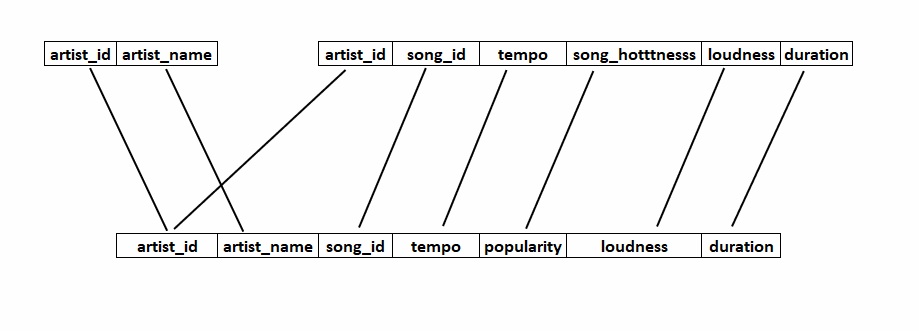
**הגשה 3- בינה עסקית**

**חלק א:**

1. לאחר בדיקה, ההוראות לתהליך ה-ETL הוגדרו היטב.
2. סכמה ויזואלית עבור הSTTM-
3. קובץ מצורף להגשה.

**חלק ב:**

1. עבור בעיית Unsupervised תהליך הKDD שנבצע הוא מסוג Descriptive באמצעות Clustering.

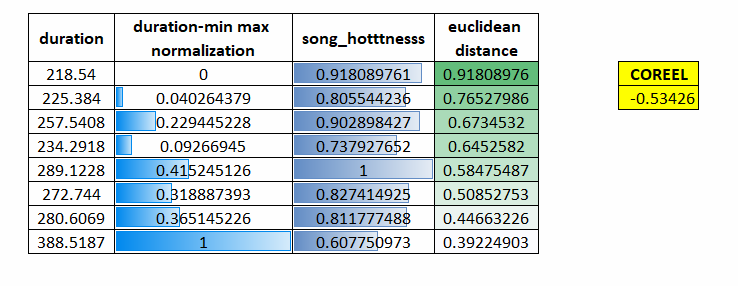
עבור בעיית Supervised תהליך הKDD שנבצע הוא מסוג Predictive באמצעות Regression.

1. א) עבור בעיית Unsupervised העמודות הרלוונטיות הן:

* Loudness, Tempo, Duration- עמודות שמכילות נתונים נומריים רציפים.
* artist\_id, artist\_name- עמודות שמכילות STR.

עבור בעיית Supervised העמודות הרלוונטיות הן:

* popularity - עמודות שמכילות נתונים נומריים רציפים.
* artist\_id, song\_id, artist\_name- עמודות שמכילות STR.

ב) **בבעיית Unsupervised**, בחרנו באקראי את להקת "Black Eyed Peas" וחישבנו עבור כל אחד מהשירים שהוציאו את המרחק האוקלידי בין משך השיר לבין הפופולריות שלו. עולה מן הנתונים כי קיים קשר בין מאפייני שירי הלהקה לבין הפופולריות. הקורלציה שלילית, ככל שמשך השיר נמוך יותר כך השיר צפוי להיות יותר פופולרי.

**בבעיית Supervised,** עבור להקת "Black Eyed Peas" ב Data Setקיימים 8 שירים שמשויכים ללהקה. ממוצע הפופולריות עבור השירים שלהם הוא 0.8264, סטיית התקן 0.1199, ועל כן הרווח סמך הינו [0.9464, 0.7065]. אנו חוזים כי הממוצע הגבוה חוזה כי השירים העתידים של הלהקה צפויים להיות פופולריים גם כן וכי מידת הפופולריות של השיר הבא ייפול ברווח סמך ברמת ביטחון 95%.

ג) מדד הדמיון עבור הDW מסתמך על Similarity Measures תוך שימוש בEuclidean Distance. הנתונים בDW שאנו משתמשים בכדי לקבץ קבוצות הם נתונים נומריים רציפים ולכן ננתח את האשכולות על פי חישוב המרחק האוקלידי.

ד) **השערות מחקר:**

**שאלת Unsupervised:** האם ניתן למצוא מאפיינים דומים בין שירים של אותו אמן? האם אמן מפרסם שירים בעלי אורך דומה? אם כן, מה מאפיין את שירים אלו?

0H - שירים של אומן אינם בעלי מאפיינים דומים.

1H- אחרת, יש תלות בין המאפיינים לפופולריות השיר.

**שאלת Supervised:** האם ניתן לחזות את מידת הפופולריות של שיר חדש של אמן על סמך שירים מהעבר שלו ומידת הפופולריות שלהם?

0H – אין קשר בין פופולריות שירי האומן שהוציא בעבר, לפופולריות של השירים העתידים שיוציא.

1H- אחרת, פופולריות שירי האומן משפיע על מידת הפופולריות של השיר החדש שיוציא.

עבור בדיקת השערות נערוך בדיקה באמצעות המבחנים שכתבנו לעיל, במידה וימצא קשר נקבל את ההשערה ברמת מובהקות של 95%.

**חלק ג:**

1. הצגה של פופולריות שירים ממוצעת לפי אומן.

SELECT A.artist\_id, A.artist\_name, S.popularity,

AVG(S.popularity) OVER (PARTITION BY A.artist\_id) AS "AvgPopularityByArtist"

FROM songs AS S INNER JOIN artists AS A ON S.artist\_id = A.artist\_id

1. הצגה של פופולריות שירים ממוצעת מצטברת לפי אומן.

SELECT A.artist\_id, A.artist\_name, S.popularity,

AVG(S.popularity) OVER (PARTITION BY A.artist\_id ORDER BY A.artist\_id ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS "CumulativeAvgPopularity"

FROM songs AS S INNER JOIN artists AS A ON S.artist\_id = A.artist\_id

1. הצגה עבור כל אמן את מספר השירים שקיבלו מעל 0.8 במדד הפופולריות

SELECT A.artist\_id, A.artist\_name, S.popularity,

COUNT(S.popularity) OVER(PARTITION BY A.artist\_id) as "NumOfPopularSongs"

FROM songs AS S INNER JOIN artists AS A ON S.artist\_id = A.artist\_id

WHERE S.popularity>0.8

1. הצגה עבור כל אמן את המשך הממוצע, הקצב הממוצע והעוצמה הממוצעת של השירים שלו.

SELECT A.artist\_id, A.artist\_name,

AVG(Duration) OVER (PARTITION BY A.artist\_id) AS "AvgDurationByArtist",

AVG(Tempo) OVER (PARTITION BY A.artist\_id) AS "AvgTempoByArtist",

AVG(Loudness) OVER (PARTITION BY A.artist\_id) AS "AvgLoudnessByArtist"

FROM songs AS S INNER JOIN artists AS A ON S.artist\_id = A.artist\_id

1. הצגה של כמות השירים שאמן הוציא כאשר ישנו קשר הפוך בין משך הזמן לקצב והעוצמה (לצורך הבדיקה השתמשנו בערכים משוערים עבור משך נמוך, עוצמה גבוהה וקצב גבוה)

SELECT A.artist\_id, A.artist\_name,

COUNT(\*) OVER(PARTITION BY A.artist\_id) as "NumOfSongs"

FROM BISONGS AS S INNER JOIN BIARTISTS AS A ON S.artist\_id = A.artist\_id

WHERE Duration < 200 AND Tempo > 170 AND Loudness > -4

1. הצגה של האחוזונים של האומנים מסודרים ע"פ משך, קצב ועוצמה.

SELECT A.artist\_id, A.artist\_name,

NTILE(10) OVER (ORDER BY Duration, Tempo, Loudness) AS RankNum

FROM BISONGS AS S INNER JOIN BIARTISTS AS A ON S.artist\_id = A.artist\_id