ולשפר את דיוק התוצאות**

**התמקדות באזורים עירוניים עם ריכוז גבוה של פשיים ובמקרים המתרכזים בסופי שבוע יכולה לשיע בгибוש אסטרטגיות מניעה מוקדמות.**