



# הקשר בין ריצות באינטנסיביות גבוהה של שחקני כדורגל להצלחת הקבוצה

פרויקט מחקר

הוגש במסגרת קורס מספר 02360502 - "פרויקט בבינה מלאכותית"

**עידו טלקר, אפק בן סימן  
בהנחיית דין צדוק**

הוגש בתאריך \_\_\_\_\_ לפרופסור שאול מרקוביץ'

## תוכן עניינים

4	1 מבוא
6	2 הגדרת הבעיה ומטרות הפרויקט
6	2.1 שאלת המחקר
6	2.2 מטרות הפרויקט
7	2.3 הגבלות
8	3 מידול הצלחת הקבוצה
8	Expected Goals xG 3.1
11	Expected Points xPts 3.2
13	4 מאגרי המידע
13	STATSport APEX 4.1
13	OPTA Stats 4.2
14	SofaScore 4.3
14	4.4 הקלטות ידאו רשםיות של המשחקים
15	5 הפיצרים
15	zone_5_distance_x 5.1
15	zone_5_time_x 5.2
15	zone_6_distance_x 5.3
16	zone_6_time_x 5.4
16	accelerations_x, decelerations_x 5.5
16	possession 5.6
17	5.7 התפלגות הפיצרים למול Gx
20	6 מתודולוגיה
20	6.1 תיאור המערכת והאלגוריתמיקה לפתורן הבעיה
20	תרשים כללי:
20	6.2 יצירת ויזואלייזציה
21	6.3 יצירת קובץ metadata
22	6.4 הצבת מקורות מידע ויצירת data frame
23	6.5 בחינת הקשרים ע"י שיטות סטטיסטיות, ליניאריות ולא ליניאריות
23	6.6 אימון מודלי חיזוי והשוואת ביצועים
24	7 ניסויים
24	7.1 אקספלורציות סטטיסטיות למול מדד Gx
26	7.1.1 בחינת מדד ההונאות Mutual Information
27	7.2 אקספלורציות סטטיסטיות למול מדד sxt
28	7.3 אקספלורציות סטטיסטיות למול עוצמת הקבוצה היריבה
36	7.4 אימון מודלי רגסיה לחיזוי Gx ו-xPt
37	7.4.1 רגסטור לינארי
38	7.4.2 יער רנדומי (Random Forest)
40	8 מסקנות
41	9 כיוונים להמשך
42	10 הצהרה בנוגע לשימוש במודלי שפה

**11 תודות**  
**12ביבליוגרפיה**

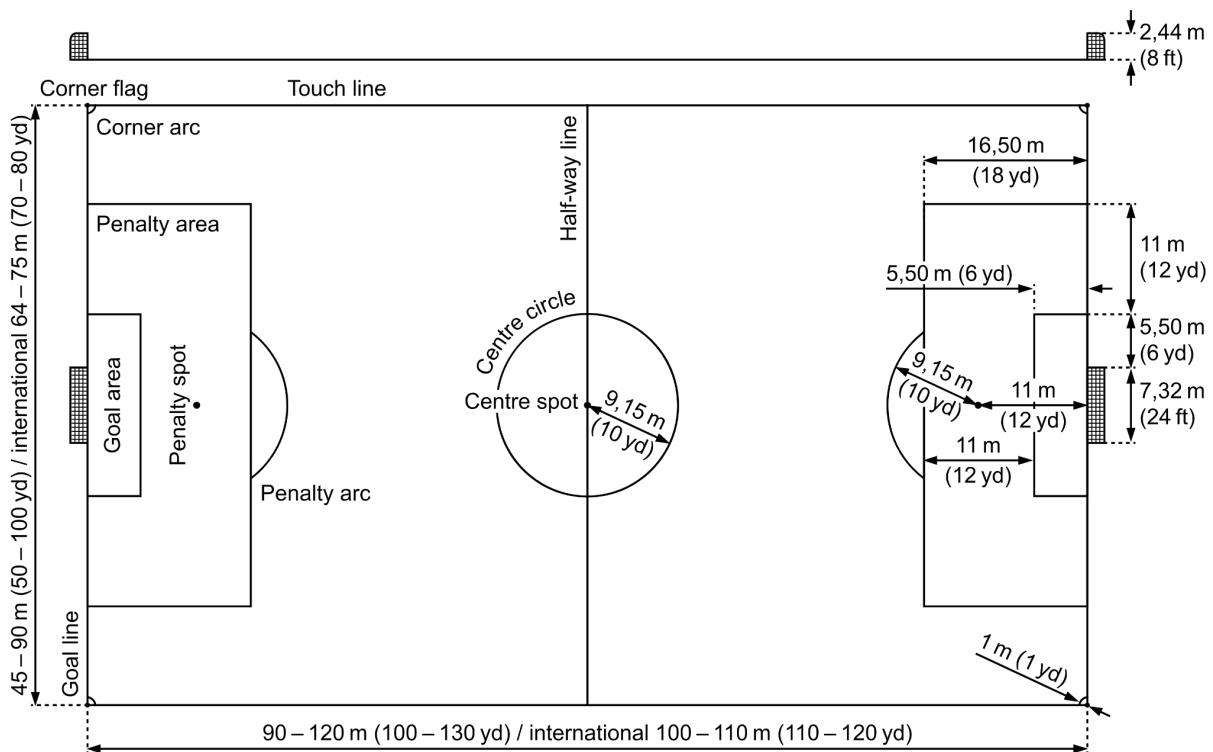
**42**

**43**

# 1 מבוא

כדורגל הוא משחק כדור קבוצתי הנחשב לענף הספורט הפופולרי ביותר בעולם ומושוכק ביותר מ-200 מדינות. המשחק מתנהל בין שתי קבוצות שכל אחת מהן מונה 11 שחקנים: 10 שחקני שדה שיכולים לגעת בכדור יחד שנמצא במגרש בכל חלק מגוף מלבד היד ושהקן ייחודי בתפקיד שוער שיכול לגעת בכדור בידו באוזור מוגדר שנקרא "רחבה" ותפקידו למנוע מהקבוצה היריבה להבקיע שערים.

המשחק משוכק במגרש מלכני כאשר ברמה המקצועית תקן מידותיו הוא כ- 100 מטרים לאורך וכ-70 מטרים לרוחב. מטרת המשחק היא להבקיע כמה שיותר שערים להכניס את הכדור לשער היריב שהוא מסגרת מתכת במידות 7.32 מטרים לרוחב על 2.44 מטרים לגובה. שנו שער אחד לכל קבוצה בצדדים המנוגדים של המגרש, על מרכז הצלע הקצרה. אורך משחק עומד על שתי מחציות של 45 דקות כל אחת ועוד תוספת זמן של דקות בודדות לכל מחצית שננתונה לדעת השופט.



מידות מגרש כדורגל תקני

כדורגל הוא ענף ספורט פיזי וטקטטי, המשלב מאיצים אנאירוביים ואירוביים לסירוגין. השחקנים נדרשים לבצע ריצות קצרות ומהירות (ספרינטים), שינוי כיוון חדים, מאבקים פיזיים ומעברים מהירים בין הגנה להתקפה תוך תיאום גובה ביניהם, ברמה טכנית גבוהה ובמהירות מחשבתית גבוהה על מנת להציגו בענף.

ברמות התחרויות הגבוהות ביותר, קבוצות כדורגל משקיעות סכומי עתק, לעיתים מילוני דולרים בשנה, במטרה למקם את ביצועי השחקנים והקבוצה כולה. השקעות אלו אינן מתמקדות אך ורק ברכש שחקנים או בוצאות מכספי, אלא גם בפיתוח תשתיות אנליטיות מתקדמות המאפשרות קבלת החלטות מבוססת-נתונים (data-driven decision making). בשנים האחרונות, חלה התקדמות משמעותית בתחום האנליה הספורטיבית, המאפשרת לחוקרים, למאנים ולאנליסטים להבין לעומק את הקשרים בין מדדים פיזיולוגיים, טקטיים ובוימכניים לבין תוצאות המשחק בפועל.

קבוצות מ Każויות עושות שימוש מגוון רחב של כלים טכנולוגיים הכוללים מערכות GPS, חיישנים ביומטריים, ניתוח וידאו, ובינה מלאכותית. טכנולוגיות אלו מאפשרות ניתוח מדויק של עומס עבודה, מהירות, תאוצה, טווח תנעה ומדדים נוספים, לצד ניתוח טקטי של דפוסי משחק והבנת דינמיקות קבוציות. הנתונים הנאספים מתרגמים לתובנות אופרטיביות המסייעות בהתאמת תוכניות האימון, בשיפור קבלת ההחלטה של מאנים ובמונעת פציעות הנובעות מעומס יתר. השקה זו כוללת גם גיוס של אנשי מקצוע ייעודיים כגון אנליסטים, מדען נתונים ומהנדס ביצועים, אשר תפקידם לנתח ולפרש את הנתונים לצורכי יישום מעשי בשטח.

דוגמה בולטת ליישום טכנולוגיות אלו היא מערכת **APEX** של חברת **STATSports**, הנחשבת לאחת המובילות בעולם בתחום ניתוח הביצועים הספורטיביים. מערכת זו מבוססת על טכנולוגיית GPS ברמת דיוק גבוהה, ומאפשרת מדידה רציפה של מגוון פרמטרים פיזיים במהלך אימונים ומשחקים. הנתונים מועברים בזמן אמת למערכת ניהול וניתוח מתקדמת, המספקת לצוות המקצועני תמונה מקייפה של מצבו הפיזי והביצועי של כל שחקן. בכך, המערכת מאפשרת ניתור עומסים, מניעת פציעות והתאמאה אישית של תוכניות אימון בהתאם לנדרונות האינדיידואליים של כל שחקן. כיום נעשה שימוש נרחב במערכת **STATSports Apex** במועדוני עילית באירופה, בהם: ליברפול, מנצ'סטר יונייטד, יובנטוס וgam בישראל קבוצות רבות רוכשות את המוצר הנ"ל וביניהן גם מועדון הכדורגל מכבי חיפה. המערכת המספקת לקבוצות היא למעשה צייף לביש שאוסף את המדדים השונים (שיפורטו בהמשך) לכל שחקן, במהלך כל משחק או אימון.



**מערכת APEX STATSports**, משאאל הצ'יפ, באמצעות הפריט הלביש אליו מוכנס הצ'יפ, מימין מערכת הבדיקה

## **2 הגדרת הבעיה ומטרות הפרויקט**

### **2.1 שאלת המחקר**

נרצה לבחון את הקשר בין המאמץ הפיזי של שחקי ני כדורגל, בדגש על ריצות באינטנסיביות גבוהה לבין הצלחת הקבוצה.

כלומר - האם קיים קשר מובהק בין עצימות הריצות של שחקי ני מכבי חיפה, כפי שהוא נמדדת באמצעות מדדים כגון: מהירות שיא, מספר ריצות בעצימות גבוהה ומרחקים מצטברים, לבין תוצאות הקבוצה?

בנוסף, המחקר בוחן האם לרמות העצימות הפיזית של שחקי ני מכבי חיפה יש השפעה על ביצועי היריבה, למשל במצבם היכולת התקפית שלה.

לפי הספרות הפיזיולוגית, ריצות באופן כללי מחולקות ל-6 איזורי מהירות (zones) ובפרט ריצות באינטנסיביות גבוהה מוגדרות לאיזורים 5 ו-6 כאשר:

איזור 5 = ריצה ב מהירות של בין 19.8 קמ"ש ל-25 קמ"ש.

איזור 6 = ריצה ב מהירות של מעל 25 קמ"ש.

שאלת המחקר הועלתה ע"י מר אורן הראל, מאמן הכשר הראשי של מועדון הcadogal מכבי חיפה ומומחה לפיזיולוגיה.

### **2.2 מטרות הפרויקט**

כיום ריצות באינטנסיביות גבוהה נתפסות כאחד הגורמים המשמעותיים ביותר ביכולת לייצר מצבים בקבוקה, אך עדין קיים חוסר בהירות מדעי לגבי הקשר ההפוך בין השניים. הבנת קשר זה עשויה לשיער לצוותים מקצועיים לאוזן בין עומס אימון, לבין דפוסים התורמים ליעילות גבוהה יותר בזמן משחק ולספר תהליכי קבלת החלטות בקבוצות מקצועיות בזמן משחק ובזמן אימון. בנוסף לכך, הבנת קיום הקשר המתואר וחזקתו עשויה לשמש הוכחת היתכנות לפיתוח מודלי חיזוי מבוסס למידת מכונה בתחום ניתוח הביצועים הספורטיביים, ולSHIPOR מודלים קיימים. לסייעים המחקר שואף:

1. לזהות האם קיים מתאם מובהק סטטיסטי בין עצימות ריצות (מהירות שיא, מספר ריצות בעצימות גבוהה, מרחק מצטבר) לבין מדדי הצלחה קבוצתיים.
2. לבחון האם עצימות הריצה של מכבי חיפה משפיעה על הביצועים התקפיים או ההגנתיים של היריבה.
3. לבנות בסיס לפיתוח מודלי חיזוי מבוסס למידת מכונה, אשר יאפשרו לחזות את תרומות העצימות הפיזית לתוצאות עתידיות וליעילות משחקית.

### **2.3 הגבלות**

הנתונים נאספו ממשחקי ליגת העל ומשחקי הגביע בשתי עונות סדרות של מכבי חיפה בלבד (עונה 24-25, 23-24).  
כתוצאה לכך התוצאות שיפורטו בהמשך נוטות להתאמה אל שיטת המשחק ואופיים של השחקנים ששיחקו בקבוצה בעונות אותן בחרנו.

כפי שצוין קודם לכן, הנתונים הפיזיולוגים נאספו באמצעות מדידות GPS של מערכת STAT Sport APEX ותכלנו שגיאות נתונים אלה וביתר שאת מכיוון שבמהלך שתי העונות המذוברות, באיזורים שונים במדינת ישראל הופעלו משבשי GPS משיקולים בטחוניים.  
במהלך שלב ניקוי הנתונים ניקנו למיטב הבנתנו את כל השגיאות הללו אך יש להכיר בכך שיתכן ולא כולם נוקו.

לבסוף, כאשר מבצעים ניתוח סטטיסטי על משחק כדורגל, למעשה מזנחים פרמטרים אקראיים שעלו לחשיפה על תוצאת המשחק, מקרים: נוכחות אחדי הקבוצה או אחדי היריבה ורמת העידוד שלהם במהלך, תנאי מגז אויר משתנים, תנאי המגרש המשתנים, טוויות שיפוט וכדומה. למעשה כל ניתוח מדעי סביר נוטה להזניח גורמי רעש שונים כיוון שתת חלקם קשה עד לא ניתן למדוד או למדוד, וחילוקם האخر מייצר רעש גדול מדי נתונים הנבדקים. כתוצאה לכך יתכן שתוצאות המוצגות במסגרת פרויקט מחקרי זה לא יעדמו בהלימה מושלמת עם התוצאות בפועל בהינתן ולידציה על עונות משחק נוספות.

### **3 מידול הצלחת הקבוצה**

ישנן דרכים שונות למדוד הצלחה של קבוצת כדורגל, למשל: ניתן להסתכל על תוצאות יירות, כלומר מספר הנצחות/תיקו/הפסדים במהלך העונה.

דרך אחרת היא להסתכל על מספר הנקודות שהקבוצה השיגה במהלך העונה (כאשר נצחות מעניק 3 נקודות, תיקו מעניק נקודה אחת והפסד לא מעניק נקודות כלל).

דרך נוספת היא מספר השערים שהובקו על ידי הקבוצה במסגרת המשחקים שלה.

לעומת המددים הטריוויאליים הנ"ל, בחרנו לבחון 2 מددים שונים, מבוססים ניתוח סטטיסטי של מצב הביקעת שערים במהלך משחק כדורגל. הבחירה בממדים אלה נבעה מכיוון שהם מהווים ממדים מהימנים ולא רועשים במיוחד: אינם מושפעים מאיכות יכולת ההבקעה של השחקנים (שעשויہ ליצור יותר/פחות שערים בפועל) ובכך מונטראלים אלמנטים של: מזל, טעויות לא מחויבות, ביצועים אינדיבידואליים של שחקני הקבוצה או של השוער היריב.

נדגיש כי בחירה זו אינה בחירה נפוצה ולא מצאנו מחקר אחר שעשה את ההשוואה הזה קודם לכן.

#### **Expected Goals xG 3.1**

מדד שערים צפויים (Expected Goals=Gx). הוא מדד המעריך את סיכוי ההצלחה להבקיע שער מכל ניסיון ההבקעה במהלך משחק, תוך התחשבות בגורמים כמו מיקום ניסיון ההבקעה, סוג המסירה שהובילה לניסיון ההבקעה, שחקני יריים מסביב לכדור בזמן ניסיון ההבקעה, איבר הגוף שאיתו ביצע השחקן את ניסיון ההבקעה ועוד. המدد מעניק לכל ניסיון ההבקעה ציון הסתברותי בין 0 (שום סיכוי להבקעת שער) ל-1 (שער ודאי).<sup>(1)</sup>

ה-Gx מספק תובנות לגבי עילוות התקפה והאם קבוצה או שחקן מנצלים הזדמנויות כפי שצפו.

מדד ה-Gx מתימר למדל באופן סטטיסטי כמה שערים קבוצה או שחקן היו צפויים להבקיע כפונקציה של מספר ואיכות ניסיונות ההבקעה שהתרחשו. מספק מספר דוגמאות להמחשת המدد:

בשנת 2020 לאו מס' הנקשב לאחד משחקני הכדורגל הגדולים בכל הזמנים, שיחק בקבוצה הגדולה הספרדית ברצלונה. עת נציג את כל המצבים שייצר במסגרת 2 משחקים עוקבים שבהם שיחק כאשר כל מצב מתואג לפי המיקום במשחק ממנו נבעט הכדור, סוג המהלך או האיבר בגוף שהוביל במצב, התוצאה בפועל של המצב ומה ערך ה-Gx שהוא שוחש עבורה.

נבחן כי ערך ה- $xG$  גבוהה יותר ככל שמצב קרוב יותר לשער, ומהלך שMOVIL אליו כולל בתוכו פחות שחקני יריבה (לדוגמה: פנדל שבשרטוט הוא ריבוע אדום בולט עבورو ערך ה- $xG$  המתקבל הוא הגבוה ביותר בסדר גול וערך 0.74).

## Lionel Messi

Barcelona

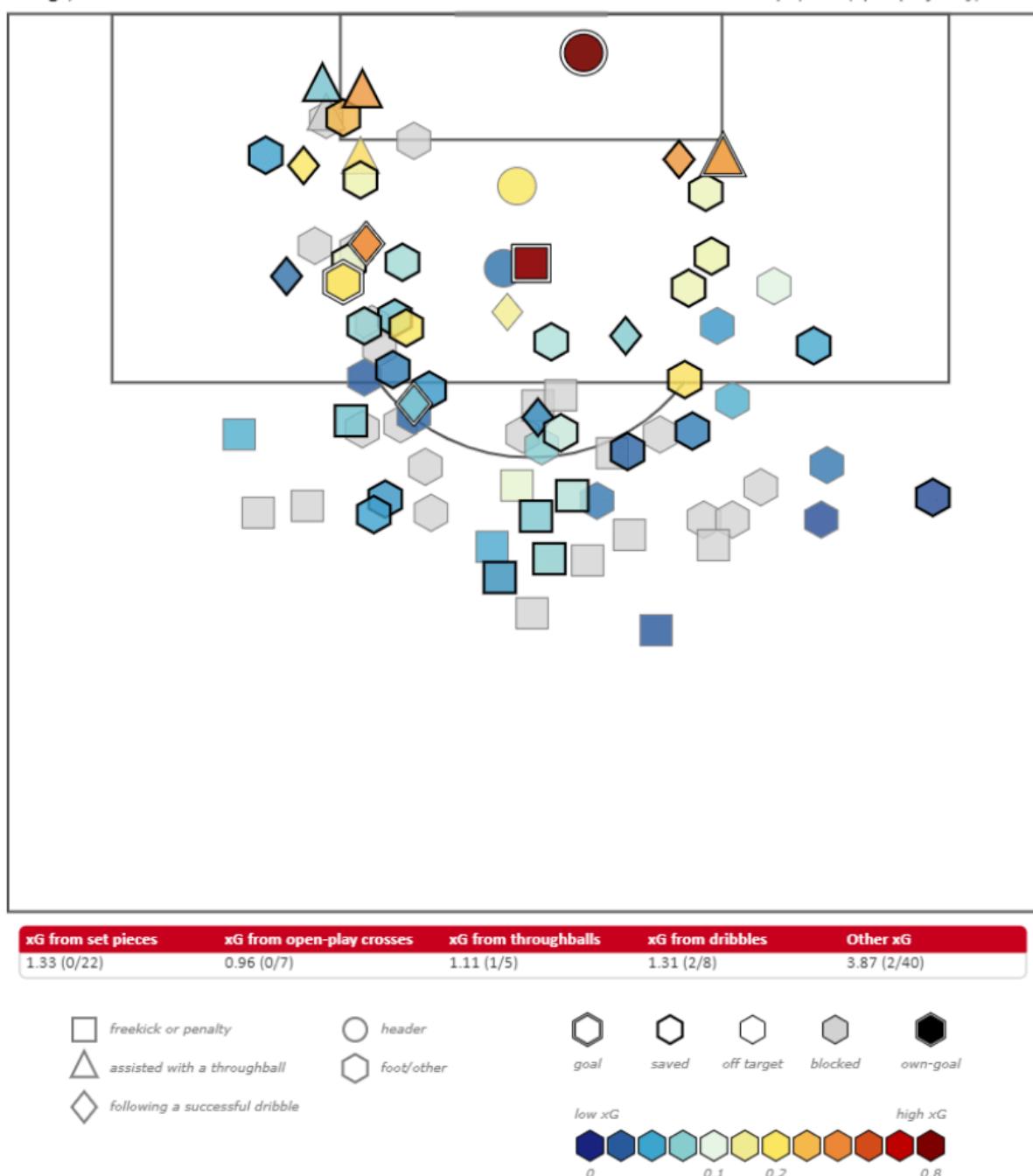
La Liga, 2020-09-01 - 2020-12-23

**Exp. goals: 8.49 (5 goals / 81 shots)**

+2.28 (2 goals / 3 penalties)

ExpG/shot (all except pens & OG): 0.105

ExpG/shot (open play only): 0.121



מצבי ההבקעה שנרשמו ב-2 משחקים עוקבים ע"י לאו מס' שחקן ברצלונה, הצורה מעידה על כיצד נסיוון ההבקעה (איבר בגוף / סוג המהלך), המסגרת מעידה על התוצאה בפועל (gól, הצלחה, חסימה, בעיטה שאינה במסגרת שער היריב), הצביע מסמן את גודל ה- $xG$  שחושב עבור אותו מצב הבקעה.

דוגמה נוספת הממחישה כיצד ניתן להשתמש במידד זה כדי להשווות ביצועי קבוצות הינו כאשר מחשבים את ה- $xG$  המצטבר של הקבוצה במשחק. כך מקבלים אינדיקציה לאיכות ההזדמנויות שיצרה הקבוצה. השוואת ממד זה למספר השערים בפועל מאפשרת לנתח את איכות הגימור (Finishing) של הקבוצה:

- **מצד אחד: יצירת הזדמנויות גבוהה, מימוש נמוך:**
  - $xG$  מצטבר של 2.8 לעומת שער אחד שהובקע בפועל.
  - המשמעות: הקבוצה יצרה הזדמנויות הביקעה איכותיות רבות (שבממוצע אמרות היו להניב כמעט 3 שערים), אך לא הצליחה לתרגם אותן לשערים. זהה עדות לגימור חלש של השחקנים במהלך משחק.
- **מצד שני: יצירת הזדמנויות נמוכה, מימוש גבוהה:**
  - $xG$  מצטבר של 0.2 לעומת שני שערים שהובקעו בפועל.
  - המשמעות: הקבוצה יצרה מעט מאוד הזדמנויות איכותיות, אך השחקנים הצליחו אותן ב.openg חרג ויצא דופן (למשל, באמצעות בעיטות מרוחק שהפתיעו את השוער). זהה עדות לגימור מצוין, ככלומר שהשחקנים עשו ברמה גבוהה בהרבה מהממוצע הנוכחי עליו המידד מבoso.

ברמה הפרטנית, השתמש ב- $xG$  כדי להציג בקצרה השוואת ביצועים של שחקני עילית בעונת המשחקים 2022-2023 הנтונה במקור (1):



אוסף ממצבי ההבקעה שbowcouro ע"י שני השחקנים, נקודות באדום מסמינות מצבים שנכנסו לשער, גודל כל נקודה מעיד על גודל ה- $xG$  עבור אותו מצב.

נבחן כי שני השחקנים עשו מספר זהה של 63 בעיטות במהלך עונת המשחקים.

ניתוח השחקן השמאלי מצבע על  $xG$  מצטבר גבוה (10.2) ומיד כי הבעיות שלו נבעטו ברובם ממצבים נוחים ואיכותיים יותר (בין היתר מרחק קצר יותר לשער). הוא הבקיע 12 שערים כאשר צפוי ההבקעה היה 2.8.

מסקנה: הגימור שלו היה טוב מהממוצע הנוכחי (הבקיע יותר ממה שחוזה המידד).

ניתוח השחקן הימני:  $Gx$  מצבבר נמוך (2.8) מעיד כי הביעות שלו נבעטו ברובן ממצבים קשים יותר (בין היתר מטווחים ארוכים). הוא הוביל 3 שערים כאשר צפוי ההבוקעה היה 2.8.

מסקנה: גם הגימור שלו היה מעט טוב מהממוצע המוצעני (הוביל יותר ממה שחזזה המדד), אך נראה שההבדנויות שלו היו פחות טובות באופן משמעותי דבר שיכול להוביל על יכולות גימור גבוהות או מנגד, יכולות מיקום ויצירת מצבים פחותות לעומת השחקן הראשון.

לאחר פגישה והציג התוצאות לצוותים המוצעים במכבי חיפה, הוכנו לנסות לחזות את מדד  $xPts$  בנוסף ל- $Gx$  שכן מדד ה- $Gx$  אינו לוקח בחשבון את הקבוצה היריבה ועלול לייצר תמונה מצב שאינה שלמה:  $Gx$  מודד את צפוי הפרודקטיביות התקפית של קבוצה בודדת ולאו דווקא משליך על תוצאה המשחק כולה.

### Expected Points xPts 3.2

מדד הנקודות הצפויות ( $xPts$ ) - מדד סטטיסטי אשר ממדל את מספר הנקודות שכל קבוצה הייתה צפואה להישג בכל משחק.

מדד זה מבוסס על מדד השערים הצפויים ( $Gx$ ) של כל קבוצה בנפרד. נציג כי מדד זה משמש חברות מסחריות כמו צחמן שהן מציאות לקבוצות כדורגל שונות ולכן אין מפרטנות את דרך חישובו. לכן, במסגרת מחקר זה פיתחנו שיטת חישוב עבורי באופן הבא:

בנייה סימולטור המבוסס על שיטת מונטה קרלו המחשב את מדד  $xPts$  כך -

כל בעיטה שקיבלה  $Gx$  מסוים, נגריל את הסתברות הביקעת השער כאמור מקרי ברנולי עם  $Gx=k$  (בין 0 ל-1) להבקעת שער. כך הסימולטור מגറיל הסתברות שער לכל הממצבים של שתי הקבוצות.

לבסוף, הסימולטור קובע את תוצאה המשחק לפי הניקוד שהתקבל בסימולציה: אם לשתי הקבוצות בסיום הגמרה אותו מספר שערים - התוצאה היא תיקו. אחרת, הקבוצה שלא שער ניצחה והשנייה הפסידה.

כל משחק מסומלך כנ"ל 10,000 פעמים תוך שמירת מספר הפעמים שכל קבוצה ניצחה או שהמשחק הסתיים בתיקו. לבסוף, מספר הנקודות שכל קבוצה מקבלת מחושב ע"י תוחלת הנקודות מכל הగזרות (כפי שציינו, ניחסן מזכה ב-3 נקודות, תיקו ב-1, הפסד ב-0):

$$(\#wins * 3 + \#draws * 1) / 10,000$$

הסימולטור השיג תוצאות קרובות מאוד לტבלת האמת עבור עונות 23-24, 24-25 ו-25-26 שעליהן בוצעו כל הניסויים, הן במקומי הקבוצות והן במספר הנקודות לקבוצה.

Team	Model	Games	Total_xPts	xPts_per_Game	Actual_Pts	Pts_Diff	Actual_Pos	Pred_Pos	Pos_Diff
Maccabi Tel Aviv	Binomial	36	78.8437	2.190103	85	6.1563	1	1	0
Maccabi Haifa	Binomial	36	71.1755	1.977097	74	2.8245	2	2	0
Hapoel Be'er Sheva	Binomial	36	58.4736	1.624267	61	2.5264	3	3	0
Hapoel Haifa	Binomial	36	47.1048	1.308467	59	11.8952	4	5	-1
Maccabi Bney Reine	Binomial	36	42.8190	1.189417	44	1.1810	5	7	-2
Bnei Sakhnin	Binomial	36	44.0929	1.224803	44	-0.0929	6	6	0
Hapoel Jerusalem	Binomial	33	41.9060	1.269879	43	1.0940	7	10	-3
Maccabi Petach Tikva	Binomial	33	34.5898	1.048176	40	5.4102	8	13	-5
Maccabi Netanya	Binomial	33	51.2309	1.552452	38	-13.2309	9	4	5
Ashdod SC	Binomial	33	36.9597	1.119991	37	0.0403	10	11	-1
Beitar Jerusalem	Binomial	33	42.1840	1.278303	36	-6.1840	11	9	2
Hapoel Hadera	Binomial	33	36.3904	1.102739	36	-0.3904	12	12	0
Hapoel Tel Aviv	Binomial	33	42.6159	1.291391	33	-9.6159	13	8	5
Hapoel Petach Tikva	Binomial	33	33.7588	1.022994	24	-9.7588	14	14	0

תוצאות הסימולטור עבור עונת המשחקים 2023/24 - ניתן לראות את הPUR את המיקום המקורי של כל קבוצה בפועל, ואת הPUR את המיקום המשוער לפי מודל הדפס אל מול המיקום בפועל של כל קבוצה.

Team	Model	Games	Total_xPts	xPts_per_Game	Actual_Pts	Pts_Diff	Actual_Pos	Pred_Pos	Pos_Diff
Maccabi Tel Aviv	Binomial	35	75.4330	2.155229	80	4.5670	1	1	0
Hapoel Be'er Sheva	Binomial	35	67.7699	1.936283	78	10.2301	2	2	0
Maccabi Haifa	Binomial	35	63.8438	1.824109	61	-2.8438	3	3	0
Beitar Jerusalem	Binomial	36	59.1850	1.644028	53	-6.1850	4	4	0
Hapoel Haifa	Binomial	36	48.6273	1.350758	52	3.3727	5	6	-1
Maccabi Netanya	Binomial	36	45.0104	1.250289	45	-0.0104	6	7	-1
Hapoel Jerusalem	Binomial	33	39.4034	1.194042	44	4.5966	7	8	-1
Maccabi Bney Reine	Binomial	33	50.1840	1.520727	41	-9.1840	8	5	3
Hapoel Ironi Kiryat Shmona	Binomial	33	29.1042	0.881945	37	7.8958	9	14	-5
Bnei Sakhnin	Binomial	32	29.8421	0.932566	36	6.1579	10	13	-3
Ashdod SC	Binomial	33	37.0181	1.121761	35	-2.0181	11	11	0
Ironi Tiberias	Binomial	33	35.0556	1.062291	35	-0.0556	12	12	0
Maccabi Petach Tikva	Binomial	33	38.6282	1.170552	33	-5.6282	13	9	4
Hapoel Hadera	Binomial	33	38.0929	1.154330	27	-11.0929	14	10	4

תוצאות הסימולטור עבור עונת המשחקים 2024/25

## 4 מאגרי המידע

לצורך מענה על שאלת המחקר השתמשנו ב-4 מאגרי מידע שונים המתכליים את משחקן מכבי חיפה ב-2 עונות משחקים: עונת 23-24 ועונה 24-25. המאגרים:

### STATSport APEX 4.1

מידע פיזיולוגי שmagiu מגופיות ספורט שחקני מכבי חיפה לובשים בזמן המשחק. גופיות אלו מנטרות בתדרות של מאות שנייה: מהירות אבסולוטית, מיקום GPS ותאוצה זוויתית ב-3 צירים עבור כל שחקן. הנ"ל מידע שימוש את צוות האימון, סופק לנו ע"י מר אורי הראל ונמדד ע"י מכשיר לביש על גופו של השחקן. מידע זה התקבל בפורמט csv ששימש לבנית פיצ'רים עבור ה-data frame בו השתמשנו. קיבלנו דגימות אלו בעבר 62 משחקים שונים מתוך משחק ליגת העל ומשחקי הגביע של מכבי חיפה כאשר בכל משחק, קיבלנו את כל המידע הפיזיולוגי של שחקני הסגל, כולל שחקני הספסל שמתחלפים במהלך המשחק עם 11 השחקנים שהתחילו אותו. (モותר להחליף עד 5 שחקנים במהלך המשחק, בכל זמן שהוא). סה"כ: כ-18 קבצי csv לכל משחק, כל אחד מהם במשקל של כ-1 GB.

1	Time	Lat	Lon	Speed (m/s)	Heart Rate	Hacc	Hdop	Quality of S	No. of Sat	Instantane	Accl X	Accl Y	Accl Z	Gyro X	Gyro Y	Gyro Z
253488	38:04.0	32.78339	34.96547	2.994447	0	0	0.8	263	17	1.444446	-0.28987	-0.87035	-0.15811	-31.5	48.65	37.87
253489	38:04.0	32.78339	34.96547	2.994447	0	0	0.8	263	17	1.444446	-0.22765	-0.89963	-0.23058	-60.9	-28.56	46.55
253490	38:04.0	32.78339	34.96547	2.994447	0	0	0.8	263	17	1.444446	-0.09882	-0.91061	-0.36746	-114.45	-110.46	60.97
253491	38:04.0	32.78339	34.96547	2.994447	0	0	0.8	263	17	1.444446	0.08784	-0.80374	-0.57901	-149.45	-125.58	77.84
253492	38:04.1	32.78339	34.96547	2.530558	0	0	0.8	263	17	4.638893	0.19032	0.221064	-0.8418	-122.85	-21.21	106.68
253493	38:04.1	32.78339	34.96547	2.530558	0	0	0.8	263	17	4.638893	0.083448	0.961848	-0.74737	-85.89	51.87	82.53
253494	38:04.1	32.78339	34.96547	2.530558	0	0	0.8	263	17	4.638893	0.009516	1.502796	-0.51386	-13.51	87.43	33.04
253495	38:04.1	32.78339	34.96547	2.530558	0	0	0.8	263	17	4.638893	-0.02416	1.572336	-0.26132	63.14	62.51	-1.47

דוגמה למידע גלומי עבור שחקן בודד במהלך משחק

מידע זה מספק מקור מצוין למדידה הפיזיולוגית שאנו צריכים שכן זמן התגובה האנושי בספורטאי עילית הינו 0.15 שניות וכאן אנו מקבלים דוגימה בכל 0.01 שניה. (2)

### OPTA Stats 4.2

נתונים מדויקים המספקים לקבוצות ליגת העל ע"י חברת OPTA. בשיתוף מכבי חיפה השגנו לכל משחק את דוח OPTA עבורו המכיל: הסטברות ניצחון לכל קבוצה, GA עבור כל מצב בעיטה במהלך, סוג התקפות מהן הגיעו הביטה, בעיטות עונשין וכוכלי. בו השתמשנו על מנת להתאים GA לכל משחק ששימוש כ-label לחיזוי ע"פ ההתאמת הזמן בין חותמת הזמן הקיימת בכל דוגימה של המידע הפיזיולוגי שתואר מעלה למידע הגלומי של OPTA עבור משחק ייחיד המגלים את כל ניסיונות ההבקעה במהלך. בנוסף, השתמשנו בנתון ה-Gx המציג כאן לכל ניסיון הבקעה על מנת ליצור את סימולטור ה-Gx שתיארנו מעלה.

TimeStamp	GameClockTeam	Position	FieldWidt	FieldLocat	EventX	EventY	MfromGeo	Gtx	Gky	homeCurr	awayCurr	PosNum	ShotPatter	ShotBody	FromCom	ShotPlayStyle	Pressure	PtryBtwm	sequenceSeconds	possessionSeconds	Goalinout	GoaloutxG
24/08/2024 16:05	347 Bnei Sakhri Centre Foi Centre Thi Attacking1	89	50	13.2	98.6	50.4	0	0	0	3	Penalty	Sh Right Foot			Low	1	0	267	27.6	48.7	0.79	
24/08/2024 16:09	2200 Bnei Sakhri Centre Foi Centre Thi Attacking1	98	59	3.7	97.7	54.2	1	0	0	69	Free Kick	Sh Right Foot		Free Kick Set Piece	Low	0	2	27.8	27.8	56.3	0.31	
24/08/2024 16:21	2303 Bnei Sakhri Centre Foi Left Third Attacking1	71	75	36.6	98.5	51.8	1	0	0	73	Regular	Pl Right Foot		Open play	Low	1	8	23	63.9	48.9	0.02	
24/08/2024 16:30	2877 Bnei Sakhri Centre Foi Centre Thi Attacking1	86	31	17	98.2	47.3	2	0	0	99	Regular	Pl Right Foot		Open play	High	2	14	22	19	48.4	0.04	
24/08/2024 16:35	4274 Maccabi P Right Back Centre Thi Attacking1	93	54	7.9	100	50.4	2	0	0	138	Regular	Pl Head		Open play	High	2	13	13	20.8	57.5	0.1	
24/08/2024 16:36	4388 Maccabi P Right Back Centre Thi Attacking1	77	56	25.1	94.9	52.9	2	0	0	144	Regular	Pl Right Foot		Open play	High	3+	18	18	80.6	36.8	0.03	
24/08/2024 16:38	2099 Bnei Sakhri Central Mi Centre Thi Attacking1	75	52	28.4	98.4	51.7	1	0	0	65	Regular	Pl Right Foot		Open play	Moderate	2	9	17	19	49.6	0.03	
24/08/2024 16:38	2288 Bnei Sakhri Central Mi Left Third Attacking1	89	67	14.9	0	0	1	0	0	73	Regular	Pl Left Foot		Open play	High	2	8	23	10.1	53.6	0.06	
24/08/2024 16:41	566 Bnei Sakhri Left Winge Left Third Attacking1	79	71	25.8	98.2	52.9	1	0	0	9	Regular	Pl Right Foot		Open play	Low	2	31	63	47.2	41.2	0.03	
24/08/2024 16:43	2623 Bnei Sakhri Left Winge Centre Thi Attacking1	93	37	10	95.5	42.3	1	0	0	79	Regular	Pl Right Foot		Open play	Moderate	1	11	72	6.3	53.5	0.38	
24/08/2024 17:06	3012 Maccabi P Central Mi Centre Thi Attacking1	94	58	7.5	99.1	51.4	2	0	0	106	From Corn Head	Yes		Corne	High	2	2	77	19	52.8	0.05	
24/08/2024 17:08	3013 Maccabi P Central Mi Centre Thi Attacking1	93	57	8.6	99.9	51.1	2	0	0	106	From Corn Right Foot	Yes		Corne	High	3+	0	77	19	48.7	0.07	
24/08/2024 17:08	3026 Maccabi P Central Mi Centre Thi Attacking1	92	42	9.2	99.8	48.4	2	0	0	106	Throw-in 5 Head	Yes		Throw-in	High	2	5	77	19.4	58.8	0.05	
24/08/2024 17:09	4030 Maccabi P Central Mi Centre Thi Attacking1	85	40	17.9	99.8	48.9	2	0	0	132	Fast Break	Right Foot		Fastbreak/Counter	Moderate	2	2	11	5.6	58	0.11	
24/08/2024 17:17	5657 Bnei Sakhri Left Attack Centre Thi Attacking1	76	58	27.6	0	0	2	0	0	183	Regular	Pl Right Foot		Open play	Low	3+	34	40	2.5	47.7	0.03	
24/08/2024 17:25	5514 Maccabi P Centre Fo Left Third Attacking1	74	75	33.3	98.6	50.4	2	0	0	176	Regular	Pl Left Foot		Open play	High	2	9	9	19	50.4	0.01	
24/08/2024 17:25	5559 Bnei Sakhri Central Mi Centre Thi Attacking1	73	64	31.9	97.6	50.6	2	0	0	179	Regular	Pl Right Foot		Open play	Low	2	30	30	44.4	35.1	0.02	
24/08/2024 17:27	2513 Bnei Sakhri Left Back Centre Thi Attacking1	93	60	8.8	98.3	51.4	1	0	0	77	From Corn Head	Yes		Corne	High	2	2	75	44.4	42.2	0.03	
24/08/2024 17:29	4103 Bnei Sakhri Left Back Centre Thi Attacking1	81	30	24.4	97.7	46.7	2	0	0	133	Regular	Pl Left Foot		Open play	Low	2	45	45	31.9	59.2	0.04	
24/08/2024 17:31	1306 Maccabi P Centre Mi Centre Thi Attacking1	88	68	16.7	100	50.5	1	0	0	34	Regular	Pl Left Foot		Open play	High	2	14	14	52.8	58	0.03	
24/08/2024 17:50	1795 Bnei Sakhri Right Wing Centre Thi Attacking1	85	40	17.8	98.1	50.5	1	0	0	51	Regular	Pl Left Foot		Open play	High	3+	6	6	45.8	56.3	0.04	
24/08/2024 17:51	3542 Bnei Sakhri Right Wing Centre Thi Attacking1	80	35	23.7	98.8	50.7	2	0	0	117	Regular	Pl Left Foot		Open play	High	2	6	59	19	49	0.02	
24/08/2024 17:53	3979 Bnei Sakhri Right Wing Centre Thi Attacking1	82	35	21.5	97.4	47	2	0	0	129	Regular	Pl Left Foot		Open play	Moderate	2	12	12	54.2	58	0.04	

דוגמה למידע מטור **cs** זה כפי שהתקבל עבור כל משחק:

## SofaScore 4.3

הנתונים שקיבלנו היו אונימיים פרט לכך על מנת לשמור על חיסיון השמות של השחקנים. בפועל, לטובות ניקיון המידע ויצירת הפיצרים, הצלבנו מיקומים של שחקנים כפי שהתקבלו מוגפיות המשחק בין הרכיבי הקבוצות בכל משחק כדי לקבוע ידנית את תפקידו (3). בנוסף, ממנו ייצרנו את הפיצ'ר possession.

## 4.4 הקלטות ידאו רשומות של המשחקים

סרטוני מנהלת הליגה מענות משחק **24-25**, **23-24**, **you** ב-**YouTube**. צפינו בהם על מנת ליצור קובץ metadata מדויק באופן ידני שבאמצעותו סיננו רק את דקוט המשחק הרלוונטיות מטור המידע הפיזיולוגי שסופק לנו מכבי חיפה.

דוגמה לdataset הסופי שיצרנו בהמשך.

## 5 הפיצרים

את הפיצרים ייצרנו בחלוקת ל-3 קבוצות תפקידים טקטיים מרכזיים במשחק ה��ורגל: שחקני הגנה (defenders), שחקני קשר (midfielders) ושחקני התקפה (attackers). בנוספ', שמרנו את סכום הפיצרים של כל הקבוצה לכל פיצר שייצרנו. טכנית, הנתונים של כל שחקן נסכמים על אינטראול הזמן שהגדכנו (מחצית או משחקים מלאים) והנתון שמתקיים בסכום עבור תפקידו של השחקן, לפי תפקיד x כאשר:  $x \in \{defenders, attackers, midfielders\}$ , ייצרנו לצורך האקספלורציה 2 מගרי נתוניים: עבור מחצית (סכימה על אינטראול של כ-45 דקות) ועבור משחקים מלאים (סכימה על אינטראול של כ-90 דקות).

### **zone\_5\_distance\_x 5.1**

סה"כ המרחק שיכסו השחקנים שתפקידם x באינטראול הזמן הרלוונטי (מחצית או משחק שלם) כאשר נעו ב מהירות שבין 19.8 ל-25 קמ"ש.

$$\text{zone\_5\_distance} = \text{zone\_5\_distance\_defenders} + \\ \text{zone\_5\_distance\_midfielders} + \text{zone\_5\_distance\_attackers}$$

### **zone\_5\_time\_x 5.2**

סה"כ הזמן שבו השחקנים שתפקידם x שהו ב מהירות שבין 19.8 ל-25 קמ"ש באינטראול הזמן הרלוונטי (מחצית או משחק שלם).

$$\text{zone\_5\_time} = \text{zone\_5\_time\_defenders} + \text{zone\_5\_time\_midfielders} + \\ \text{zone\_5\_distance\_attackers}$$

### **zone\_6\_distance\_x 5.3**

סה"כ המרחק שיכסו השחקנים שתפקידם x באינטראול הזמן הרלוונטי (מחצית או משחק שלם) כאשר נעו ב מהירות שמעל 25 קמ"ש.

$$\text{zone\_6\_distance} = \text{zone\_6\_distance\_defenders} + \\ \text{zone\_6\_distance\_midfielders} + \text{zone\_6\_distance\_attackers}$$

## **zone\_6\_time\_x 5.4**

סה"כ הזמן שבו השחקנים שתפקידם X שהו ב מהירות שמעל ל-25 קמ"ש באינטראול הזמן הרלוונטי (מחצית או משחק שלם).

$$\text{zone\_6\_time} = \text{zone\_6\_time\_defenders} + \text{zone\_6\_time\_midfielders} + \text{zone\_6\_distance\_attackers}$$

הערה: תכונות זמן נסכמו ע"י אינטראול הדגימה, כלומר, נוסף  $dt=0.01$  עבור המהירות הרלוונטית אם השחקן היה ב מהירות באזור 5 או 6. תכונות מרחק חושב ע"י:  $speed*dt$  כאשר speed היא מהירות השחקן הרלוונטי ו-  $dt=0.01$ .

המרחק והזמן נוספים לפיצ'ר הרלוונטי רק אם המהירות באותה דגימה הייתה באזור מהירות 5 או 6.

## **accelerations\_x, decelerations\_x 5.5**

מדדים לתנועות מתרוצצות של שחקנים, הגעה לתאוצה שגודלה גבוהה, חיובית או שלילית בפרק זמן קצר. חושבו על יד:

1. מיצוע של דגימות מידע ה-GPS בכל חלון של עשירית שנייה (כלומר מיצענו כל 10 דגימות לדגימה אחת).
2. עבור כל דגימה שהתאוצה בה עולה מעל 3 מטר לשנייה (הרף הסטנדרטי בספורט הספורט ב-STATSport שמספקת את המידע משתמש (4)) הסתכמנו על 5 הדגימות הבאות וביצענו vote major ביןיהן לפי המהירות כדי להבין האם זאת הייתה תאוצה או תאומה. מספרי התאוצאות שהתקבלו דומים בסדר הגודל לאלו ששיתפו איתנו מכבי חיפה במערכת המידע אצלם מתוך משחקי אימון.

$$\text{accelerations} = \text{accelerations\_defenders} + \text{accelerations\_midfielders} + \text{acceleration\_attackers}$$

$$\text{decelerations} = \text{decelerations\_defenders} + \text{decelerations\_midfielders} + \text{decelerations\_attackers}$$

## **possession 5.6**

פיצ'ר רציף בין 0 ל-1 המתאר מהו החלק יחסי של המשחק שבו הקבוצה החזיקה בצד. (לדוגמה: מכבי חיפה החזיקה בצד פי 3 מיריבתה). אותו הכנסנו ידנית ע"י בדיקה של כל משחק ב-SofaScore dataframe ל-dataframe של משחקים מלאים בלבד, שכן הנתון מהאתר ק"י'ם עבור משחקים מלאים.

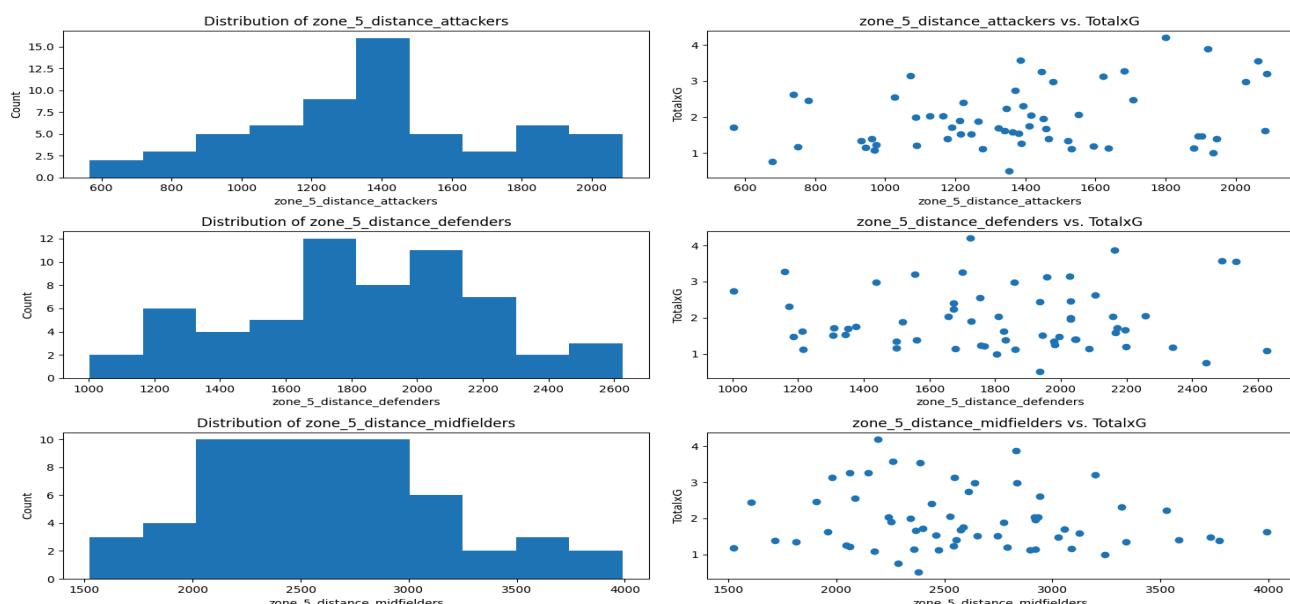
נבחן כי יש לנו בסה"כ 25 פיצרים אשר רובם תלויים זה בהז, לדוגמה: תאוצה היא נגזרת של מהירות בזמן, פיצרים הסוכמים מרחקי ריצה לכל הקבוצה תלויים בפיצרים המחשבים מרחק ריצה לכל תפקיד (זו אפילו תלות לנארית), פיצרים הסוכמים מרחק באיזור מהירות מסוים מאוד תלויים בזמן שננסכם באותו אזור מהירות. נתיחס לכך בשלב מאוחר יותר.

## 5.7 התפלגות הפיצרים למול Gx

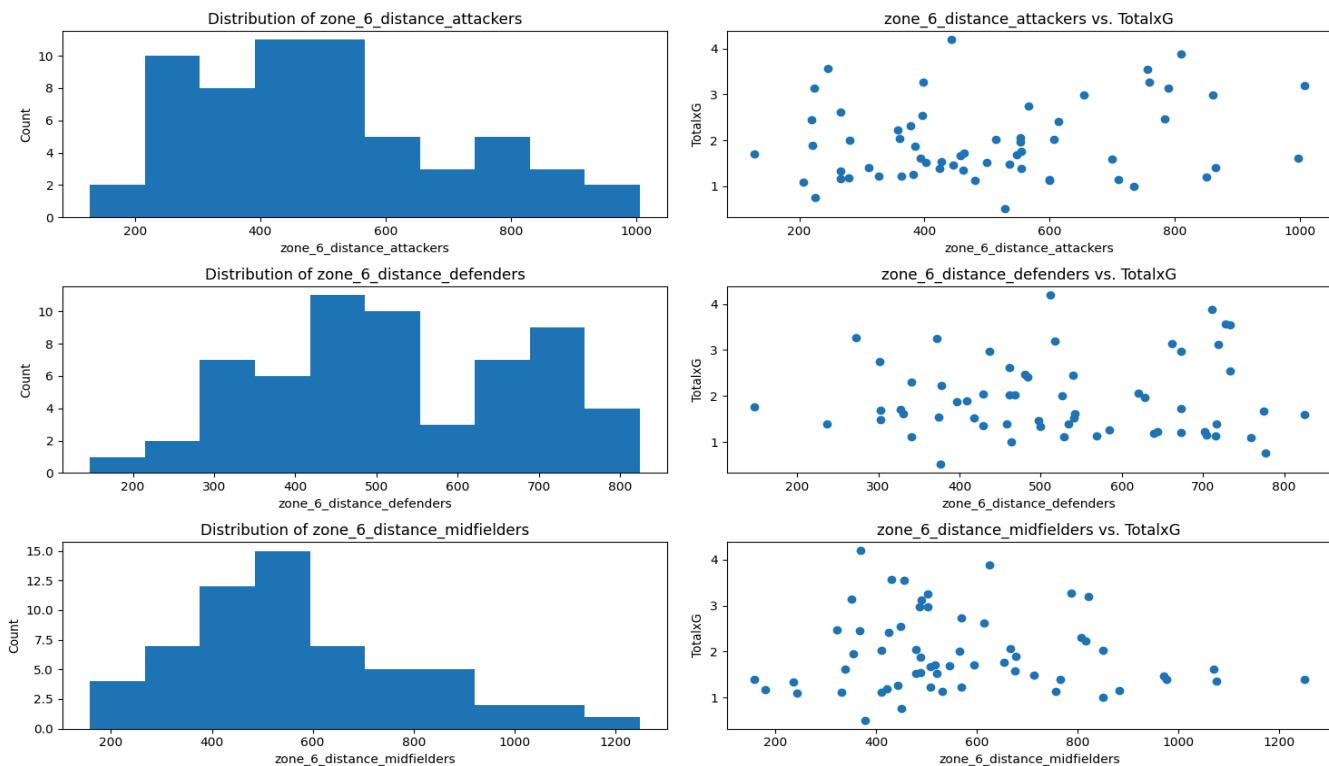
לצורך ההמחשה של נראות הפיצרים, נציג כיצד כל פיצר מתפלג בפני עצמו וכי צד מתפלג למול Gx לאורך משחקים שלמים. לא נראה את התפלגות פיצרי זמן מכיוון שהוא דומה מאוד לזה של פיצרי המרחק (וקורלציה ביניהם שואפת ל-1).  
בכל האירומים הבאים -

**מצד שמאל** רואים היסטוגרמה פשוטה של הפיצר כאשר count הוא כמות המופעים של הפיצר בערך הספציפי. (לדוגמה, בטבלה הראשונה: ישנו 2 משחקים מלאים בהם סך המרחק שביצעו שחקני התקפה באיזור 5 הוא בין 500 ל-700 מטרים.  
**מצד ימין** רואים את התפלגות הפיצר לעומת גודל Gx שהוא label אותו אנו רוצים לחזות:

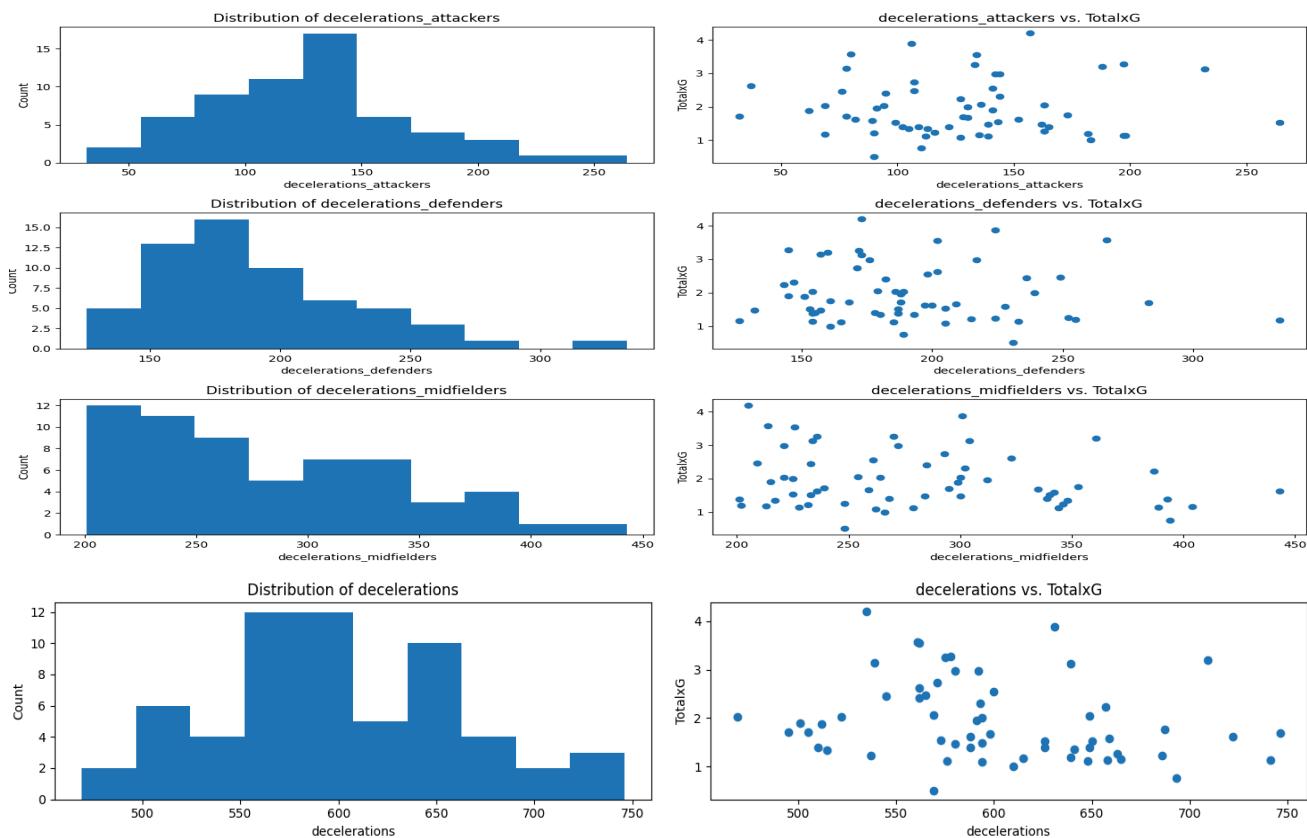
התפלגות פיצרים של מרחק בזון 5, ותציגו אל מול גודל Gx כולל שנמדד בכל משחק. עבור 3 העמדות השונות.



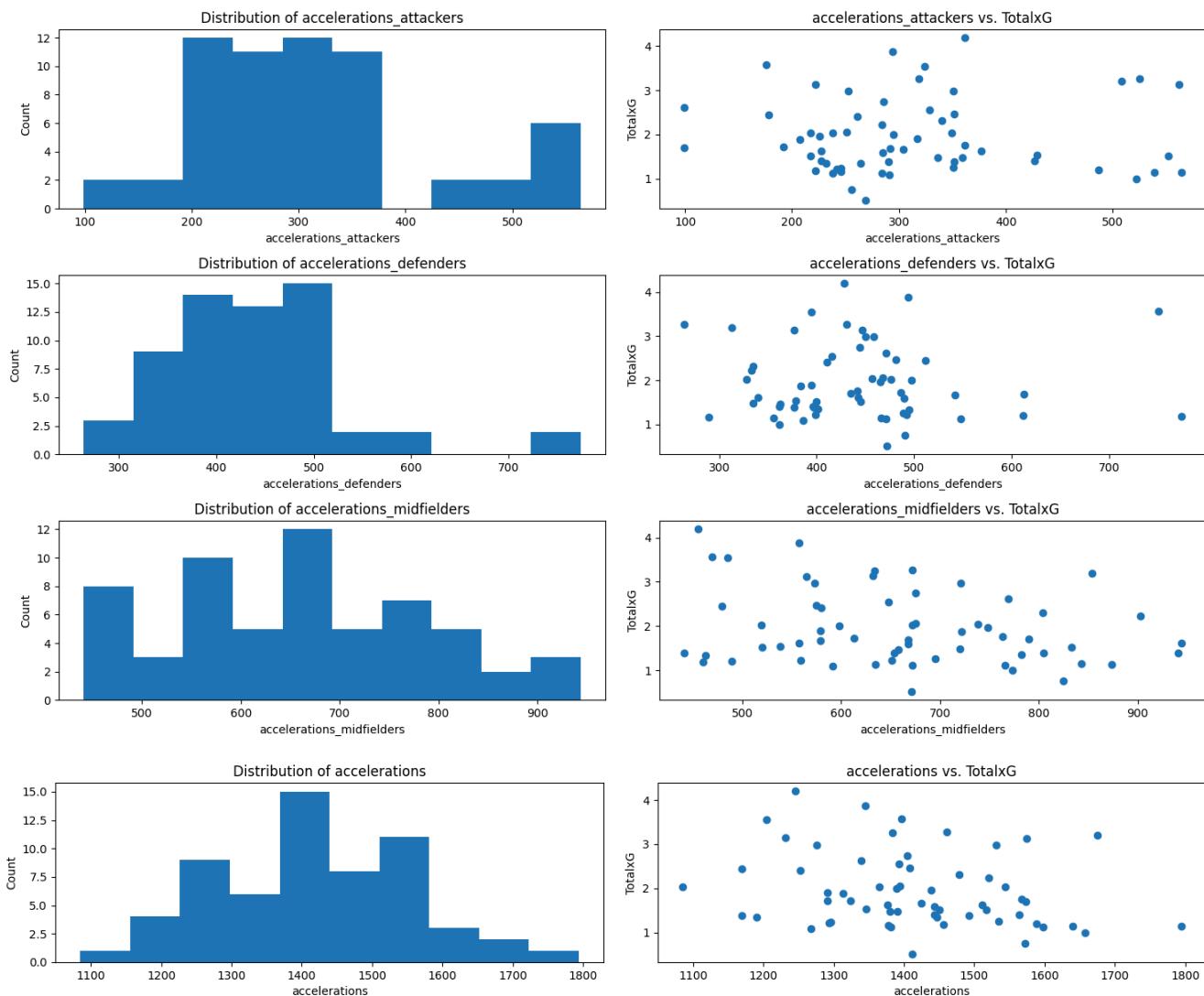
התפלגות פ'יצרים של מרחקゾן 6, ותצוגה אל מול מדד Gx הכלול שנמדד בכל משחק. עבורי 3 העמדות השונות.



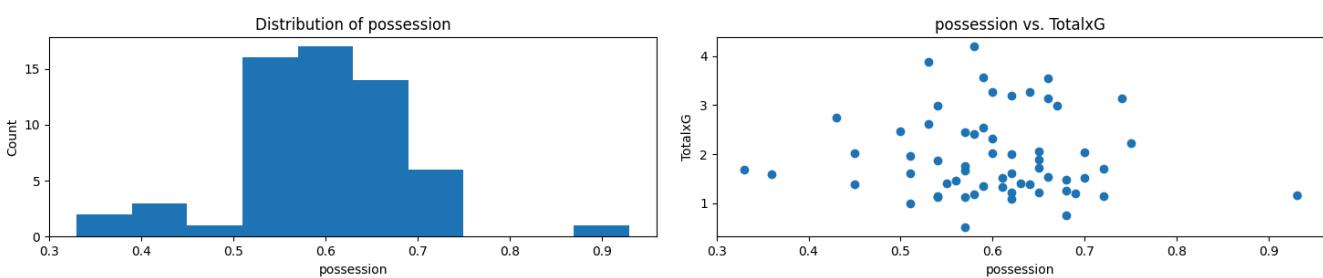
התפלגות פ'יצר התואטה, ותצוגה אל מול מדד Gx הכלול שנמדד בכל משחק. עבורי 3 העמדות השונות ולכל הקבוצה.



**התפלגות פיצ'ר התאוצות , ותצוגה אל מול מדד הGx הכללי שנמדד בכל משחק. עבור 3 העמדות השונות ובסך הכל לכל הקבוצה.**



**התפלגות פיצ'ר החזקת הכדור, ותצוגה אל מול מדד הGx הכללי שנמדד בכל משחק.**



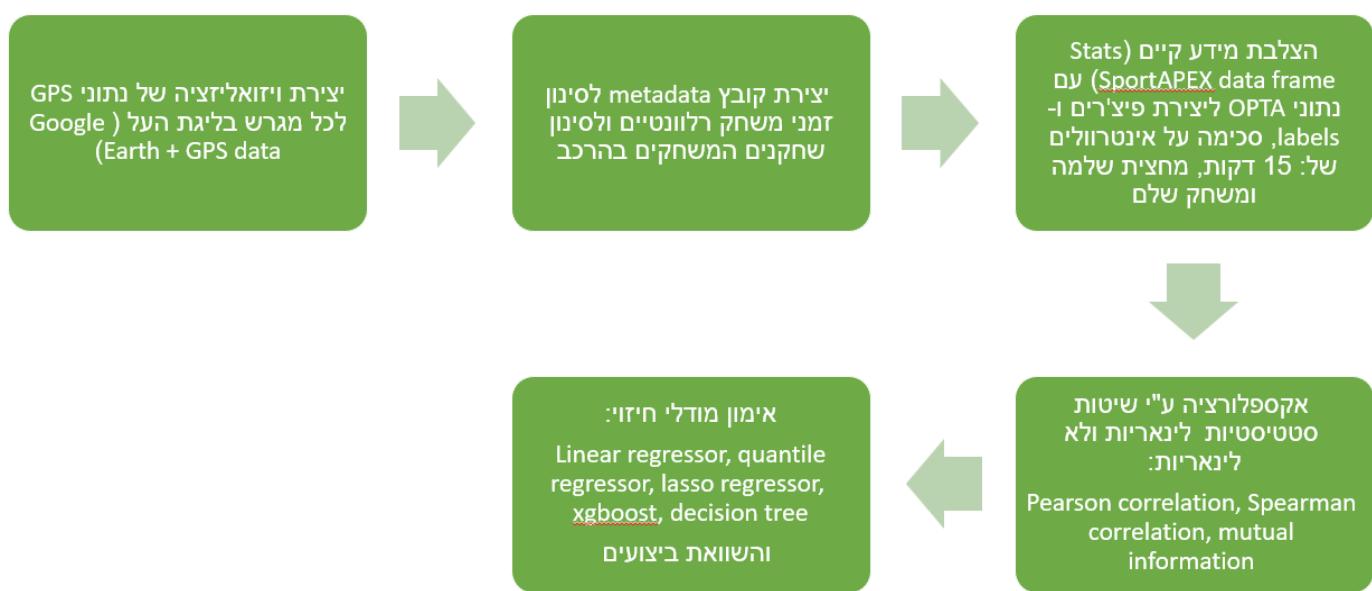
ניכר כי ישנו באופן יחסי מיעוט דוגמאות וקשה להגיד מהי התפלגות הפיצ'רים לה הינו מצפים (נורמלית / מעריכית/ לוגריתמית וכו').

## 6 מетодולוגיה

בפרק זה נתאר את שיטת הפעולה בה נקטנו על מנת לבחון את הבעה שתוארה בפרק 2, למשה יצרנו פיףליין שאפשר לנו לבחון את הקשרים הנבחנים במספר זוויתות שונות תוך שינוי מינימלי, בפרק 7 ניתן לראות כיצד שניים שביצעו ותוצאותיהם.

### 6.1 תיאור המערכת והאלגוריתמייה לפתרון הבעיה

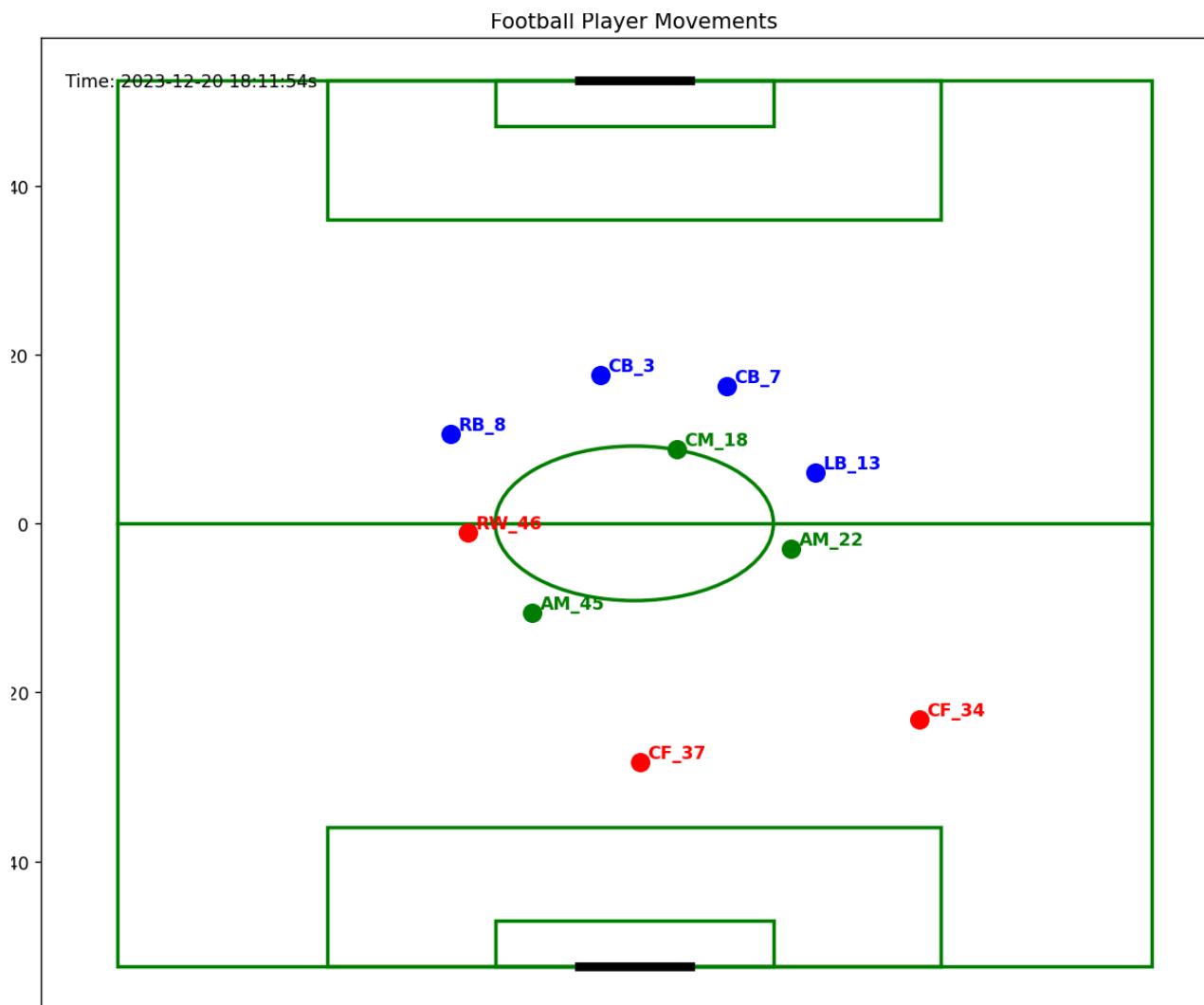
תרשים כללי:



### 6.2 ייצור ויזואליזציה

על מנת להבין את המידע בצורה אינטואטיבית, השתמשנו ב- Google Earth כדי למצוא מרכז מגזרים שעיליהם שוחקו המשחקים ואת ציון המגרש (זווית בה נמצא קו ישר בין מרכז שני שעריו המגרש). על פניהם נתונים אלו נירמלנו את המידע ה-GPS-י שהוא בידינו וייצרנו ויזואליזציה של תנועת השחקנים המתואמת במשחק על פני המגרש. שחקני התקפה מסומנים באדום, שחקני קישור בירוק ושחקני הגנה בכחול.

הנ"ל היה חשוב מאוד כדי לקבל תחושה ראשונית לאיך נראה המידע איתנו עובדים ולהשוו עלי ציוני פיתוח שונים לפרויקט.



דוגמה מתוך כלי היזואליזציה המתוואר, שחקני התקפה בצביע אדום, שחקני קישור בירוק ושחקני הגנה בכחול,  
מידות המגרש מנורמלות למידות מגרש סטנדרטי

מרכז המגרש מ- Google Earth אינו מדויק די וביצעוו שלב fine tuning נוסף בו הצלבנו  
ידנית את השחקן הבועט בצדור בפתחת המחזית בה מכבי חיפה מתחילה מודיאו של  
השחקן למול היזואליזציה הראשונית וכך הגיעו לנרטול מדויק למרცח המגרש.

מצין גם שכלי זה גילה לנו שגיאות במידע GPS שהוא קשה לזהות ללא המידע היזואלי,  
שגיאות אלו נבעו ממספר סיבות כגון מגרשים שחילק מהאזרחים שלהם בנויים מבטון  
ומפריעים לקליטת האות או שיבושי GPS שהתרחשו ברחבי המדינה במקביל למשחקים.

### 6.3 ייצור קובץ metadata

כדי לשנן מידע GPS ולהישאר רק עם החלקים הרלוונטיים ליצירת פיצרים, תחילת  
תיגנו לכל משחק את האצטדיון בו שוחק, לפי הרישומים של הנהלת הליגות לכדורגל  
בישראל. באמצעות כלי היזואליזציה שתואר בסעיף 6.2, ניתן לטעון חלקים שונים של  
המשחק לפי שעת התחלתה וסוף. כדי לזהות את שעות ההתחלה והסיום המדוייקות של כל

מחצית (מידע שאינו מופיע ברשות), ביחסויה של עשירית שנייה, טענו את שעת תחילת המשחק הרשמית לכל, בעזרתו ניתן לzechot תנוצה מתואמת של כל השחקנים במשחק בו זמן. במקביל צפינו בהקלות וידעו מלאות של המשחקים, המתפרסמים על ידי מנהלת הליגות לכדורגל בישראל על מנת לzechot את הזמן המדויק של מחצית ארכה, באמצעות ההצלבה זו תייגנו ידנית את השעות המדויקות בהן התחלת והסתיימה כל מחצית עבור 62 משחקים מלאים.

game_folder_name	stadium	start_csv_time	first_half_finish_csv_time	second_half_start_csv_time	second_half_finish_csv_time	red cards
2023-09-03-September 3, 2023-RawDataExtended	Teddy_stadium	17:00:58	17:48:15	18:04:11		18:55:01
2023-09-17-September 17, 2023-RawDataExtended	HaMoshava_stadium	17:01:00	17:51:05	18:07:13		19:04:58
2023-09-27-September 27, 2023-RawDataExtended	Sammy_Ofer_stadium_before_Nov	17:02:36	17:51:02	18:08:06		18:59:56
2023-10-01-October 1, 2023-RawDataExtended	Netanya_stadium	17:16:33	18:04:35	18:20:24		19:11:31
2023-11-25-November 25, 2023-RawDataExtended	Sammy_Ofer_stadium_before_Nov	18:00:47	18:46:45	19:02:29		19:52:36
2023-12-03-December 3, 2023-RawDataExtended	Sammy_Ofer_stadium_before_Nov	18:31:25	19:16:25	19:33:40		20:23:55
2023-12-10-December 10, 2023-RawDataExtended	Sammy_Ofer_stadium_before_Nov	18:30:40	19:18:50	19:34:25		20:26:25
2023-12-17-December 17, 2023-RawDataExtended	Sammy_Ofer_stadium_before_Nov	18:32:00	19:19:00	19:34:45		20:24:50
2023-12-20-December 20, 2023-RawDataExtended	Sammy_Ofer_stadium_before_Nov	18:00:34	18:47:24	19:03:50		19:51:20
2023-12-24-December 24, 2023-RawDataExtended	Netanya_stadium	18:31:40	19:19:40	19:36:15		21:21:40
2023-12-27-December 27, 2023-RawDataExtended	Sammy_Ofer_stadium_before_Nov	17:59:50	18:48:00	19:03:50		19:55:00
2023-12-31-December 31, 2023-RawDataExtended	Bloomfield_stadium	18:30:10	19:18:20	19:34:25		20:25:00
2024-01-03-January 3, 2024-RawDataExtended	HaMoshava_stadium	18:15:15	19:02:25	19:18:10		20:09:40
2024-01-07-January 7, 2024-RawDataExtended	Sammy_Ofer_stadium_before_Nov	18:31:40	19:18:40	19:35:30		20:25:00
2024-01-10-January 10, 2024-RawDataExtended	Bloomfield_stadium	18:31:05	19:20:00	19:36:40		20:27:40
2024-01-14-January 14, 2024-RawDataExtended	Sammy_Ofer_stadium_before_Nov	18:30:40	19:16:40	19:33:05		20:21:10
2024-01-17-January 17, 2024-RawDataExtended	Doha_stadium	18:01:15	18:48:30	19:03:25		19:50:30
2024-01-21-January 21, 2024-RawDataExtended	Sammy_Ofer_stadium_before_Nov	18:31:40	19:17:45	19:33:00		20:20:50
2024-01-31-January 31, 2024-RawDataExtended	Terner_stadium	18:32:15	19:20:00	19:36:05		20:33:00 CF_37 18:55:00

דוגמה מຕוך קובץ Metadata - כל רשומה מתארת משחק בודד

## 6.4 הצלבת מקורות מידע ויצירת data frame

לכל משחק, יוצרים אינטראול סכימה מחצית שלמה (כ-45 דקות) ומ משחק שלם (כ-90 דקות) שטוכמים את הפיצרים בהם השתמשנו ([כפי שפורטו בפרק 5](#)) ואת סך ה-Gx שהצטבר במהלך מצב המשחק בהתאם לאינטראול הזמן בו הפיצרים נסכמים.

zone_6_tir	zone_5_di	zone_5_tir	zone_6_di	zone_6_tir	total_zone	total_zone	total_zone	accelerati	decelerati	accelerati	decelerati	accelerati	decelerati	accelerati	decelerati	possessio	TotalxG	opponent	xG_diff	xPts	
68.6	1457.79	242.4	457.363	60.2	6021.34	1003.5	1740.8	230.6	1425	598	542	209	579	259	304	130	0.57	1.67	0.6	1.07	2.205
70.3	1189.88	195.6	463.285	60.8	5761.37	955.3	1653	220.5	1291	505	486	188	613	239	192	78	0.65	1.72	1.88	-0.16	1.413
74.8	1087.86	180.3	280.799	37.7	5460.81	908.3	1372.62	182.7	1390	594	497	239	598	225	295	130	0.62	2	0.32	1.68	2.599
45.5	2084.88	345.9	996.534	130.6	5871.21	977.3	1876.73	247.6	1377	588	443	200	557	236	377	152	0.51	1.62	1.7	-0.08	1.266
55.2	1127.62	186.4	514.144	67.6	5025.9	836.4	1392.53	185.2	1085	469	328	154	519	221	238	94	0.6	2.03	1.02	1.01	2.159
66	1381.38	228.8	426.662	56.9	5183.97	864.3	1288.3	172.6	1346	573	379	205	538	225	429	143	0.66	1.54	0.44	1.1	2.253
66	1623.12	268.2	789.663	103	6126.22	1020.4	1998.17	263.8	1574	639	447	173	565	234	562	232	0.66	3.13	0.16	2.97	2.92
43.3	1706.15	281.5	783.263	100.3	5644.44	938.9	1586.49	207.6	1408	565	481	249	575	209	352	107	0.5	2.47	0.86	1.61	2.409
61.5	2063.1	341.9	755.76	99.2	6984.26	1164.4	1945.57	257	1204	562	395	202	485	226	324	134	0.66	3.55	0.55	3	2.881
70.1	1214.84	202.5	402.431	52.9	5904.06	984.5	1464.06	194.2	1517	650	445	153	520	233	552	264	0.7	1.52	1.01	0.51	1.735
33.3	969.333	161.4	206.08	27.1	5771.85	963.1	1209.51	160.8	1268	594	386	205	591	262	291	127	0.62	1.09	0.13	0.96	2.213
23.7	751.127	124.7	278.392	37.2	4615.42	772.2	1097.64	146.1	1456	615	774	333	460	213	222	69	0.58	1.18	0.71	0.47	1.746
56.5	1595.12	264.9	849.456	109.6	6581.84	1096.1	1945	255.3	1588	639	612	255	489	202	487	182	0.69	1.2	1.17	0.03	1.377
56.6	1221.89	202.8	613.314	79.5	5334.87	890.9	1522	199.9	1252	562	411	182	580	285	261	95	0.58	2.41	0.87	1.54	2.535
89.7	1363.6	224.7	699.352	90.2	6654.24	1104.2	2199.36	287.6	1443	659	490	228	668	342	285	89	0.36	1.59	1.68	-0.09	1.285
65.2	1477.9	245.8	654.19	86	6172.42	1029.3	1812.63	239.3	1276	580	450	217	573	221	253	142	0.54	2.98	0.63	2.35	2.789
48.8	1450.24	240.6	552.195	72.7	6402.36	1067.2	1535.77	204.1	1439	591	465	188	748	312	226	91	0.51	1.96	0.67	1.29	2.369
84.5	1921.68	317.7	809.654	105.4	6915.7	1149.3	2144.72	284.3	1345	631	494	224	557	301	294	106	0.53	3.88	1.1	2.78	2.749
21.3	1467.1	245	553.685	72.1	4744.34	795.2	949.182	125.1	1170	510	377	187	441	201	352	122	0.45	1.39	1.41	-0.02	1.334
73.4	1322.21	220.5	547.08	71.1	5248.95	877.5	1396.05	184.1	1573	746	613	283	668	335	292	128	0.33	1.69	0.87	0.82	2.013
45.1	1532	252.7	599.837	77.8	5219.88	869.7	1273.1	168.7	1381	576	471	185	672	279	238	112	0.54	1.12	0.93	0.19	1.439
64.5	1244.37	205.9	499.791	65.3	5204.41	868.6	1397.9	184.2	1450	626	400	187	833	340	217	99	0.61	1.52	2.14	-0.62	1.162
46.7	1071.75	178.8	224.092	30.1	5079.28	847.1	1236.69	164.2	1231	539	377	157	632	304	222	78	0.74	3.14	0.08	3.06	2.939

דוגמה מຕוך dataframe הסופי שיצרנו, אשר שימש לניטוח הבעיה - כל שורה מתארת משחק בודד

## 6.5 בוחינת הקשרים ע"י שיטות סטטיסטיות, ליניאריות ולא ליניאריות

סעיף זה מתאר את תהליכי הבנית הסטטיסטי שנועד לבחון את עוצמתו וכיוון הקשרים בין המדרדים הפיזיים שנאספו (כגון מרחק ריצה, ספרינטים, דופק וכו') לבין ממד' ההצלחה במשחק (כגון  $\text{Gx}$  או תוצאה המשחק בפועל).

תחילה, ערכנו בדיקת מתאם פירסון (Pearson's Correlation) בין כל אחד מה משתנים הפיזיים העצמאיים לבין משתני התוצאה התלויים, במטרה לבדוק אילו פעולות פיזיות מראות קשר ישיר ומובהק סטטיסטית עם ממד' ההצלחה.

לאחר מכן, ביצענו רגסיה ליניארית: שיטה זו מאפשרת לבחון את השפעתם המשולבת של מספר משתנים פיזיים על משתנה התוצאה, תוך נטרול השפעה של משתנים מתערבים. המטרה הייתה לבנות מודל ליניארי המסביר את השונות במידדי ההצלחה באמצעות שילוב אופטימלי של מדרדים פיזיים.

לבסוף ביצענו ניתוח לא ליניארי באמצעות עצי החלטה ואנSEMBLE (ensemble) של עצי החלטה. בצד רגאל קשרים רבים אינם ליניארים (לדוגמה, השפעת הדופק על ביצועים עשויה להיות שונה בעצימות נמוכה מאוד וגובהה מאוד). במטרה לזהות כללים לא ליניארים הקשרים בין הביצועים הפיזיים להצלחה.

## 6.6 אימון מודלי חיזוי והשווואת ביצועים

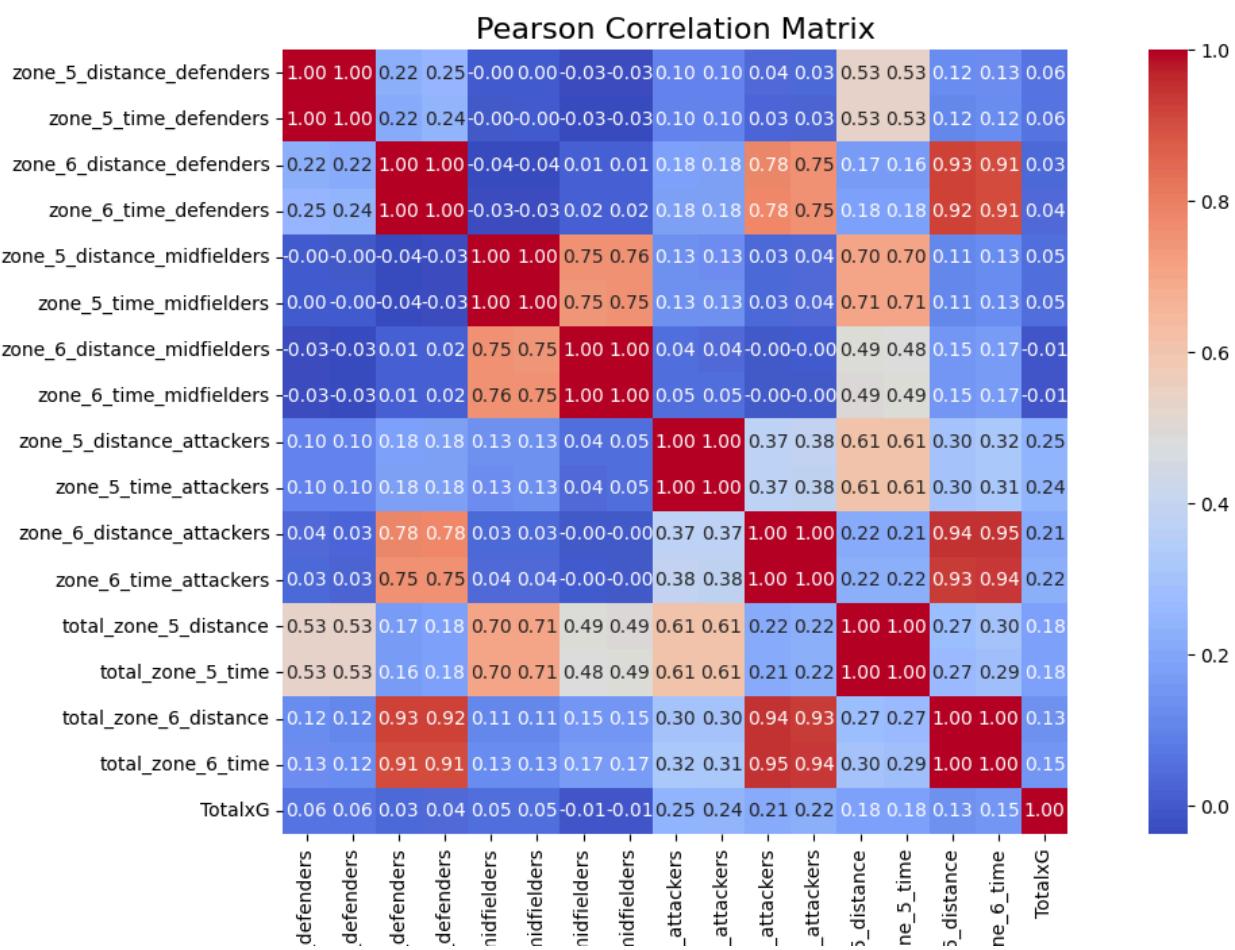
לאחר שזיהינו שאכן קיימים הקשרים שעוניינו אותנו, וחלק אף חזקים כפי שיפורט בפרק 8, בחרנו לאמן מספר מודלי חיזוי על הבעיה. הקלט יהיה הפיצרים שתוארו בפרק 5 והמודל ינסה לחזות את ממד' ההצלחה השונים, בפרויקט נבחנו מגוון מודלים, הכללים בין היתר רגסור ליניארי עם פונקציות  $\text{loss}$  שונות, עץ רגסיה ומודלים מבוססי עצי החלטה כגון Random Forest, XGboost.

כפי שיתואר בפרק 7, המודלים אומנו תוך שימוש בשיטות ואלידציה נפוצות והוערכו על קבוצת המבחן הבלתי נראית באמצעות ממד' ביצועים שונים.

## 7 ניסויים

### 7.1 אקספלורציה סטטיסטיות למול מדד Gx

עבור סכימה של מחצית שלמה (כ-45 דקות), ביצענו בדיקה של הקורלציה הלינארית בין הפיצרים לעצם ובין הפיצרים לבין Gx שראים בעמודה הימנית ביותר. ישנה קורלציה של מעל 0.2 בין ריצות באזוריים 5 ו-6 של שחקי התקפה לבין מדד Gx המעיד על יכולת ההגעה של הקבוצה לנצח הבקעה.



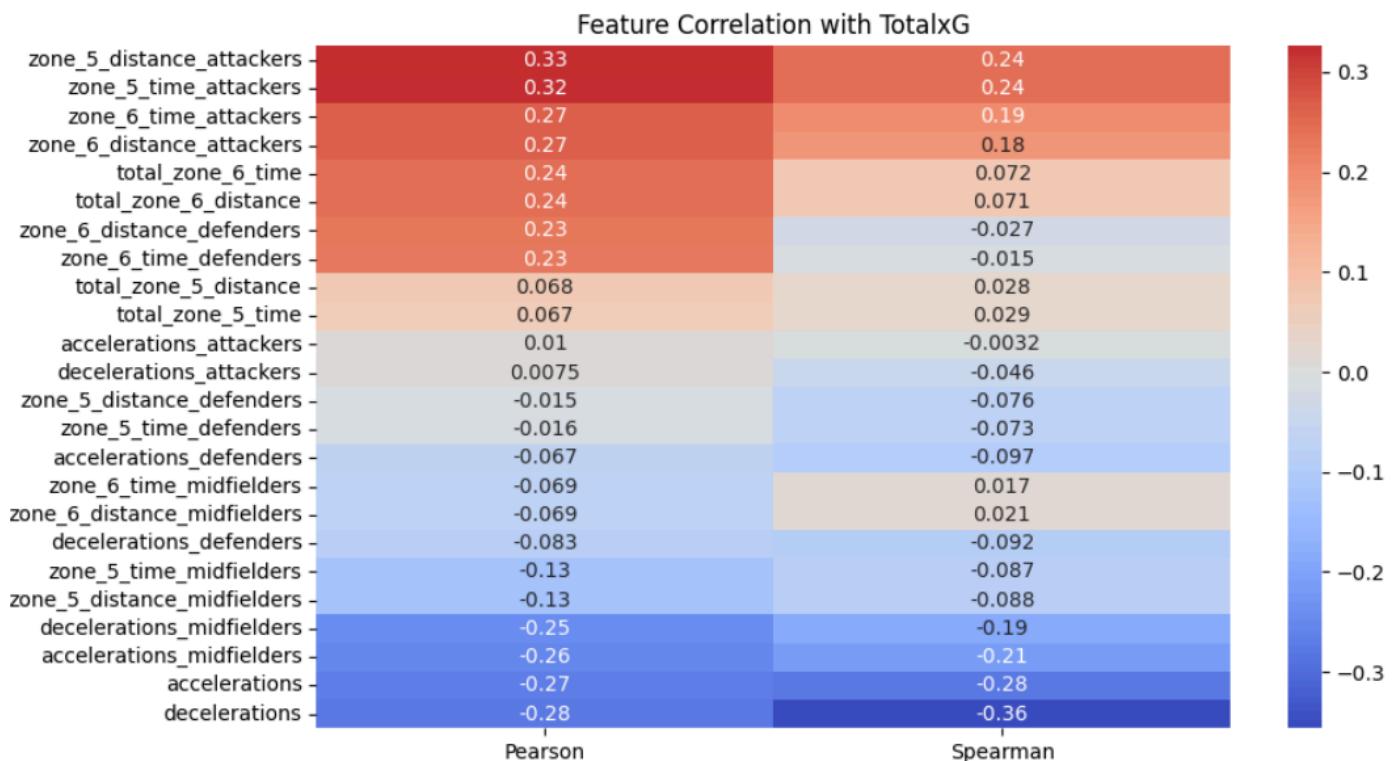
הנ"ל גרם לנו לבצע את השלב הבא של אימון מודלי רגרסיה מסוגים שונים הן על dataset של מחזיות מלאות והן על dataset של משחקים מלאים.

הסיבה לבחירה ב-2 datasets שונים הייתה המיעוט היחסי במידע: בידינו 62 משחקים מלאים.

יצירת dataset שטוכם מחזיות תיצור 124 דגימות, אך על פניו הקורלציות הלינאריות מעט חלשות משל dataset המוצר בסכימה על משחקים מלאים. לעומת זאת, עבור משחקים מלאים הקורלציות חזקות יותר, אך האחרון מותיר בידינו 62 דגימות בלבד dataset הסופי.

בסיומו של דבר, את התוצאות המובהקות ביותר ביותר קיבלנו על dataset של משחקים מלאים, בהן נדון מיד.

א) עבור סכימה של משחקים מלאים, הקורלציית שהתקבלו (ליינארית = Pearson ולא ליינארית Spearman):

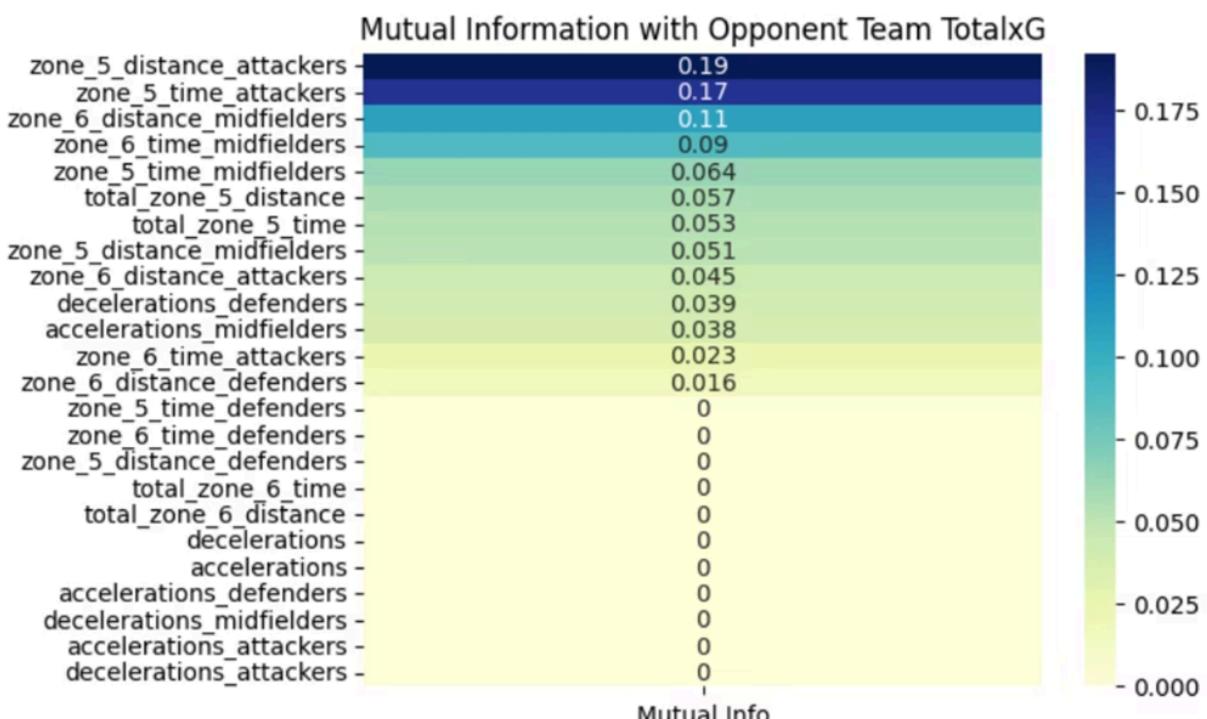
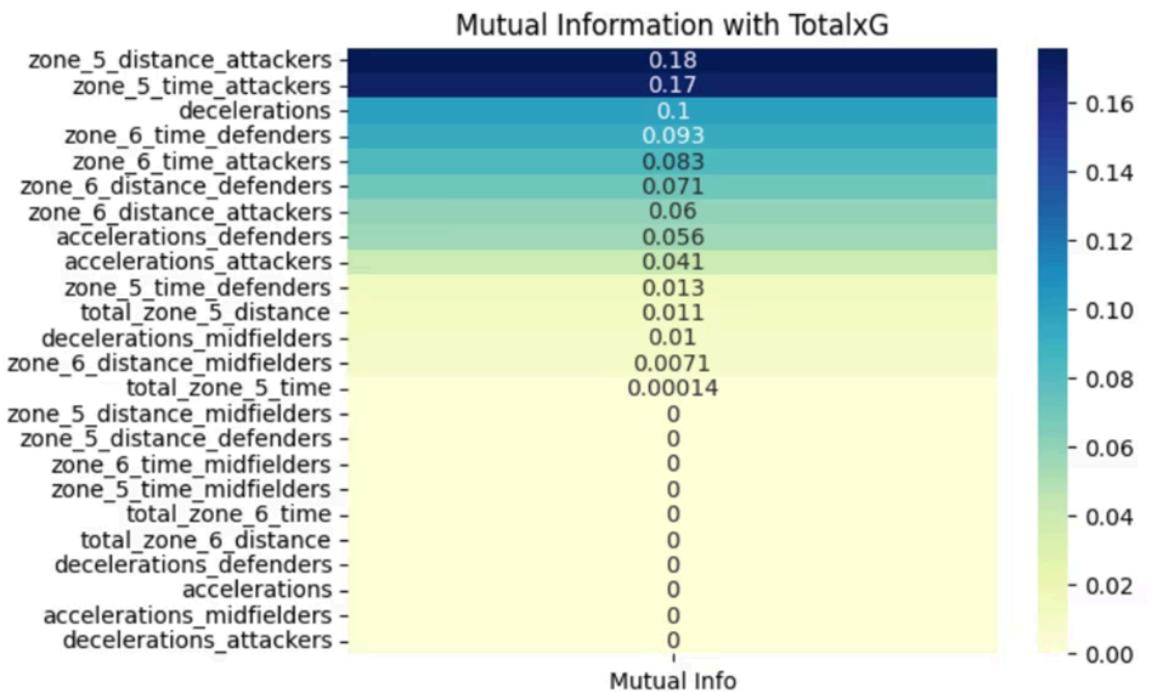


ניתן לראות כי נוסף כאן לעומת האקספלורציהות הקודמות 2 הפיצ'רים:  
accelerations, decelerations

הסיבה להוספת פיצ'רים אלו היא שבבחינה שביצענו בשלב מאוחר יותר של ניסיון לבניית מודלים לא מצאנו מודל שהשיג תוצאה R squared שמצוין שמידع נוסף (פיזיולוגי באופיו גם כן) יכול להוביל לשיפור המודלים שנבחנו.

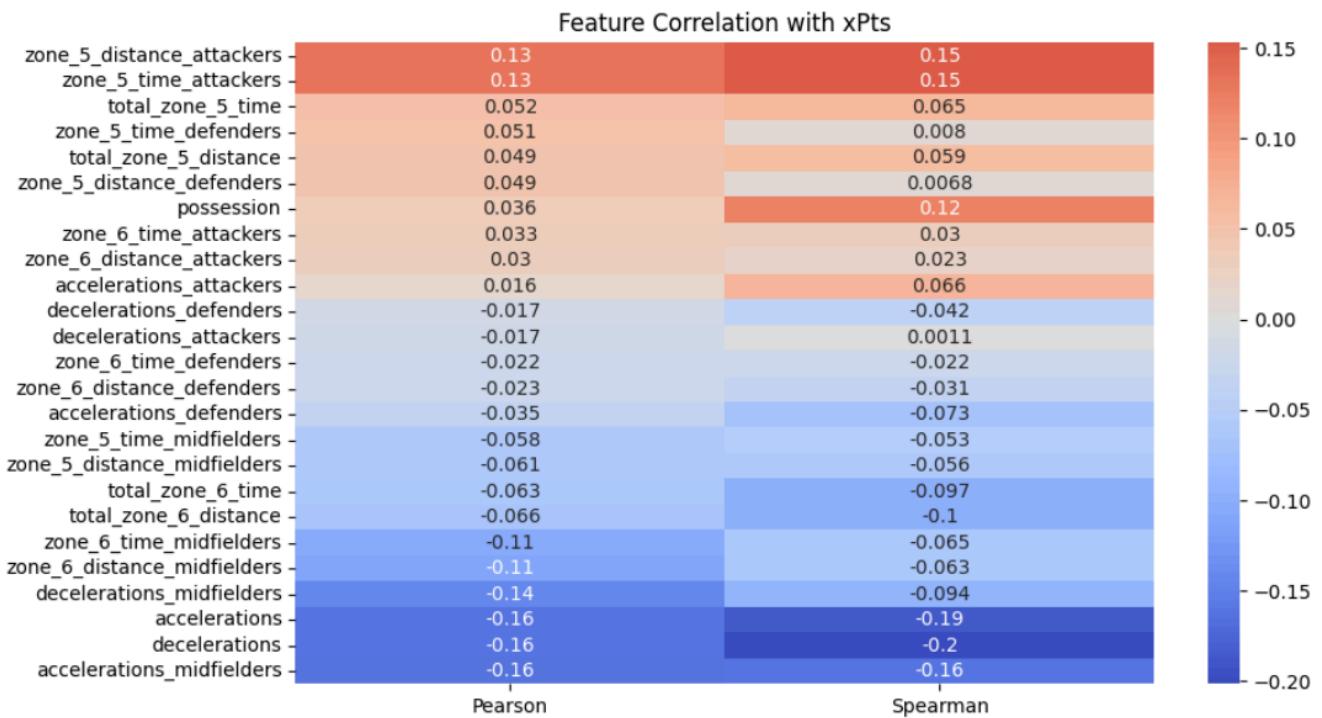
### 7.1.1 Mutual Information

ניסוי נוסף שביצענו היה בחינת mutual information שהוא מדד מבוסס אנטרופיה מותנת<sup>(5)</sup> על מנת להבין מי הפיצרים המכילים הכי הרבה מידע עבור חיזוי  $G_x$ . ביצענו את הבדיקה גם עבור  $G_x$  של הקבוצה היריבה וכן מבחנים שוב כי ריצות מהירות באזורי 5, 6 של שחקני התקפה הן הפיצר החשוב ביותר, יחד עם כמות התאונות שבוצעה.



## 7.2 אקספלורציה סטטיסטיות למול מדד xPts

לאחר שהשתכנענו שניתן להשתמש בסימולטור שבנינו על מנת לחזות xPts, ייצרנו באמצעותו את ה-label xPts וסתכלנו על קורלציה לינארית ומונוטונית בין הפיצרים לבינו:



התוצאות שאנו רואים כאן עוקבות עם מה שראינו עד כה, אם כי פחות מובהקות. הסיבה לכך היא שמספר הנקודות הצפוי הוא מודד מורכב יותר לחיזוי מאשר Gx שמעיד על יכולת הקבוצה להגעה למצוות הבקרה איקוטיים. בכל זאת:

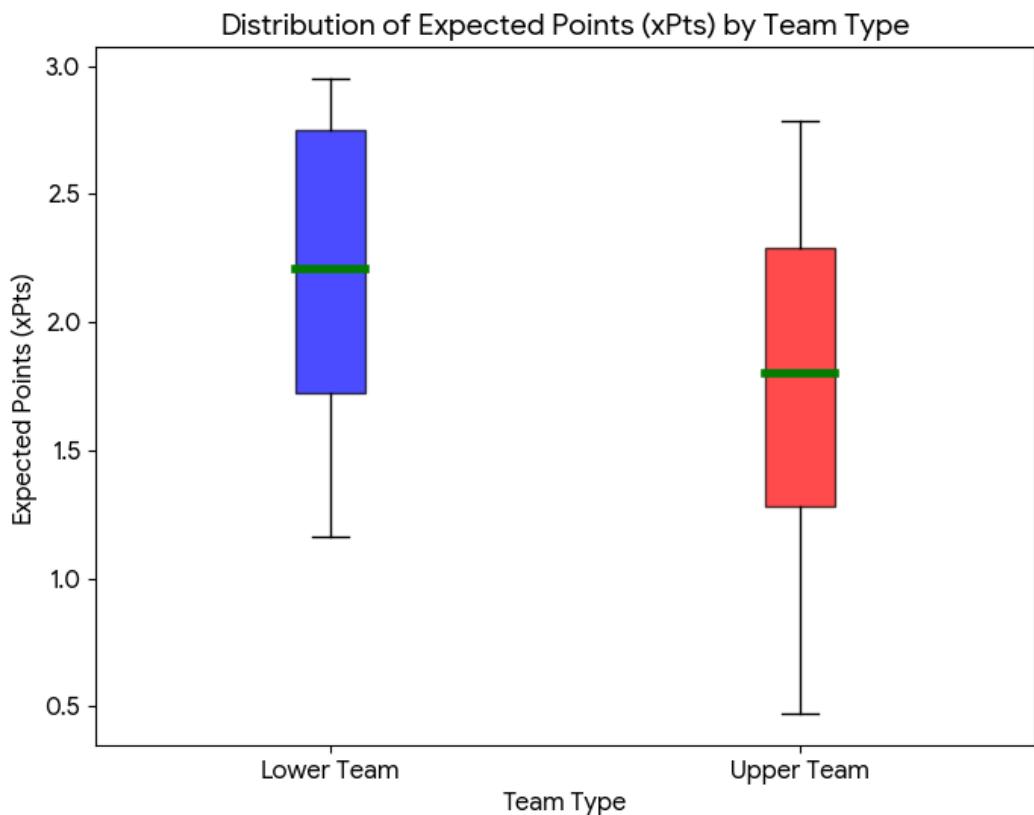
1. ישנו מתאם חיובי ביןוני - חלש בין ספרינטים של שחקני התקפה באזורי 5 לבין מספר הנקודות הצפוי.
2. ישנו מתאם שלילי ביןוני בין מספר התאצויות והתקאות שביצעו השחקנים לבין מספר הנקודות הצפוי. הנ"ל עשוי לנבוע מתכיפות שנייה כיוון גבולה שעשויה להיגרם מאיבוד כדור המוביל לשינוי כיוון פטאומי. כמו כן, ריבוי תאצויות של שחקני הקישור הוא הגורם המוביל מבין כלל התאצויות לקורלציה השלילית.

### 7.3 אקספלורציה סטטיסטיות למול עוצמת הקבוצה היריבה

במטרה להבין האם התוצאות משתנות למול קבוצות חלשות או חזקות, חילקנו את הקבוצות לפי איקוטן האובייקטיבית: קבוצה שסימנה את עונתה עם מספר נקודות מתחת לחציון הנקודות שצברו קבוצות בליגה תיקרא "קבוצת פלייאוף תחתון".

קבוצה שסימנה את עונתה עם מספר נקודות מעל לחציון הנקודות שצברו קבוצות בליגה תיקרא "קבוצת פלייאוף עליון".

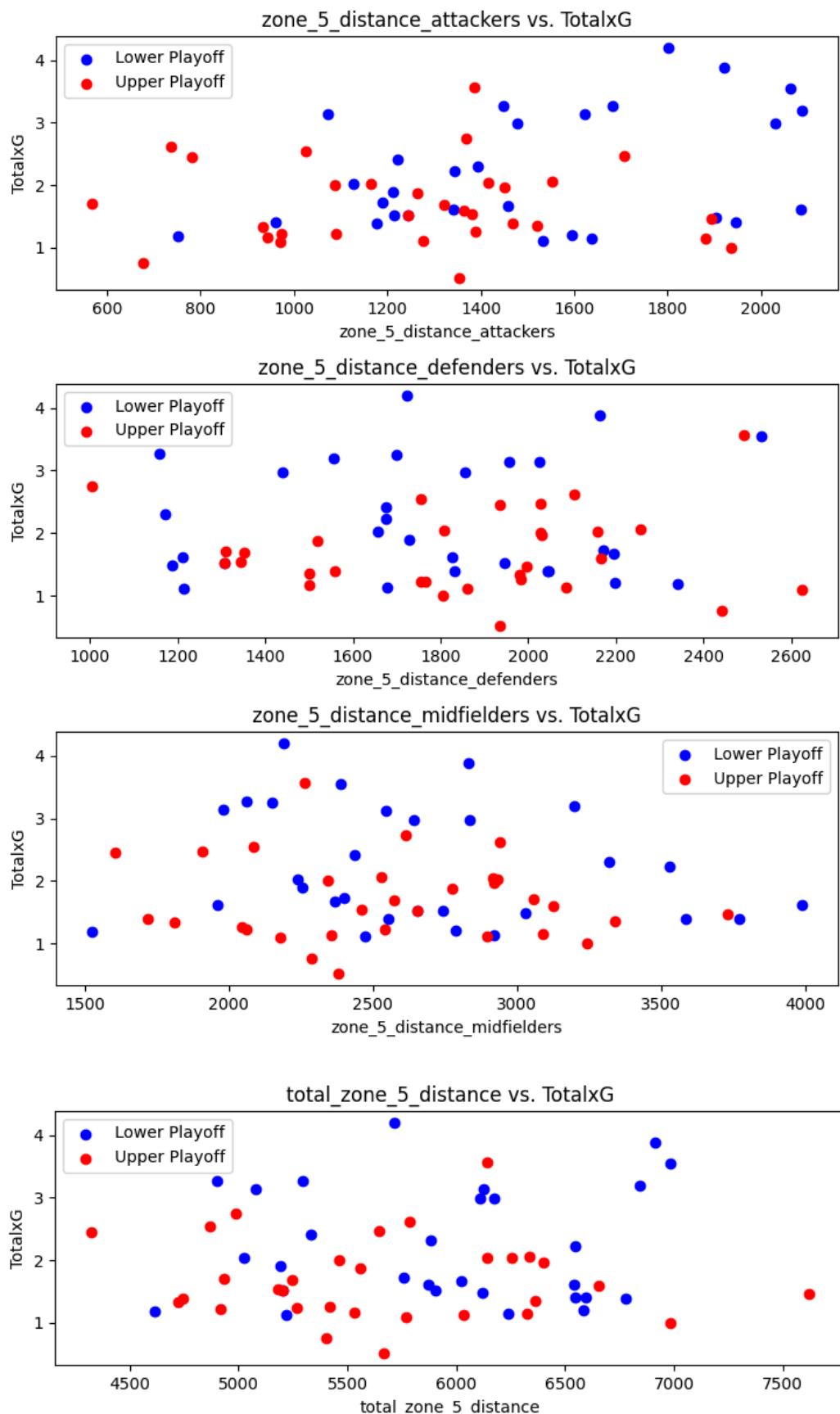
לצורך המחשה, נראה תחיליה את ערכי ה $x$ pts שהושגו ע"י מכבי חיפה נגד קבוצות מ2 הסוגים על מנת להמחיש את הבדלי הקושי במשחקים ביניהם:



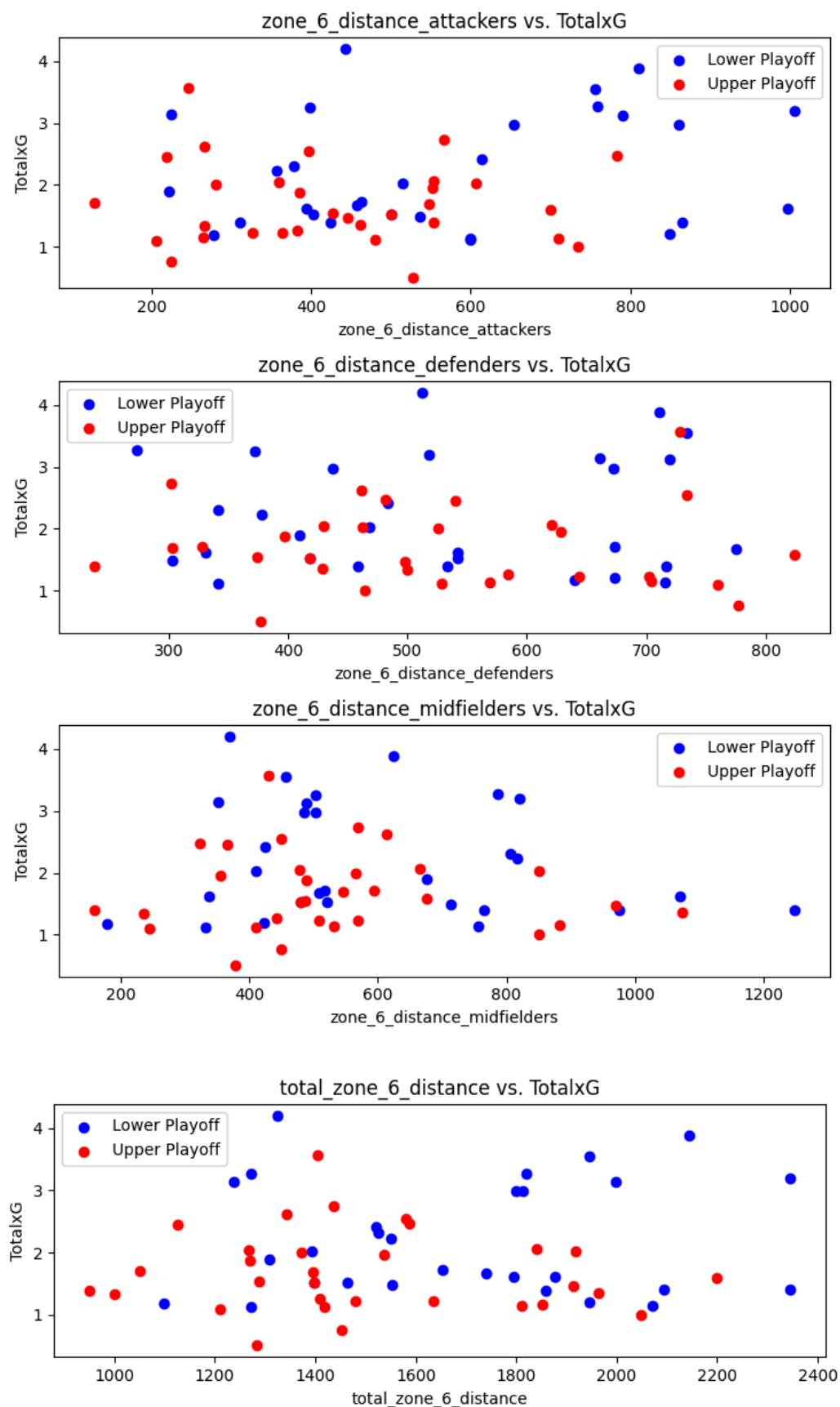
ניתן לראות שנגד קבוצת פלייאוף עליון, מדד הנקודות הצפויות נמוך יותר באופן ממוצע מאשר נגד קבוצת פלייאוף תחתון, בנוסף גם ערכי הקיצון נמוכים יותר.

נראה כיצד מפולוגים הפיצרים שלנו למול  $Gx$  ועוצמת היריבה, הנ"ל הוא חלוקה של התפלגות הפיצרים כפי שנראית בפרק התפלגות הפיצרים:

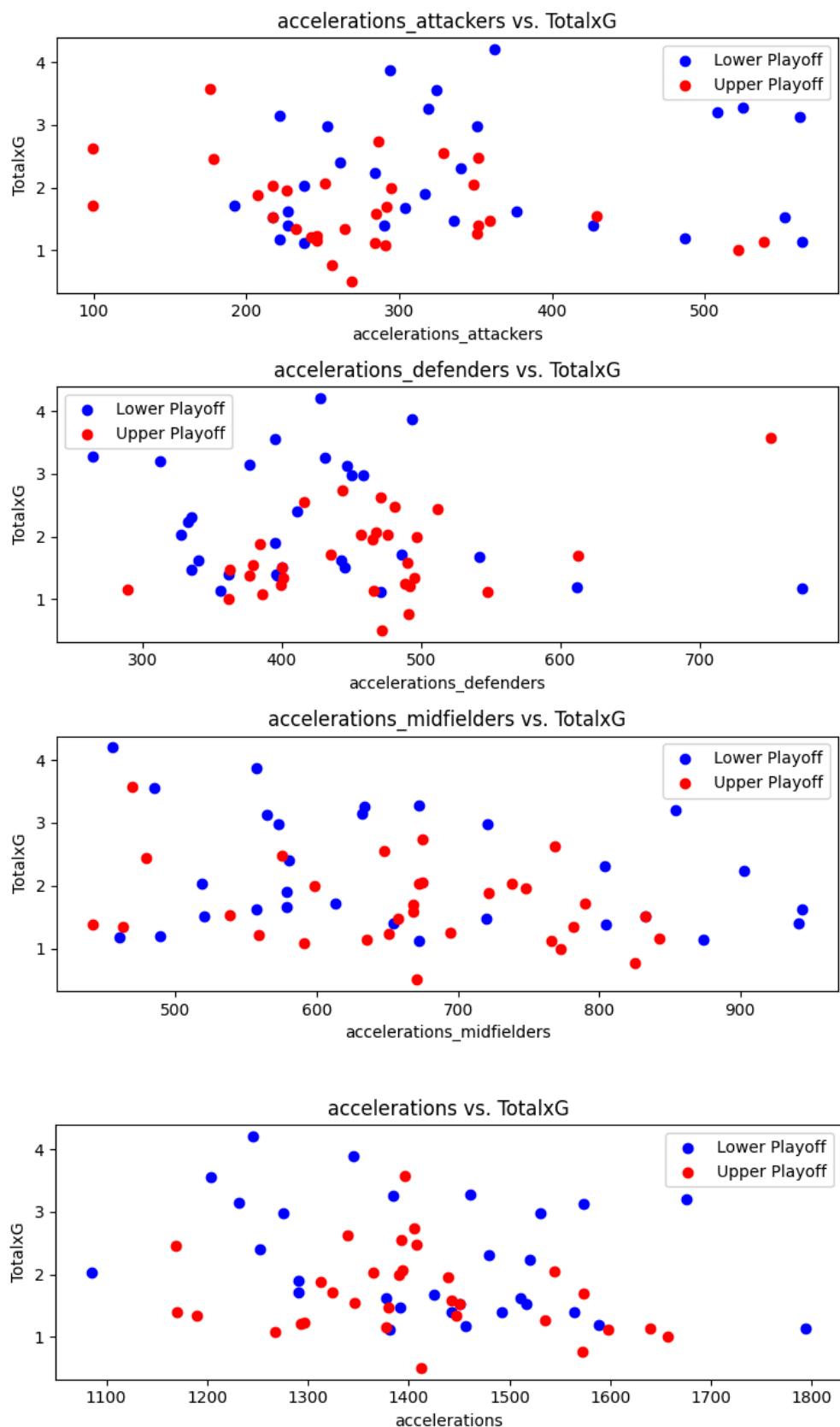
התפלגות פ'יצרים של מרחק בזון 5, ותצוגה אל מול מדד הGx הכלול שנמדד בכל משחק. עבור 3 העמדות השונות ובשם הכל לכל הקבוצה. בכחול - משחקים נגד קבוצות פלייאוף תחתון באדום - משחקים נגד קבוצות פלייאוף עליון



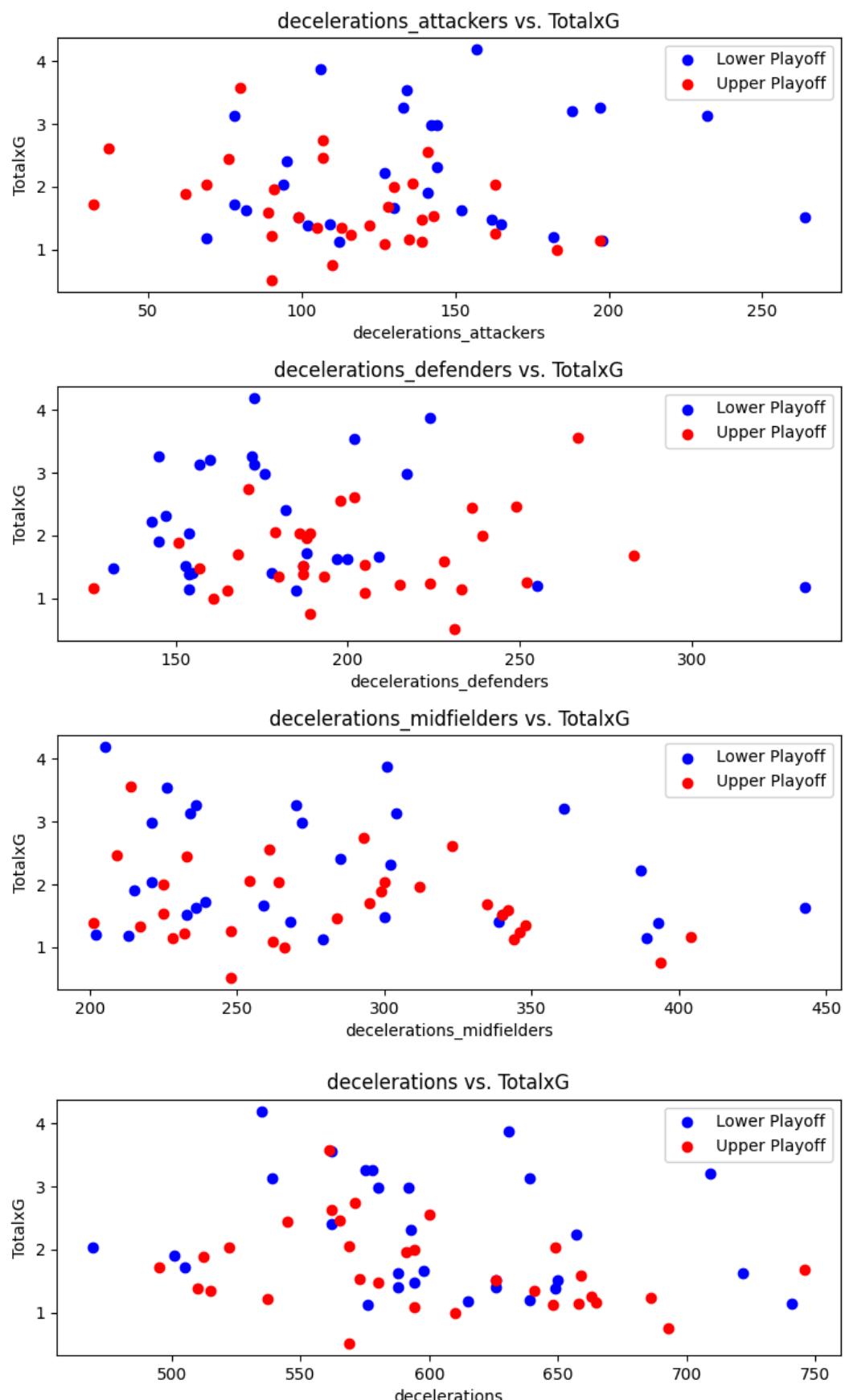
התפלגות פ'יצרים של מרחקゾן 6, ותצוגה אל מול מדד הGx הכלול שנמדד בכל משחק. עבור 3 העמדות השונות ובשם הכל לכל הקבוצה. בכחול - משחקים נגד קבוצות פלייאוף תחתון באדום - משחקים נגד קבוצות פלייאוף עליון



התפלגות פ'יצר התואזה, ותציגו אל מול ממד ה $\text{Gx}$  הכלל שנמדד בכל משחק. עבור 3 העמדות השונות ובסס הכל לכל הקבוצה. בכחול - משחקים נגד קבוצות פלייאוף תחתון באדום - משחקים נגד קבוצות פלייאוף עליון



התפלגות פ'יצור התואמת, ותציגו אל מול מד Gx הכלל שנמדד בכל משחק. עבור 3 העמדות השונות ובסס הכל כל לכל הקבוצה. בכחול - משחקים נגד קבוצות פלייאוף תחתון באדום - משחקים נגד קבוצות פלייאוף עליון

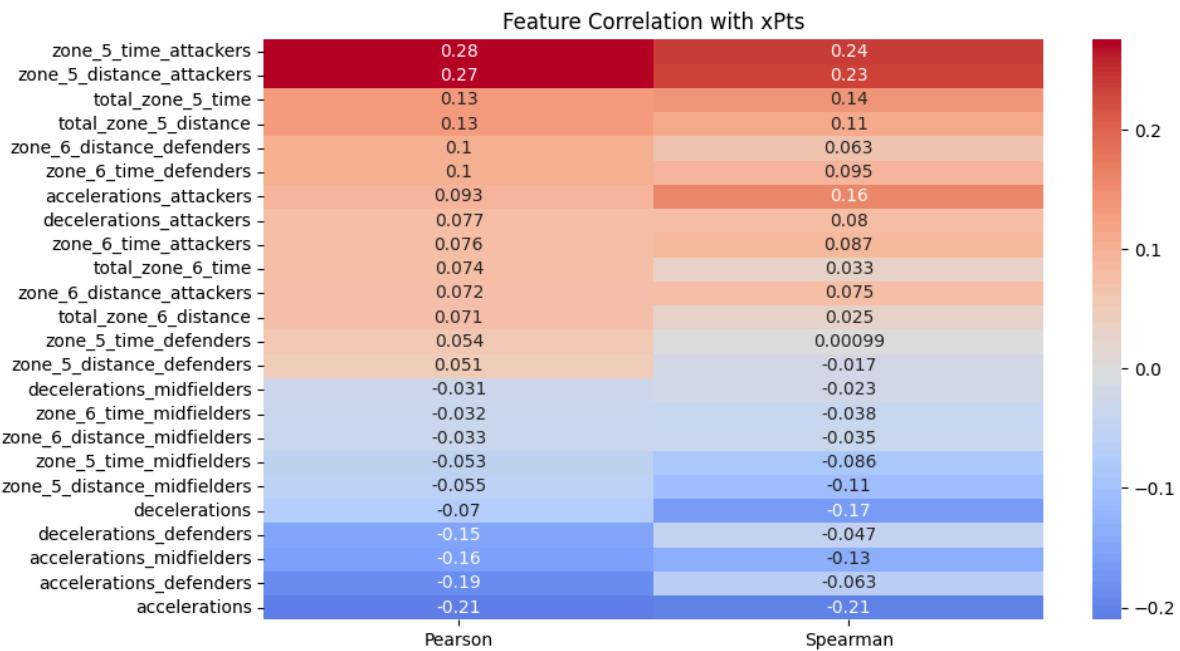


ניתן לראות שבפועל, נראה כי על הdataset הקטן שלנו ישנו פיצ'רים שהשפעתם על  $G_x$  מגיעה מהתפלגיות שונות לחלווטין, לדוגמה: `acceleration_defenders` שלו קורלציה חיובית עם  $G_x$  עבור קבוצות פלייאוף עליון אך שלילית עבור פלייאוף תחתון.

כמו כן, נראה כי עבור `total_zone_5_distance` גם כן הקורלציה שונה בסימנה בין קבוצות פלייאוף תחתון לעליון.

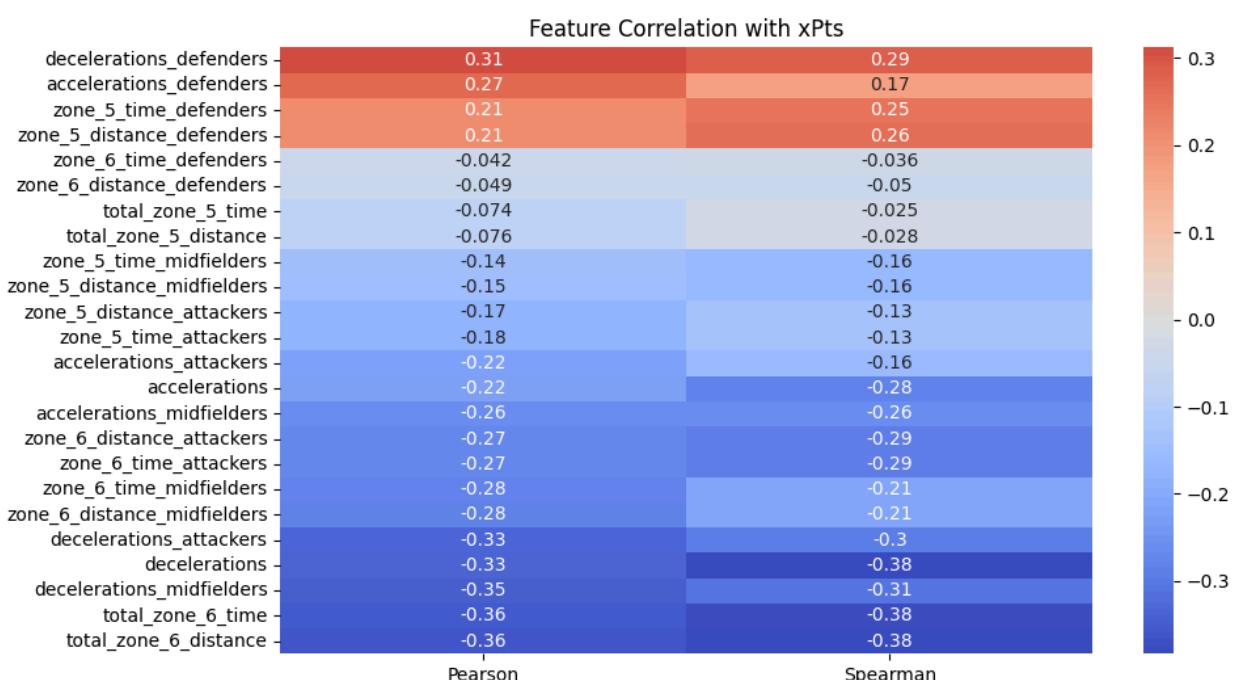
לראיינו, שימוש בעוד כ-3 עוגנות לסת הדוגמאות היה מספק בהירות רבה בהרבה על התפלגות הפיצ'רים האמיתית ומרכזת גורמי רעש שצינו וביניהם: זהות שחקנים אינדיבידואלית, שיטת משחק וכו'.

למול **קבוצות פלייאוף תחתון**, קורלצית הפיצ'רים ל-Pts-x:



זו תוצאה קוגנטיבית עם כלל המידע שראינו עד עתה ומפתחה יחסית בעוצמתה: ישנה קורלציה חיובית ביןונית-గבוהה בין ריצות באזור 5 של שחקני התקפה לבין  $\text{t}_{\text{PA}}$  וכן בין קורלציה שלילית ביןונית-גבוהה בין תאוצות של שחקני קישור והגנה לבין מספר הנקודות הצפוי. לראייתנו, למול קבוצות חלשות ישנו ערך גבוה במיוחד לריצות שחקני התקפה על מנת לנצח.

למול קבוצות פלייאוף עליון, קורלצית הפיצרים ל-Pts-x:



תוצאה זו מפתיעה בקורסילציה החיובית הבינונית - גבולה בין תאוצות ותאותות מגנים לבין Ptsx. כמו כן, קיימת קורלציה ביןונית בין זמן הריצה של שחקי הגנה באזרע 5 לבין יכולת לנצח.

ישנה קורלציה שלילית משמעותית (מעל 0.3) בין ריצה בספרינט באזרע 6 לבין Ptsx וקורסילציה שלילית ביןונית-نمוכה בין ריצה באזרע 5 של שחקי קישור והתקפה לבין Ptsx בגיןוד למול קבוצות חלשות ולמול כלל המשחקים. כמו כן, התוצאה של קשר שלילי ביןוני-חזק בין תאוצות ותאותות שחקי קישור נשמרת באופן אחיד ללא תלות בעוצמת היריבה.

יתכן שלמול קבוצות חזקות עדיפה שיטת משחק אחרת ששמה דגש על אינטנסיביות גבולה בריצה של שחקי הגנה על מנת לנצח משחקים ולא דרך ריצות מהירות של שחקי התקפה (למול קבוצות חזקות ניכר כי הקורלציה מתהפקת והיא ביןונית שלילית ביחס ל-Ptsx).

לצערנו בפועל לмерות הקורלציות המבטיחות בעוצמתן לפיצרים עבור קבוצות פלייאוף תחתון - באימון המודלים, בחלוקת לפלייאוף עליון ותחתון קיבלנו כ-30 דוגמאות בכל dataset, מה שלא מאפשר אימון מודל בצורה אמינה והביא ערכי squared R שליליים מאוד לכל מגוון השיטות השונים.

## 4.7 אימון מודלי רgresיה לחיזוי $G_x$ ו- $x_{\text{pts}}$

בשלב זה, ייצרנו מספר מודלים שונים על מנת לחזות את תוצאה  $G_x$  כתלות בפיצ'רים הקיימים. ניסינו מספר סוגי מודלים:

1. רgresור לינארי עם פונקציות loss שונות: Net Regressor
2. עץ רgresיה
3. אנסמבל עצים: Random forest, xgboost

ביצענו חלוקת train-test סטנדרטית ועבור חיפוש היפר פרמטרים של המודלים השונים (כולם מלבד MSE Regressor) השתמשנו ב-Grid Search המחלק את ה-train לסטים של validation, train לבסוף, מדדנו את ביצועי המודלים שלנו במדד MSE, MAE, MSE squared על ה-test.

עבור ה dataset הסוכם מחציות (121 דגימות סה"כ) לא הגיעו לאף מודל שקיבל ערך R squared חיובי מעל 0.1 ש מבחינותנו היה סוף מובהקות סביר ולכן החלטנו להשתמש dataset של משחיקים שלמים עבור המשך הניסויים במודלי למידה.

להלן נציג את המודל המצליח ביותר מבחינת שגיאה על test שייצרנו וננתח את חשיבות הפיצ'רים במודלים שונים.

#### 7.4.1 רגרסור לינארי

עבור הניסויים על dataset הסוכם משחקים מלאים (שלهم מראש ראיינו קורלציה גבוהה יותר ביחס ל-G<sup>x</sup>), באופן מפתיע המודלים שניבאו באופן טוב ביותר את ה-test היו מודלי רגרסיה פשוטים:

**עבור G<sup>x</sup>:** הגענו עם רגרסור לינארי (עם MSE סטנדרטי) לערך R squared של 0.4 ושגיאת חיזוי ממוצעת של 0.5 G<sup>x</sup> עבור משחק שלם (על test dataset). הגבלנו מודל זה להשתמש רק ב-12 הפיצ'רים הקורלטיביים ביותר ל-G<sup>x</sup> לינארית (Pearson):

```
zone_5_distance_midfielders', 'zone_5_time_midfielders',
'zone_6_distance_midfielders', 'zone_6_time_midfielders',
'zone_5_distance_attackers', 'zone_5_time_attackers',
'zone_6_distance_attackers', 'zone_6_time_attackers',
'accelerations', 'decelerations', 'accelerations_midfielders',
'decelerations_midfielders'
```

מבח'יננו זו הצלחה מרשימה שכן לחזות את ערך G<sup>x</sup> של קבוצה רק מנתונים פיזיולוגיים אינה ממשימה טריומאלית, שהרי יש גורמים טכניים נוספים רבים המשפיעים על הפרודקטיביות התקפית של קבוצה כאשר את חלקם ציינו. מנקודת מבטו: עומס משחק הקבוצה בנקודת זמן, איכות המסירה של שחקני הקישור, זמן משחק משותף של שחקני הקבוצה (המשפיע רבות על תיאום) ועוד.

**עבור Pts<sup>x</sup>:** לא הצלחנו להגיע בשיטות השונות למודל שימושי ערך R squared Chiobi של 0.05 ואנו משערים שהדבר נובע מהקורלציות החלשות יותר שקיימות אליו. יתכן שעם הגדלת dataset ניתן היה להגיע למודל מדויק יותר.

ראינו, הסיבה בגינה מודל רגרסיה פשוט היה המצליח ביותר במקרה זה ונבעה מכך שהוא כמעט מושך יחסית של 62 משחקים ולכן מודלים יחסית חלשים לא נטו לבצע Overfit על המידע שהוא קיים ב-train dataset שלהם. טיעון זה מתחזק לאור העובדה שמודל הרגרסיה הטוב ביותר שלא הוגבל במספר הפיצ'רים שלו הגיע למדד R squared של 0.3 (ושגיאת ממוצעת של 0.62 לניבוי G<sup>x</sup>).

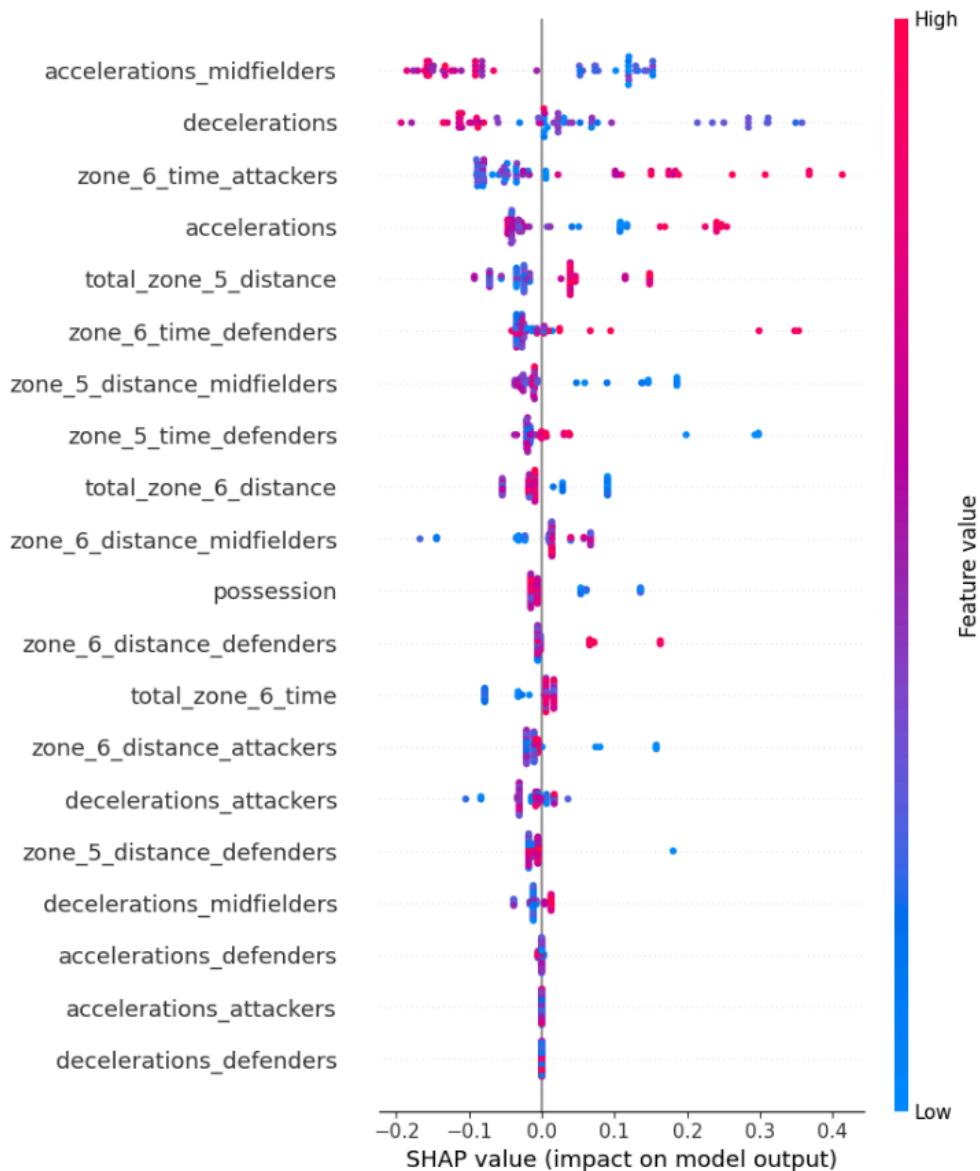
בנוסף, הדאטה הפיזיולוגי שהיא בידינו ככה"נ מספיק מובהק כדי לנבأ G<sup>x</sup> כמעט פרודקטיביות התקפית של מכבי חיפה לאורך 2 העונות שנבדקו, אך אינו מצליח לתפוס את המרכיבות הגדולה יותר של Pts<sup>x</sup> שנועד לנבأ מהי התוצאה הסופית של המשחק. הנה'ל מתישב עם העובדה שהקורלציות של הפיצ'רים פחות מובהקות ביחס אל Pts<sup>x</sup> כאשר מסתכלים על כלל המשחקים ללא חלוקה לפלייאוף עליון ותחתון.

## 7.4.2. יער רנדומי (Random Forest)

עם גַּם ערך Random Forest הגענו לערך squared R של 0.3 ושגיאת חיזוי ממוצעת של 0.57 גַּם עבור משחק שלם (על test dataset).

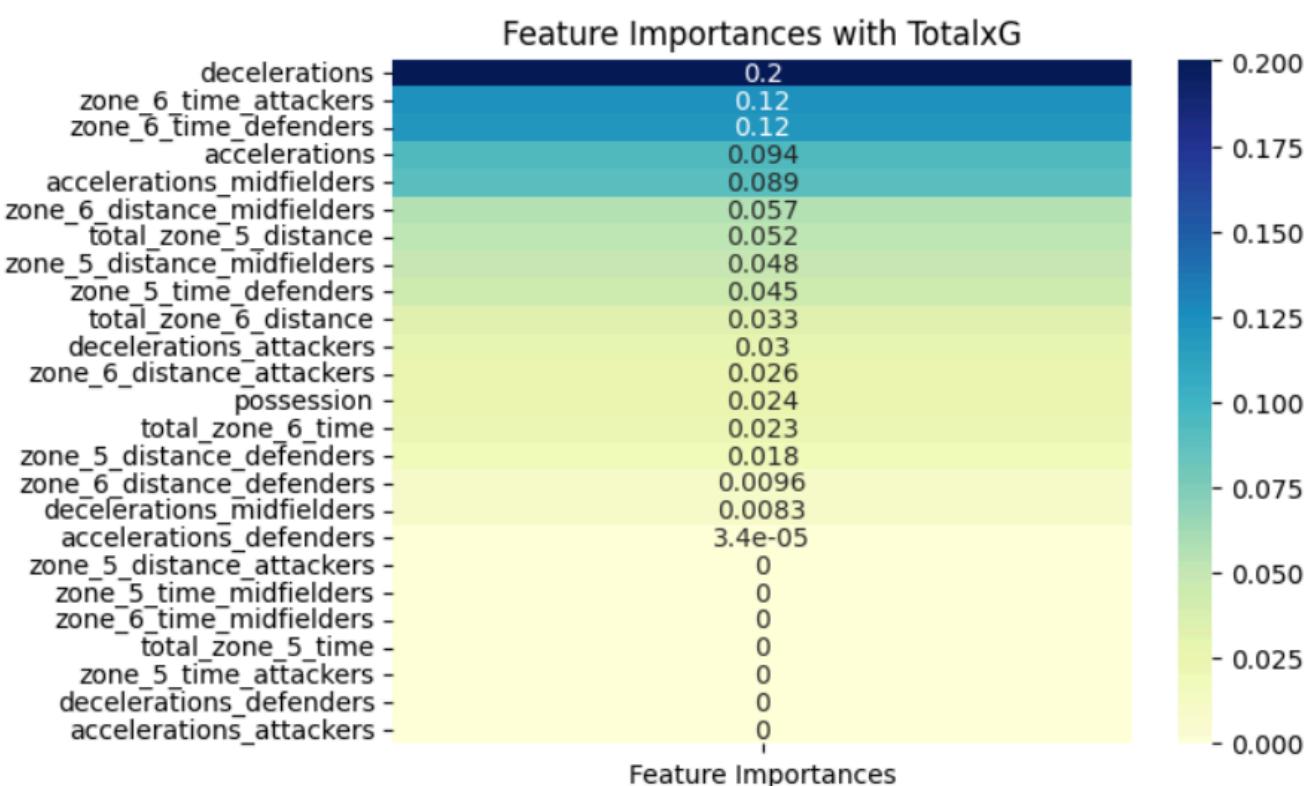
המודל משתמש ב- 5 עצי החלטה שעומקם עד 3, אלו פרמטרים שנמצאו במסגרת Grid Cross Validation על train dataset כפי שתיארנו מעלה. זהו מודל טוב יחסית ולכן התעניינו בהבין את חשיבות הפיצרים כפי שקיים בו. עם זאת, יש לקחת ניתוח זה בעירובן מוגבל שכן בוצע על dataset קטן כאמור וערך squared=0.3 R אינו גבוה במיוחד.

ניתוח (SHAP) מTARGET את השפעתו של שינוי feature מסוים על הערך הנוכחי של המודל ע"י הרצת אבולוציות שונות של שינוי המתקובל בערכי החיזוי של המודל כתלות בשינוי ערכי הפיצר הנבדק. עבור המודל שלנו:



ניכר כי המודל מיחס דוקא את החשיבות הגדולה ביותר למספר התאוצאות שביצעו שחקני קישור ולמספר התאוצאות שביצעו במהלך ע"י כלל השחקנים. הנ"ל בהלמה לקורלציות המונוטוניות והlienarיות של הפיצ'רים כפי שראינו.

לאחר מכן, זמן ריצה של שחקני התקפה באזורי 5 ו-6 מהווים פיצ'רים חשובים מאוד למודל. הנ"ל בהלמה גם ל-*feature importance* שהתקבל שהוא בעצם הלימה לערכים המוחלטים של הקורלציות שראינו בניתוח הסטטיסטי - דבר המחזק את המסקנה כי אכן ישנו קשר בין ריצות בmahioriyot 5 ו-6 לבין יכולת ההבקעה של הקבוצה. כמו כן, נבחן שהמודל כלל לא בחר בפיצ'רים מסוימים כי להבנתנו - אין בהם מידע נוסף מלבד המידע הקיים בפיצ'רים שבחר.



## 8 מסקנות

ישנו קשר ביןוני - חזק בין ריצה באזוריים 5 ו-6 של שחknim לבין Gx, ביחוד של שחknim התקפה והגנה. ישנו היפוך למול דатаה סט קטן (30 דוגמאות) של משחקים מול קבוצות פלייאוף עליון עבור ריצות שחknim התקפה בלבד. ניתן להסיק כי על מכבי חיפה בפרט ועל קבוצות הדרוגל בישראל ככל לחשוב כיצד ניתן לאמן את שחknim על מנת לפתח מהירויות גבוות לאורך זמן ומרחק רב כדי למקסם את יכולת לנצח.

כמו כן, ישנו קשר שלילי ביןוני - חזק בין מספר התאותות והתאותות שմבצעים שחknim קישור לבין Gx. זהו ממצא מעניין ועשוי להעיד על כך שאם נדרשים שינויים כיוון רבים ומהירים מצד קשרים - לקבוצה סיכוי גבוה יותר להפסיד (אולי במסגרת תנוועה לקבלת כדור ואולי מסיבות אחרות).

מגמות אלו במדוק נצפו גם במדד  $S_{Px}$  אך בצורה פחותה. (קורלציה לינארית של  $c-0.2$  במקום  $0.3$  עבור Gx). כמו כן, יכולת הניבוי של המודלים השונים עבור  $S_{Px}$  הייתה כאמור פחותה משל Gx.

אנו מדגישים כי נקודת חולשה של הממחקר היא שבוצעו על יחסית מעט דатаה (62 משחקים) והוא מותאם ספציפית למכבי חיפה שמנמה נלקח עיקר המידע.

בחולק של ניתוח לפי עוצמת יריבה (פליאוף עליון / פלייאוף תחתון) - בכל dataframe הינו דגימות עבור כ-30 משחקים בכל אחד ולראיתנו לאו דווקא ניתן להכליל את המסקנות ממנה.

לראייתנו, שימוש בעוניות נוספות ובפריסור שייצורנו של קבצי המידע יכול להוביל לחיזוק או חילשת המסקנות ויצירת מודל גרסיה החוצה Gx בצורה טובה משאצלחנו לייצר (אולי אף מודלים מתקדמים מכך). להערכתנו, שימוש בכ-5 עוניות לביצוע ניסויים דומים לשילנו יסייעו לרמת ודיות טוביה בהרבה (סדר גודל של לפחות 150 דגימות משחקים מלאים).

## **9. ציונים להמשך**

חלוקת תפקידי השחקנים לשחקני קו ושחקני מרכז מגרש עבר פיצרים במקום עבר  
עמדת לפי תפקיד.

שימוש במידע נוספת משחק נוספות ומקבוצות נוספות על מנת לחזק או להחליש  
את המסקנות שהתקבלו מן המחקר.

שימוש במידע GSP-י של קבוצות יRibot על מנת להשתמש במדדים פיזיים מדויקים  
של שתי הקבוצות לחיזוי וקורסיבציה עבר PtGx. היתכן שהקבוצה שרצה יותר  
באזרחים 5 ו-6 מנצחת באופן עקוב?

## **10 הצהרה בנוגע לשימוש במודלי שפה**

השתמשנו ב-5 chat GPT 4,5 ל课文 עזרה בבנייה המחברות עבור הניסוי שמבצע  
לולאה לבדיקת מודלי גרסיה שונים.

בנוסף, השתמשנו ב-5 chat GPT 4,5 על מנת ליצור את גודל ופורמט ההפסה של  
פונקציית היזואלייזציה שבנוינו (project\_animation).

## **11 תודות**

פרויקט זה בוצע בהנחייתו של:  
מר דין צדוק, מהפקולטה למדעי המחשב בטכניון.

וביעוזו של:  
מר אורי הראל, מאמן המקצועי הראשי של מועדון הcadorgal מכבי חיפה.

הפרויקט לא היה מתאפשר ללא שיתוף הפעולה הפורה של מועדון הcadorgal מכבי  
חיפה, אשר סייפק גישה לנוטונים, תמייה מקצועית וליווי לאורך שלבי העבודה השוניים.  
אנו מבקשים להודות גם לטכניון, על המסגרת האקדמית, המשאבים והתמייה  
שהוענקו במהלך הפרויקט.

תרומת הנוטונים, הידע והניסיונו של כל אחד מן הגורמים הללו היו בסיסי משמעותית  
להצלחת הפרויקט, ואני מוקירים תודה عمוקה על המעורבות, העזרה והאמון.

## 12 ביבליוגרפיה

<u>What Is Expected Goals (xG)?   Opta Analyst</u>	.1
<u>reaction time aspects of elite sprinters</u>	.2
<u>מכבי חיפה - Sofa Score</u>	.3
<u>MDIDAT תאוצאות בספורטאים - STATSports</u>	.4
<u>מדד mutual information</u>	.5