# פרויקט ביג דאטה –

# רקע:

# שאלת מחקר – איזה סטטיסטיקות מנצחות משחק כדורסל.

בעבודה זו נחקור את האפשרות שישנם סטטיסטיקות מסוימות שאם ננתח אותם נוכל לדעת מה תהיה תוצאת המשחק.

# תיאור המידע:

את המידע בו עשינו שימוש אספנו מהאתר: [www.stats.nba.com](http://www.stats.nba.com) - אתר המכיל סטטיסטיקות על אינספור מדדים.  
המידע שנאסף מכיל נתונים על 4 עונות (2013-2017), הנתונים כוללים:

|  |  |
| --- | --- |
| מדד | פירוט |
| TEAM\_NAME | שם הקבוצה. |
| FG3M | מספר שלשות שנקלעו. |
| FG3A | מספר שלשות שנזרקו. |
| FG3\_PCT | אחוז קליעת שלשות. |
| FTM | זריקות חופשיות שנקלעו. |
| FTA | זריקות חופשיות שנזרקו. |
| FT\_PCT | אחוז קליעת זריקות חופשיות. |
| OREB | ריבאונד מתקיף. |
| DREB | ריבאונד מגן. |
| REB | מספר ריבאונדים. |
| AST | אסיסטים. |
| STL | גניבת כדור. |
| BLK | בלוקים. |
| TO | איבוד כדור. |
| PF | פאול אישי. |
| PTS | נקודות |
| FG2M | מספר 2 נקודות שנקלעו. |
| FG2A | מספר 2 נקודות שנזרקו. |
| FG2\_PCT | אחוז קליעת 2 נקודות. |
| OUTCOME | תוצאה. |
| יתר העמודות הן אחוז המדדים לעומת הקבוצה השנייה. | |

יש לציין שחלק מהנתונים לא היו זמינים באופן ישיר ונדרשנו לחשב אותם בעזרת הנתונים הקיימים.

# תיאור כללי של הצעדים בביצוע הפרויקט:

## איסוף מידע:

בשלב זה נעזרנו ב API המאפשר שליפת סטטיסטיקות מאתר ה- NBA. הכלי נכתב בשפת Python ולכן הקוד לאיסוף המידע שלנו נכתב גם הוא ב Python.   
התוכנית שכתבנו מקבלת כארגומנט טווח של שנים ואוספת את המידע הרלוונטי לעונות בשנים האלה.  
כאשר האיסוף מסתיים התוכנית מסדרת את המידע שנאסף בקובץ csv שיתאים לעבודה עם R.

## בניית טבלאות לעבודה עם R:

## זיקוק מדדים והתמקדות במדדים הכי חזקים:

על מנת לבחון אילו מדדים הם הממדים החזקים ביותר שכדאי להשתמש בהם, השתמשנו בשיטת ה PCA (Principle Component Analysis).

## ניבוי:

# תיאור תוצאות:

# דיון בתוצאות:

# הקוד שהיה בשימוש: