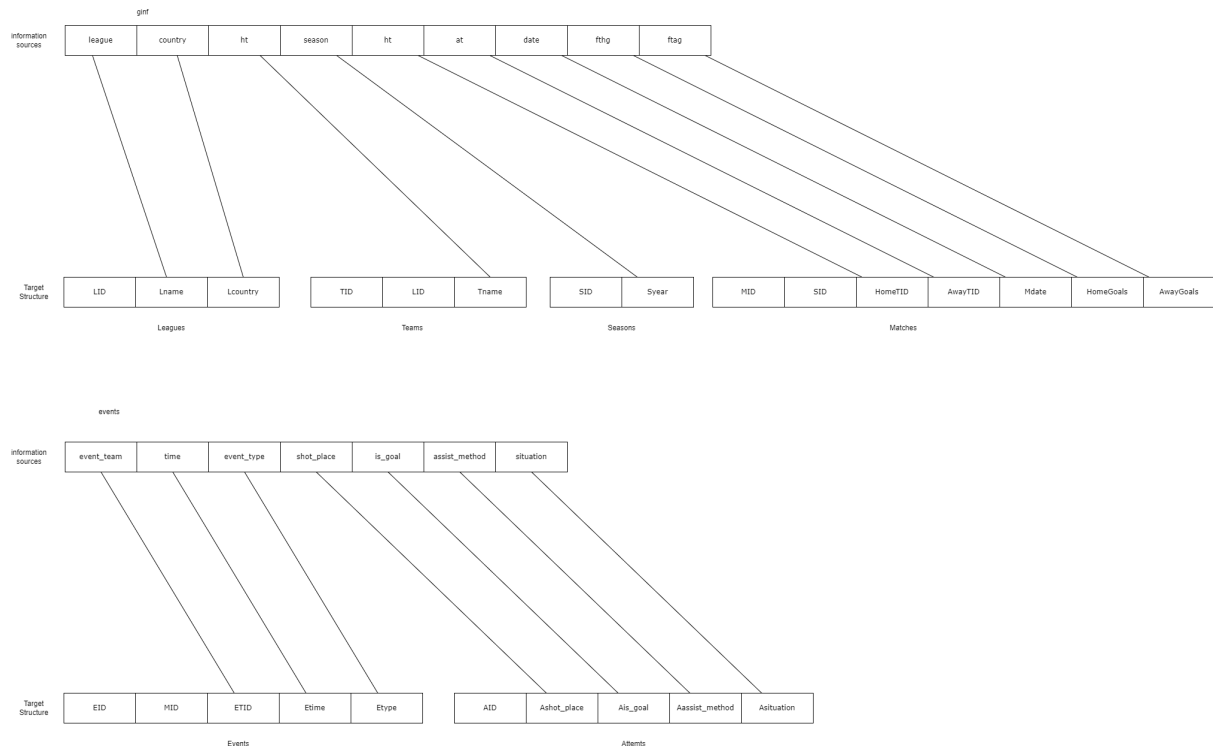


פרויקט בינה עסקית- מטלה 3

חלק 1



מסמך STTM מצורף בקובץ נפרד

חלק 2

1. בתהליך הKDD נבצע ניתוח נתונים לDW כדי לענות על השאלות העסקיות שהגדרנו. בשביל שאלה האם כדאי לבשל באמצעות מסירה עדיף על בישול באמצעות שליחת כדור, נחקור בשיטת Descriptive, באמצעות Clustering. ולשאלה לאיזה קבוצות ריאל מדריד דומה על סמך עבירות, כרטיסים ושערים, נחקור בשיטת Predictive, באמצעות אלגוריתם KNN.
2. בשאלה הראשונה, נשתמש מטבלת Attempts בשדות Ais_goal (בינארי), Assist_method (קטגוריאלי), נשתמש בטכניקת Clustering. נחשב את אחוז הגולים באמצעות בישול במסירה ביחס לכלל הניסיונות בעיטה שנוצרו כתוצאה ממסירה ואת אחוז הגולים באמצעות שליחת כדור ביחס לכלל הניסיונות בעיטה שנוצרו כתוצאה משליחת כדור, אם אחוז הגולים באמצעות מסירה הוא 40% ואחוז הגולים באמצעות שליחת כדור הוא 15% נסיק כי יש יותר סיכוי לגול אם הבישול הוא באמצעות מסירה מאשר באמצעות שליחת כדור.
- בשאלה העסקית השניה נשתמש מטבלת Events בשדות ETID (נומרי), Etype (קטגוריאלי), נשתמש באלגוריתם KNN. נבדוק את הדמיון בין הקבוצות באמצעות אלגוריתם KNN על סמך המדדים עבירות כרטיסים ושערים. נחשב את המרחק האוקלידי בין הקבוצות לריאל מדריד. אם המרחק האוקלידי של ברצלונה מריאל מדריד הוא 2.3 והמרחק האוקלידי של t.o.פ הוא 7.8 נסיק כי ברצלונה יותר דומה לריאל מדריד מאשר t.o.פ.
3. מדד דמיון לDW שלנו הוא Jecard. נוכל למדוד באמצעותו את הדמיון בין ערכים קטגוריאליים Etype (עבירות שבוצעו) ETID (מסמל את ID של הקבוצה שביצעה את העבירה) וכך לראות את הדמיון בין קבוצות מבחינת עבירות.

4.שאלה עסקית ראשונה:

H0: אין הבדל בין בישול באמצעות מסירה לבישול באמצעות שליחת כדור

H1: יש הבדל בין בישול באמצעות מסירה לבישול באמצעות שליחת כדור

נחליט לקבל את ההשערה באמצעות מבחן t.

שאלה עסקית שניה:

H0: אין דמיון בין קבוצת ריאל מדריד לקבוצות אחרות מבחינת עבירות

H1: יש דמיון בין קבוצת ריאל מדריד לקבוצות אחרות מבחינת עבירות

נחליט לקבל את ההשערה באמצעות מבחן אנובה לקבוצות.

חלק 3

עבור שאלת מחקר 1:

כמות הפעמים שהכניסו גול באמצעות בישול מסוים חלקי כמות הפעמים שניסו להבקיע באמצעות אותו בישול(עם גול ובלי גול)

```
SELECT Aassist_method, COUNT(AID) as numberOfAttempts, SUM(Ais_goal) as  
,numberOfGoals  
CAST (SUM(Ais_goal) AS float)/CAST (COUNT(AID) AS float))*100  
as Percentage  
FROM Attempts  
;GROUP BY Aassist_method
```

כמות הפעמים שהכניסו גול באמצעות בישול מסוים חלקי כמות הפעמים שניסו להבקיע באמצעות אותו בישול (עם גול ובלי גול) בקבוצה מסוימת.

```
SELECT ETID ,Aassist_method, COUNT(AID) as numberOfAttempts,  
,SUM(Ais_goal) as numberOfGoals  
CAST (SUM(Ais_goal) AS float)/CAST (COUNT(AID) AS float))*100  
as Percentage  
FROM Attempts, Events  
ON Attempts.AID = Events.EID  
WHERE ETID = 2  
;GROUP BY Aassist_method
```

כמות הפעמים שהכניסו גול באמצעות בישול מסוים חלקי כמות הפעמים שניסו להבקיע באמצעות אותו בישול בטווח דקות מסוים במשחק.

```
SELECT Aassist_method, Etime, COUNT(AID) as numberOfAttempts,  
,SUM(Ais_goal) as numberOfGoals  
CAST (SUM(Ais_goal) AS float)/CAST (COUNT(AID) AS float))*100  
as Percentage
```

```
FROM Attempts, Events
ON Attempts.AID = Events.EID
WHERE Etime > 46
;GROUP BY Aassist_method
```

עבור שאלת מחקר 2:

group by ע"י ID קבוצה שביצעה את האירוע וסכימת האירועים.

```
(SELECT DISTINCT ETID, Etype, COUNT(Etype
OVER(PARTITION BY ETID, Etype) as TotalEventsPerTeam
;FROM Events
```

```
(SELECT DISTINCT ETID, Etype, COUNT(Etype
OVER(PARTITION BY ETID, Etype) as TotalEventsPerTeam
FROM Events
;WHERE ETID = 1
```

```
(SELECT ETID, Sum(Ais_goal
FROM Events
LEFT JOIN Attempts
ON Attempts.AID = Events.EID
;GROUP BY ETID
```

קישור ל-GitHub:

<https://github.com/OrHaru/BICourseProject>