**Homework 2 – Business Intelligence**

**40205**

**מרצה: אור יצחק פרץ**

מגישים:

נדב שפינדל 205612468

דור רון 311129936

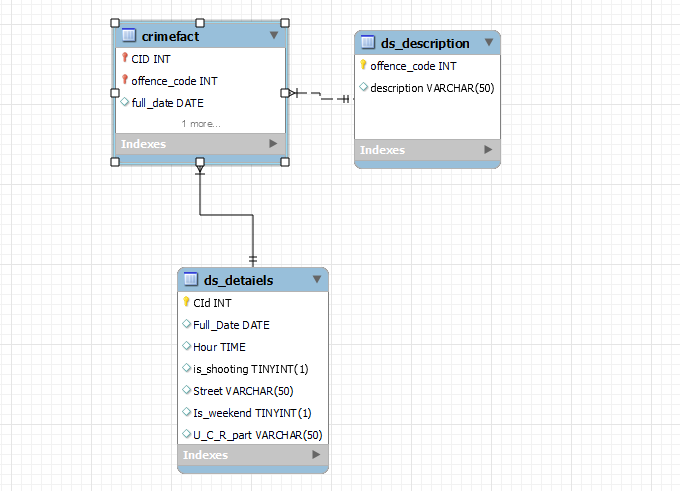
ענבר מוצרי 315073353

רועי הניגסברג 205387848

**חלק 1 – Data Warehouse**

1. בחרנו בסכימת הכוכב כיוון שאין לנו הרבה דאטה, והדרך הפשוטה ביותר להבין את הנתונים שלנו יהיו ע"י בניית טבלה אחת מרכזית (DW) וממנה יצאו כלל הטבלאות המשניות.
2. הData Warehouse שלנו מכיל שלוש טבלאות,  
   א. טבלת ds\_description המכילה את offence code, description המתארת את כל סוגי הפשיעה והתיאור שלהם. [מפתח ראשי offence code].

ב. טבלת ds\_Detaiels המכילה את כל הנתונים הרלוונטיים עבור כל פשיעה, מספקת מידע מעמיק באשר לזמן ולמיקום התרחשות הפשעים האם כללו ירי וכו'. [מפתח ראשי CID]  
ג. טבלת crimeFact – אשר מכילה את המפתחות הראשיים של כל טבלה CID, offence\_code.

****

1. ERD-
2. באמצעות טבלת ds\_Detaiels ניתן לבחון באופן פשוט את הפשיעה בבוסטון לפי רחובות ושעות. כך שבעזרתה ניתן לזהות בקלות מתי באיזורים מסויימים יש בהם אחוזי פשיעה גבוהים יותר ועל כן דורשים תגבור של מערכת אכיפת החוק כדי למנוע פשעים דומים בעתיד.
3. מצורף כקובץ טקסט.

**חלק 2 – ETL**

**1+2**. הגדרת התהליך-

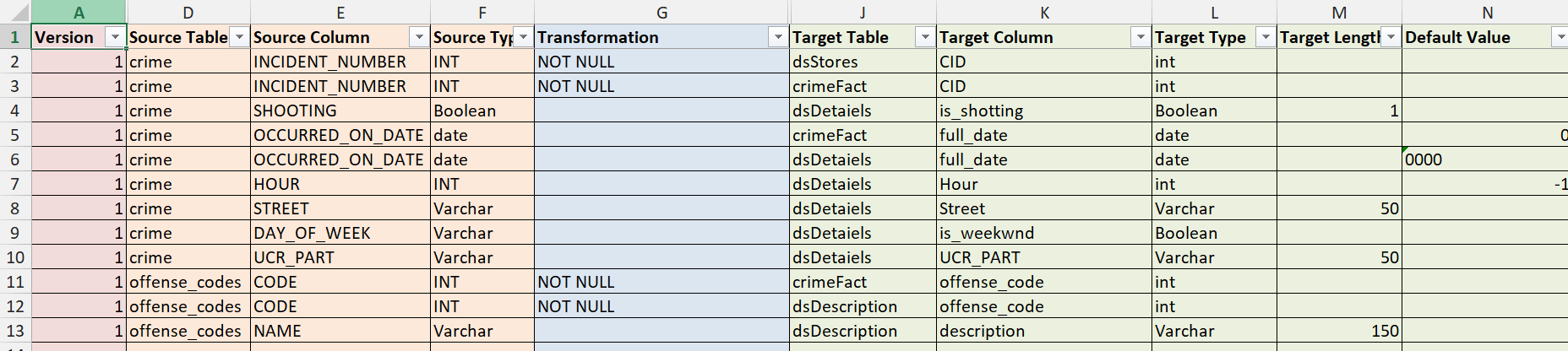
E- נמשוך את המידע מקבצי הcsv את המקור של טבלאת 'offense\_code' נמפה במלואה, יש צורך בקוד של הפשע ובתיאור שלו.

בטבלת 'crime' נמפה רק את העמודות הרלוונטיות, למשל עמודה REPORTING\_AREA לא רלוונטית כי אנו רוצים לצמצמם את הפשעים במחוזות שהאירועים מתרחשים בהם ולא איפה

שדווחו, בנוסף עמודות Month,Year לא רלוונטיות כי ישנה עמודה המאגדת את כל פרמטרים אלו OCCURRED\_ON\_DATE , העמודות המתארות קואורדינטות (Lat/Long/Location) מתארות מיקום- נרצה להגביר אכיפה ברחוב מסוים ולא בנקודה ספציפית. ההמרה מערך הקואורדינטה לרחוב המצוין דורש טרנספורמציה מורכבת ואנו לא צריכים זאת מכיוון שישנה עמודה המתארת את שם הרחוב.

T- בחלק של עיבוד התאריך ניקח את עמודת DAY\_OF\_WEEK ונמיר את עמודה זו לעמודה נומרית האם זה סוף שבוע או לא. נבדוק שאכן הגדרת המשתנים מתבצעת כראוי כמו משתנה שלא יכול להיות ריק אכן לא יעבור לטבלה המשויכת לו אם הוא לא עונה על התנאי.

L- נטען את כל העמודות בטבלאת הfact והdims.

**3.**

**מצורף כקובץ**