**Homework 3 – Business Intelligence**

**40205**

**מרצה: אור יצחק פרץ**

מגישים:

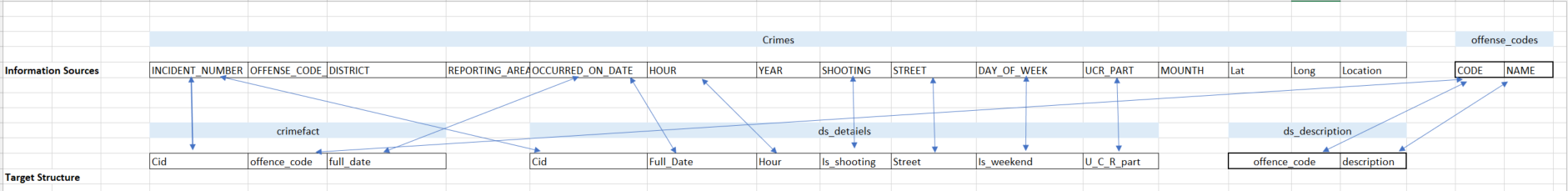
נדב שפינדל 205612468

דור רון 311129936

ענבר מוצרי 315073353

רועי הניגסברג 205387848

**חלק 1 – STTM**

1. בדקנו ואכן תקין.
2. 
3. מצורף גם כקובץ אקסל.

**חלק 2 – Data Mining Techniques**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | OFFENSE\_CODE | YEAR | MONTH | HOUR | Lat | Long |
| mean | 2317.547 | 2016.560586 | 6.609719 | 13.118205 | 42.214381 | -70.908272 |
| std | 1185.286 | 0.996344 | 3.273691 | 6.294205 | 2.159766 | 3.493618 |
| min | 111 | 2015 | 1 | 0 | -1 | -71.178674 |
| q1 | 1001 | 2016 | 4 | 9 | 42.297442 | -71.097135 |
| median | 2907 | 2017 | 7 | 14 | 42.325583 | -71.077524 |
| q3 | 3201 | 2017 | 9 | 18 | 42.348624 | -71.062467 |
| max | 3831 | 2018 | 12 | 23 | 42.395042 | -1 |

1. אנו נבצע תהליך KDD מסוג Data/Pattern Analysis כדי **לחזות (predictive)** אירועי פשיעה במטרה למנוע אותם.
2. הטכניקה שאנו נשתמש בה היא עץ החלטה. הדאטה שלנו מחולק לנתונים קטגוריאליים או מספריים. באמצעות השימוש בעצי החלטה נוכל לבחון אילו מאפיינים משפיעים הכי הרבה על רמת הפשיעות (בהתאם לשאלות המחקר שלנו).  
   תרחיש ראשון – נרצה לדעת לחזות אירועי ירי, לשם כך נבחר את המאפיין shooting כמשתנה שאנו רוצים לחזות, ואת המאפיינים, Full date, HOUR, Location כדי ללמוד אילו מאפיינים חוזים יותר טוב ירי. לדוגמא, נוכל ללמוד כי מתוך 215 אירועי ירי שמצוינים בסט הנתונים שלנו, 169 מהם התרחשו בין השעות 00:00-02:00  
   תרחיש שני – חלוקת סט הנתונים שלנו לפי ימי השבוע, כך נוכל ללמוד האם יש שוני בין פשעים שמתרחשים באמצע השבוע לעומת פשעים בסוף השבוע. לדוגמא, בימים ראשון עד חמישי רב הפשיעות הן קלות ומרחשות באזור מסוים, וסוף השבוע הפשעים מתרחשים באזורים אחרים.
3. מדד הדימיון שבחרנו הינו מדד Jaccard – כיוון שהנתונים שלנו הם בעיקר קטגוריאליים ונרצה לחשב את הדמיון בין הקטגוריות השונות כדי לבחון דמיון בין אירועי הפשיעה השונים.
4. עבור השאלה: בהינתן מיקום, שעה ותאריך האם ניתן לחזות תלונה על פשע הכוללת ירי?  
   0H – למיקום שעה ותאריך אין השפעה על תלונות שכוללות ירי.  
   1H – למיקום שעה ותאריך יש השפעה על תלונות שכוללות ירי.  
   נבחן את כל מקרי הירי שיש לנו, אם נראה שיש הרבה מקרים כוללים ירי בעלי מיקום/שעה/תאריך דומים, נוכל לקבוע כי יש למאפיינים אלו יש השפעה על מקרי ירי

מהם המאפיינים של תלונות בסופי השבוע מול שאר ימות השבוע?  
0H – אין הבדל במאפיינים בין פשעים באמצע שבוע לפשעים בסוף שבוע.  
1H – יש הבדל במאפיינים בין פשעים באמצע שבוע לפשעים בסוף שבוע.

נבחן את כלל הפשעים לפי החלוקה שצוינה לעיל אם נוכל לזהות תבניות מאפיינים על פשעים שמתרחשים באמצע השבוע ו/או בסוף השבוע, כנראה שיש הבדל ויש לדחות את השערת האפס.

**חלק ג'**

1. מהם שמות הרחובות שבהם התרחשו אירועי ירי?

Select Street, count(\*)

from ds\_details

where is\_shooting = ‘Y’, Group by ‘Street’

1. אילו פשיעות בוצעו בסוף שבוע וכמה פעמים?

Select det.is\_weekend, cf.offence\_code,count(\*)

from ds\_details det join crimeFact cf on det.CID=cf.CID

where is\_weekend = ‘Y’

group by IS\_WEEKEND, cf.offence\_code

1. איזה סוגי פשיעות כללו ירי וכמה פעמים?

Select det.is\_shoou, cf.offence\_code,count(\*)

from ds\_details det join crimeFact cf on det.CID=cf.CID

where is\_shooting = ‘Y’

group by IS\_Shooting, cf.offence\_code

1. כמות הפשיעות לכל רחוב, מסודר על פי סדר יורד

Select Street, count(\*)

from ds\_details

group by Street

order by desc

1. עבור כל סוג פשיעה, כמה מתוכם כללו ירי?

Select des.offence\_code, des.description,

SUM(des.offence\_code) OVER (order by det.U\_C\_R\_part) AS shooting\_count

from ds\_description des, ds\_details det, crimeFact cf

where des.offence\_code = cf. offence\_code and cf.CID = det.CID

and is\_shooting = ‘Y’

1. עבור כל סוג פשיעה, מהו כמות התלונות שמתקבלות כל שעה?

Select des.offence\_code, des.description,

SUM(des.offence\_code) OVER (order by det.hour) AS incident\_count

from ds\_description des, ds\_details det, crimeFact cf

where des.offence\_code = cf. offence\_code and cf.CID = det.CID