

## סיכום הפרוייקט

### רקע

הישגיות של קבוצה בדרך כלל מאופיינת על ידי העובדה ש"הסכום גדול מסך כל חלקיו".

לכן, אנו צופים ש"סכימה" פשוטה של רמת יכולת השחקנים לא תתאר במדויק את רמת היכולת של הקבוצה. מתוך כך, סביר להניח שלהרכב הרמות התוך-קבוצתי יש גם כן השפעה על הישגיות הקבוצה.

נשים לב שגם לאופי העבודה יש השפעה על אופטימליות פיזור רמות היכולת התוך-קבוצתית.

מצד אחד, כאשר ניתנת לקבוצה משימה שפתרונה יכול להגיע משחקן אחד, הנצחון תלוי בשחקן עם היכולת הגבוהה ביותר. לדוגמא – מציאת פתרון לחידה לוגית.

מצד שני, כאשר ניתנת משימה שמחולקת לתתי-משימות קטנות יותר, והצלחתה תלויה ברמת הביצוע של כל פרט ופרט, נשאף באופן טבעי להומוגניות ברמת היכולת. לדוגמא, בניית פרויקט טכנולוגי בקנה מידה גדול, שדורש מכל פרט בקבוצה לבצע את משימתו בצורה היעילה והטובה ביותר.

משחק הכדורסל הינו משחק קבוצתי וכיום יש מגוון רחב של קבוצות וסגנונות משחק. בפרוייקט שלנו ננסה לאפיין את המרכיבים הדומיננטיים של קבוצה שתורמים הכי הרבה להצלחתה.

בליגת ה NBA משתתפות 30 קבוצות, כל קבוצה משחקת 82 משחקים במהלך העונה הסדירה.

מספר השחקנים הפעילים בעונה בכל הקבוצות יחד נע בין 500-600.

הניתוח שלנו בוצע רק על משחקי העונה הסדירה ולא על משחקי הפלייאוף על מנת שנוכל לנתח את המידע כשאנחנו נותנים את אותו המשקל לכל הקבוצות וכך לקבל תמונה מובהקת יותר לגביי ההבדלים ביניהן.

## שאלת המחקר

**האם יש קשר בין פערי רמות של שחקנים בקבוצת NBA לבין הדירוג שלה?**

או במילים אחרות : איזה קבוצות ב NBA מגיעות להישגים טובים יותר, קבוצות מאוזנות מבחינת רמת השחקנים בקבוצה או קבוצות שיש בהן הבדלי רמות גדולים בין השחקנים ביחס לקבוצות האחרות?

- יש לציין ששאלת המחקר המקורית שלנו הייתה:  
"האם שחקנים מובילים בקבוצות משפיעים לטובה או לרעה על היכולת של חבריהם לקבוצה?"

נאלצנו לשנות את השאלה כיוון שהמידע הזמין במקורות בהם חיפשנו לא אפשר לענות עליה.

## מקורות המידע

את המידע לקחנו מהאתר <https://www.basketball-reference.com>

השתמשנו בשני סוגי מידע מהאתר:

א. טבלה שמכילה נותנים על כל השחקנים בעונה מסוימת כמו:

שם שחקן, שם הקבוצה, גיל, מספר זריקות ל 2 ו 3 נקודות, מספר אסיסטים, ריבאונדים, חסימות, איבודי כדור וכו'.

דוגמה לטבלה בה השתשמנו - נתוני שחקנים לעונה 2018-2019 :

[https://www.basketball-reference.com/leagues/NBA\\_2019\\_per\\_game.html](https://www.basketball-reference.com/leagues/NBA_2019_per_game.html)

ב. טבלה שמכילה את שם הקבוצה והדירוג שלה בעונה מסוימת.

דוגמה לטבלה בה השתשמנו - נתונים לעונה 2018-2019 (Expanded Standings) :

[https://www.basketball-reference.com/leagues/NBA\\_2019\\_standings.html](https://www.basketball-reference.com/leagues/NBA_2019_standings.html)

השתמשנו במידע של 10 העונות האחרונות 2009-2019

## תיאור כללי של הצעדים בביצוע הפרוייקט

דיון על כיצד ננתח את הנתונים בשביל לענות על שאלת המחקר :

החלטנו להפיק מדד (plus\_minus\_sd) שיתאר לנו בצורה מספרית את הבדלי הרמות שיש לכל קבוצה - לאורך 10 עונות, לעשות ממוצע של מדד זה על פני כל העונות לכל מיקום (דירוג בטבלה), ולבדוק את הקשר בין התוצאות שהתקבלו לבין הדירוג הסופי.

הסבר על אופן חישוב מדד זה:

מדד יעילות plus\_minus – מדד זה נחשב על ידי האיחוד האירופי כמדד יעיל לשקלול יכולות השחקן והישגיו.

חישוב המדד מתבצע כדלקמן:

=PIR

(נקודות + ריבאונדים + אסיסטים + חטיפות + חסימות + עבירות שנחסטו) –

(החטאות של סלי שדה + החטאות מהעונשין + איבודים + זריקות שנחסמו + עבירות שבוצעו)

ביצענו התאמה למדד זה בהתאם לנתונים שהיו זמינים במאגר שממנו שאבנו את המידע. החישוב שלנו התבצע כך:

=PIR

( נקודות + ריבאונדים + אסיסטים + חטיפות ) –

( החטאות של סלי שדה + החטאות עונשין + איבודי כדור )

- חישוב ממוצע ה plus\_minus עבור כל קבוצה.

מתבצע על ידי הממוצע של ערכי ה plus\_minus של שחקניה.

חישוב סטיית התקן של מדד ה plus\_minus לכל קבוצה

חישוב סטיית התקן על ידי הפונקציה sd() של R, תיתן לנו את הבדלי הרמות התוך-קבוצתיים עבור כל קבוצה.

קישרנו בין הנתונים החדשים שחישבנו – מדדי ה `plus_minus` וסטיית התקן של ה `plus_minus` של הקבוצה (הבדלי הרמות התוך-קבוצתיים) לבין הנתונים על דירוג הקבוצות. השאלה העיקרית היא כיצד פערי הרמות בקבוצה משפיעים על הדירוג – שהוא המאפיין הכי חזק בהצלחת קבוצה בסוף כל עונה.

בנוסף נעזרנו בנתונים על מנת לבדוק האם ישנם פרמטרים אחרים שמאפיינים את הישגי הקבוצה.

החישוב המתואר הנ"ל עזר לנו להגיע למסקנה שישנו קשר ליניארי בין רמת הבדלי היכולות של השחקנים לבין הדירוג (מספרית). ככל שיש יותר הפרשים, הדירוג בטבלה יהיה יותר נמוך מספרית – כלומר הישגי הקבוצה יהיו גבוהים יותר.

#### תיאור התוצאות

##### הקשר בין הפרשי היכולות התוך-קבוצתיים לבין הדירוג הכללי:

הזכרנו בסעיף הקודם שנמצא קשר ליניארי הפוך בין ה `plus_minus_sd` של קבוצה לבין הדירוג המספרי שלה.

מהחישובים שביצענו הגדרנו חסם עליון ותחתון עבור הערך `plus_minus_sd` על מנת להבין את התוצאות והאיורים בצורה טובה יותר.

חסם תחתון – 3

חסם עליון – 9

חשוב לציין שהחסמים תקפים רק עבור המידע הספציפי עליו ביצענו את ניתוח המידע ובאופן כללי החסם התחתון הוא אפס ואין חסם עליון.

### בדיקת הקשר על ידי ציר x,y:

באיור 1 ניתן לראות את הקשר הנ"ל.

**בכחול** מתואר קו הרגרסיה הליניארית שנעשה בעזרת הפונקציה

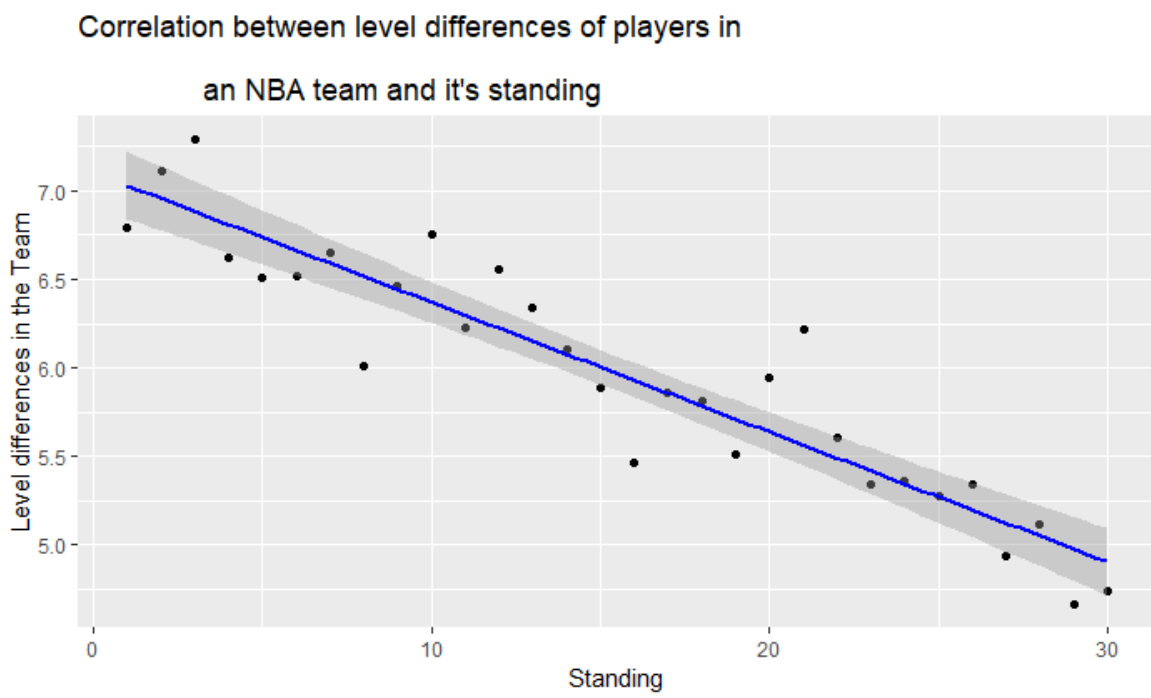
`geom_smooth` על מנת לזהות מגמה כלשהי אם קיימת.

**בציר ה-X** מתואר הדירוג המספרי של הקבוצה.

**בציר ה-Y** מתואר רמת ההבדלים הממוצעת (`plus_minus_sd`) המתאימה

לכל דירוג בציר ה X .

### איור 1



### בדיקת הקשר על ידי PCA:

דרך נוספת לבדוק את הקשר היא על ידי אלגוריתם PCA.

כך נוכל לראות בצורה ויזואלית את הקשרים בין הפרמטרים השונים.

ניתן לוודא מהימנות האיור הזה בכך שהפרמטר `plus_minus_team`

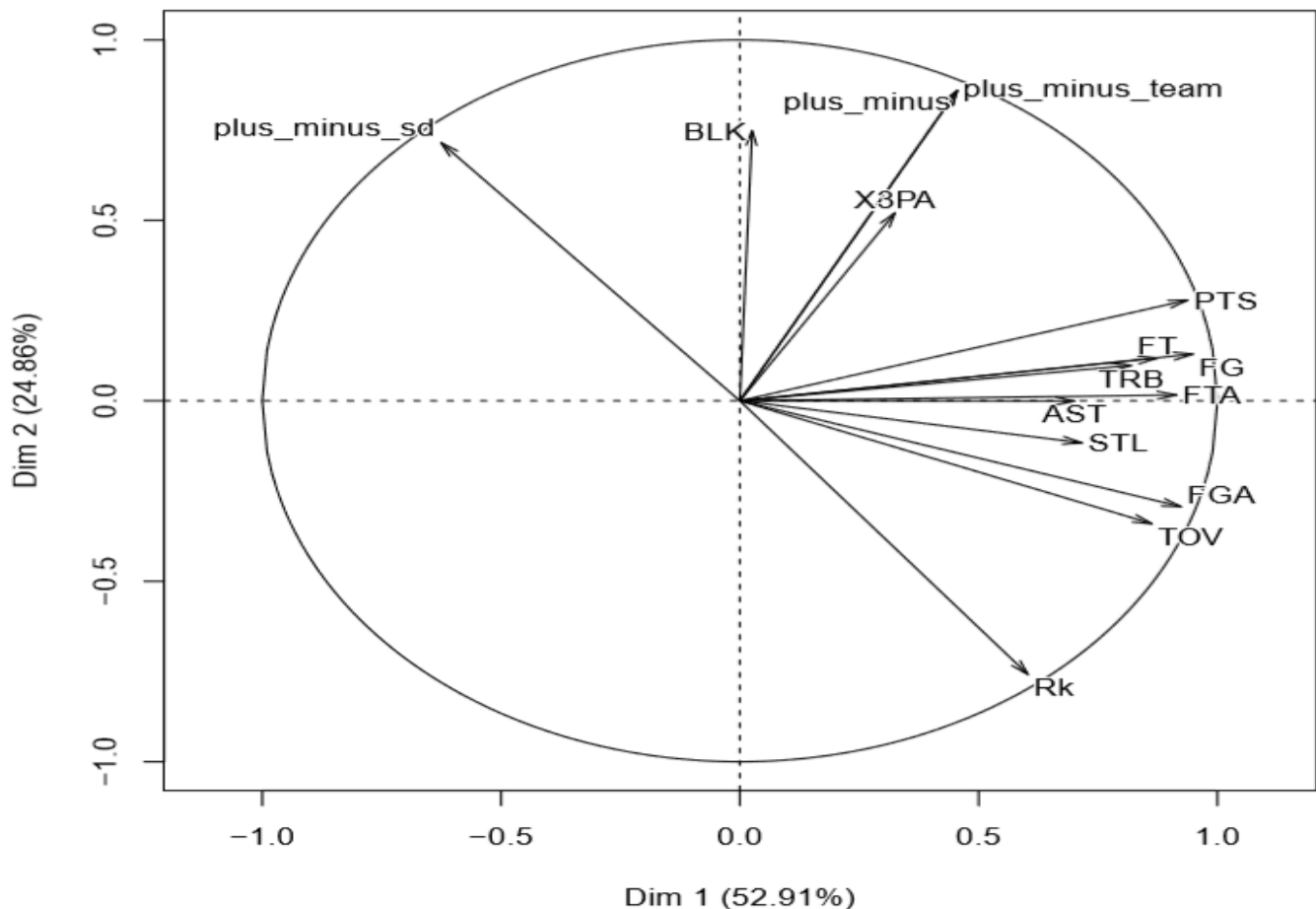
והפרמטר `plus_minus` נמצאים בקורלציה של 1 כלומר הם נמצאים באותו

הכיוון בדיוק, שכן האחד נגזר מהשני. מופיע בעמוד הבא באיור 2.

## מקרא של הפרמטרים באיור:

- Plus\_minus – המדד שמתאר את יכולת השחקן על פי התקן שהגדרנו.
- Plus\_minus\_team – ממוצע מדדי ה plus\_minus של שחקני הקבוצה.
- Plus\_minus\_sd – סטיית התקן של מדד ה plus\_minus\_team.
- Rk – הדירוג של הקבוצה בסוף העונה.
- BLK – מספר חסימות.
- X3PA – מספר נסיונות לזריקה של 3 נקודות.
- PTS – מספר נקודות.
- FT – מספר זריקות חופשיות.
- FG – מספר קליעות שדה.
- TRB – מספר הריבאונדים.
- FTA – מספר נסיונות לזריקות חופשיות.
- AST – מספר אסיסטים.
- STL – מספר החטיפות.
- FGA – מספר נסיונות לקליעות שדה.
- TOV – איבודי כדור (turnovers ליתר דיוק)

## איור 2:



## דיון בתוצאות:

כעת נדון בתוצאות וכיצד הסקנו אותן מהנתונים וניתוחם.

כפי שהוזכר מקודם, נמצא קשר ליניארי הפוך בין מדד הפרש היכולות ( $plus\_minus\_sd$ ) לבין הדירוג (מספרית). כלומר, ככל שה- $plus\_minus\_sd$  יותר גבוה, כך הדירוג המספרי יותר נמוך, משמע שהקבוצה הצליחה יותר. ניתן להסיק זאת מהתבוננות על קו הרגרסיה הליניארית באיור 1.

בנוסף, ניתן להסתכל על איור 2 ולהבחין בקשר הנ"ל באיור זה הפרמטר  $plus\_minus\_sd$  והפרמטר  $R_k$  נמצאים בכיוונים מנוגדים אחד מהשני, וניתן לאשר שאכן קיים קשר הפוך ביניהם.

ניתן לנסות לחפש עוד פרמטרים שיכולים להשפיע על הדירוג ( $R_k$ ), אך לא נראה שישנם כאלה שמשפיעים על הדירוג בצורה מובהקת.

מגישים: רון מלמד ואבישי אופיר