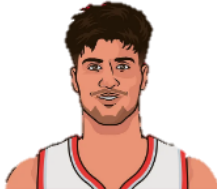


דו"ח מסכם פרויקט - קורס ויזואליזציה של מידע

1. מבוא

נושא פרויקט הוויזואליזציה: ניתוח מגמות שינוי המשחק של שחקן הכדורסל, דני אבדיה, בקריירת ה-NBA בארבע העונות האחרונות (2020-21 עד 2023-24).



דני אבדיה הוא אחד השחקנים הישראלים המבטיחים ביותר ששיחקו בליגת ה-NBA. כנציג ישראלי בליגת ה-NBA, דני מהווה שגריר של הספורט הישראלי ומקדם את התדמית של ישראל בתחום זה.

דני משמש דוגמה והשראה לספורטאים צעירים בארץ ובעולם. ניתוח ההישגים ומגמת השיפור שלו מדגים כי עם עבודה קשה, התמדה והשקעה, ניתן להגיע להישגים מרשימים. לאורך השנים, דני מציג מגמות שיפור במגוון תחומים, בפרט עלייה בנקודות ממוצעות למשחק, אחוזי קליעה משופרים, תרומה כוללת למשחק הקבוצתי ועוד.

מטרת הפרויקט היא להמחיש בצורה ויזואלית את השינויים הללו באמצעות גרפים ומסקנות.

לטובת מימוש הפרוייקט, השתמשנו בנתונים שייבאו בעזרת חבילת NBA API ועיבדנו אותם לצרכי הפרוייקט:

א. **DA_shooting_stats** - נתוני הזריקות של דני אבדיה בארבע העונות הראשונות שלו ב-NBA (2020-21 – 2023-24).

ב. **deni_stats** - נתונים של ממוצעי סטטיסטיקות (שכוללים מדדים שונים כגון: זריקות, אסיסטים, חסימות ועוד) של דני אבדיה בארבע העונות הראשונות שלו ב-NBA (2020-21 עד 2023-24).

ג. **lebron_stats** - נתונים של ממוצעי סטטיסטיקות (שכוללים מדדים שונים כגון: זריקות, אסיסטים, חסימות ועוד) של לברון ג'יימס (כיום השחקן הטוב ב-NBA) בארבע העונות האחרונות ב-NBA (2020-21 עד 2023-24).

ד. **casspi_stats** - נתונים של ממוצעי סטטיסטיקות (שכוללים מדדים שונים כגון: זריקות, אסיסטים, חסימות ועוד) של עומרי כספי (שחקן ישראלי בכיר נוסף ששיחק בעבר ב-NBA בעמדות דומות) בארבע העונות הראשונות שלו ב-NBA (2009-10 עד 2012-13).

ה. **DA_defence_offence_df** - נתונים שמכילים סטטיסטיקות ביצועים של דני אבדיה, בפילוח לרמת מספר משחק (1 עד 82), בכל אחד מארבע העונות הראשונות שלו ב-NBA (2020-21 עד 2023-24).

2. נתונים:

(1) DA shooting stats טבלת

טבלה זו מכילה 8 עמודות ו- 2267 שורות. כל שורה מסמלת זריקה של דני במהלך הקריירה:

Attribute	Munzner Attribute Type	Explanation
SEASON	Categorical Ordered	עמודת SEASON מסמלת את השנה (עונה) של הזריקה.
SHOT_DISTANCE	Quantitative	עמודת SHOT_DISTANCE מסמלת את מרחק הזריקה.
SHOT_MADE_FLAG	Ordered (Binary: 0, 1)	עמודת SHOT_MADE_FLAG מסמלת האם הזריקה נקלעה.
LOC_X	Quantitative	עמודת LOC_X מסמלת את ציר ה - x של הזריקה.
LOC_Y	Quantitative	עמודת LOC_Y מסמלת את ציר ה - y של הזריקה.
SHOT_TYPE	Categorical	עמודת SHOT_TYPE מסמלת את סוג הזריקה (כמה נקודות היא שווה).
SHOT_ZONE_BASIC	Categorical	עמודת SHOT_ZONE_BASIC מסמלת את האזור שנלקחה הזריקה.
ACTION_TYPE	Categorical	עמודת ACTION_TYPE מסמלת את אופי סגנון הזריקה.
<p>המשך ניתוח לפי Munzner:</p> <p>Data types = items = זריקה, Attributes = העמודות שפורטו DataSet type: טבלה</p>		

(2) DA defence offence df טבלת

טבלה זו מכילה 9 עמודות ו- 318 שורות. כל שורה מסמלת משחק של דני במהלך הקריירה:

Attribute	Munzner Type	Explanation
GAME_NUMBER	Quantitative Ordered	עמודת GAME_NUMBER מסמלת את מספר המשחק בעונה הסדירה (לפי שנה).
SEASON	Categorical Ordered	עמודת SEASON מסמלת את השנה של המשחק.
STL	Quantitative	עמודת STL מסמלת מספר החטיפות באותו משחק.
BLK	Quantitative	עמודת BLK מסמלת מספר החסימות באותו משחק.
offensiveRating	Quantitative	עמודת offensiveRating מסמלת את דירוג ההתקפה באותו משחק.
defensiveRating	Quantitative	עמודת defensiveRating מסמלת את דירוג ההגנה באותו משחק.
effectiveFieldGoalPercentage	Quantitative	עמודת effectiveFieldGoalPercentage מסמלת את אחוז זריקות השדה מותאם לסוג זריקה.
trueShootingPercentage	Quantitative	עמודת trueShootingPercentage מסמלת את יעילות הזריקה.
pace	Quantitative	עמודת pace מסמלת את הערכת מספר הפוזישנים של דני במשחק.
<p>המשך ניתוח לפי Munzner:</p> <p>Data types = items = משחק, Attributes = העמודות שפורטו DataSet type: טבלה</p>		

3) טבלת סטטיסטיקה לכל שחקן

לכל שחקן (דני/לברון/כספי) יצרנו אותה טבלת סטטיסטיקה המכילה 33 עמודות ומספר שורות כמס' עונות של השחקן שנדגמו (4 עונות). כל שורה מסמלת שנה של השחקן במהלך הקריירה שלו.
נפרט על 7 העמודות הבאות (העמודות החשובות): פירוט הנוגע לשאר העמודות מצורף בנספחים.

Attribute	Munzner Type	Explanation
AST_TO	Quantitative	עמודת AST_TO מסמלת את היחס בין אסיסטים לאיבודים.
USG_PCT	Quantitative	עמודת USG_PCT מסמלת את אחוז הפעמים שהשחקן מסיים את ההתקפה.
PIE	Quantitative	עמודת PIE מסמלת את השפעת השחקן למשחק (נוסחא שמוחשבת מראש).
POSS	Quantitative	עמודת POSS מסמלת את מספר התקפות של השחקן בעונה הסדירה (לפי שנה).
PTS_RANK	Quantitative Ordered	עמודת PTS_RANK מסמלת דירוג השחקן מבין כל השחקנים בכמות הנקודות בקלאץ'.
FG_PCT_RANK	Quantitative Ordered	עמודת FG_PCT_RANK מסמלת דירוג השחקן מבין כל השחקנים באחוז הקליעות בקלאץ'.
MIN	Quantitative	עמודת MIN מסמלת את מספר הדקות לממוצע למשחק.
<p>המשך ניתוח לפי Munzner: Data types = items, עונה = Attributes, העמודות שפורטו DataSet type: טבלה</p>		

3. מטלות:

א. מטלות המשתמש – שאלות עליהן הוויזואליזציה אמורה לענות:

1. דירוג ביצועים של דני אבדיה כשחקן "שלם" במטרה להבין:

1.1. את "זהות" השחקן - באילו אספקטים הוא טוב ובמה פחות.

1.2. כיצד "זהותו"/סגנון המשחק שלו השתנה לאורך העונות.

1.3. האם מגמת השיפור שלו דומה או שונה למול שחקנים אחרים (למול שחקן ישראלי עם נתונים דומים ולמול שחקן אולסטאר בעמדה דומה).

2. ניתוח ביצועי הזריקות (shooting) של דני אבדיה לאורך קריירת ה-NBA שלו עד כה

(מעונות ה-Rookie שלו ועד העונה הרביעית שלו) במטרה להבין:

2.1. מאיזה אזורים הוגא זרק, מאיזה אזורים קלע/החטיא.

2.2. מאיזה אזורים הוא הכי אפקטיבי (אפקטיבי = קולע באחוזים הכי גבוהים).

2.3. מה סוג הזריקות הכי אפקטיביות (אפקטיבי = קולע באחוזים הכי גבוהים).

2.4. האם הוא השתפר באספקט זה לאורך העונות.

3. ניתוח מגמת שינוי משחק הגנה וניתוח מגמת שינוי משחק ההתקפה:

3.1. הצגת מגמת שינוי משחק ההגנה (לפי פרמטרים המקושרים לתחום זה).

3.2. הצגת מגמת שינוי משחק ההתקפה (לפי פרמטרים המקושרים לתחום זה).

מיפוי כל מטלה למונח אבסטרקטי בהתאם לטפולוגיה של Munzner:

מטלה 1 - דירוג ביצועים של דני אבדיה כשחקן "שלם":

Action	Target
Analyse: עבור כלל תתי המטלות (1.1-1.3) - present אנחנו מציגים את הנתונים.	All Data: עבור תת מטלה 1.1 features. עבור תת מטלה 1.2, 1.3 trends – אנחנו מבצעים השוואה בין אותם מדדים בחתך לפי שנים, או השוואה בין אותם מדדים בחתך לפי שחקנים שונים, במטרה לאמוד מגמות. לדוגמה, דני מציג מגמת שיפור בביצועיו כשחקן (מגמה) כי הוא השתפר בכל 7 המדדים מעונה 2020-21 לעונת 2023-24.
Search: Lookup עבור כלל תתי המטלות - המטרה וגם המיקום של המידע ידועים. אנחנו מציגים מידע למול 7 מדדים/תכונות שהוגדרו.	
Query: עבור כלל תתי המטלות – compare עבור 1.1 אנחנו מבצעים השוואה ברמת המדדים. לדוגמה, דני מקבל ציון גבוה יותר בממד shooting לעומת play-making. עבור 1.2 אנחנו מבצעים השוואה בין עונות. לדוגמה, דני השתפר בממד ה-scoring מעונה 2020-21 לעונת 2023-24. עבור 1.3 אנחנו מבצעים השוואה בין הציונים של דני לבין הציונים של שחקנים אחרים.	Attributes: אנחנו לא מתמודדים עם תוכנה בודדת אלא עם 7 תכונות. בשלושת תתי המטלות אנחנו יכולים לבצע גם בדיקה של תלות בין תכונות גם קורולציה וגם דימיון.

מטלה 2 - ניתוח ביצועי הזריקות (shooting):

Action	Target
Analyse: עבור כלל תתי המטלות (2.1 עד 2.4) - present אנחנו מציגים את הנתונים.	All Data: עבור תת המטלות 2.1-2.3 features. עבור תת מטלה 2.4 trends – אנחנו מבצעים השוואה של אספקט הזריקות בין עונות שונות במטרה להצביע על מגמה.
Search: Lookup עבור תתי מטלות 2.1, 2.4 - המטרה וגם המיקום של המידע ידועים. Locate עבור תתי מטלות 2.2, 2.3 – אנחנו יודעים מה אנחנו מחפשים (אזור ממנו דני זורק הכי טוב, העונה בה דני זורק הכי טוב) אבל אנחנו לא יודעים איזה אזור/עונה.	

Query: עבור תתי המטלות 2.1-2.3 – summarize עבור 2.1 אנחנו מציגים את כלל המידע - כלל הזריקות (קליעות שנכנסו/החטאות) לאותה עונה. עבור תת המטלה 2.4 - compare אנחנו מבצעים השוואה של אספקט הזריקות בין עונות שונות.	Attributes: ניתן להסתכל על Attributes בשני אופנים: מצד אחד, התכונה שנמצאת במרכז המטלה היא זריקות, כלל הניתוחים שמוצגים למול תכונה זו מסבירים אותה ומציגים התפלגות של זריקות לפי תתי מאפיינים שונים (זריקות שנכנסו/החטאות, זריקות מטווחים שונים). מצד שני, ניתן להסתכל על תתי התכונות כמספר תכונות שונות, נפרדות, ולבצע למולן בדיקות ביחס של דימיון, קורולציה או תלות. לדוגמה לבדוק את הדימיון בין זריקות מאזורים שונים (5 אזורים שונים של טווח זריקה ל-3 נק').
---	--

מטלה 3 - ניתוח מגמת שינוי משחק ההגנה/התקפה:

Action	Target
Analyse: עבור כלל תתי המטלות (3.1,3.2) - present אנחנו מציגים את הנתונים.	All Data: עבור כלל תתי המטרות trends – אנחנו מעוניינים להצביע על מגמות שינוי במשחק ההגנה ובמשחק ההתקפה לאורך העונות.
Search: Lookup עבור כלל תתי המטלות - המטרה וגם המיקום של המידע ידועים.	
Query: עבור כלל תתי summarize. אנחנו מציגים את כלל הנתונים ומבצעים ניתוח מגמה למולם.	Attributes: כל גרף (הגנה/התקפה) מורכב מכמה קווי מגמות שמבוססים על כמה תכונות שונות. בכדי לענות על מטלה 3, נוכל לבצע עבור כל תת מטלה (גרף ההגנה וגרף ההתקפה) בדיקות של תלות, דימיון או קורולציה בין תכונות שמרכיבות אותם, בכדי להצביע על מגמה.

ב. רעיונות ועיצובים חלופיים:

בוויזואליזציה שלנו יש 3 גרפים. נציג עבור כל אחד מהם את הרעיונות והעיצובים החלופיים:

גרף 1: גרף שנועד להסביר את מטלה 1 - דירוג ביצועים של דני אבדיה כשחקן "שלם" והשוואה לשחקנים אחרים.

גרף נבחר – bars plot:

• **Mark** – lines (או Bars) - מייצגים את הערך של כל מדד ביצועים עבור כל שחקן.

• **Channels**:

- Position (x, y): מיקום ציר ה-X מייצג את מדדי הביצועים השונים (למשל, קליעה, ניקוד, אגרסיביות התקפית). מיקום ציר ה-Y מייצג את הערך של מדד הביצועים.
- length: מייצג את הערך של אותו משתנה (אורך יותר-ערך גדול יותר)
- Color: מאפשר הבחנה בין השחקנים השונים.

• **יתרונות:**

- מאפשר להציג ולהשוות נתונים במגוון תכונות בצורה ברורה על אותו גרף.
- ניתן לראות את ההבדלים בערכים בצורה יחסית ברורה. יותר קל למשתמש לתפוס את היחס גם אם הוא קטן.

• **חסרונות**

- כאשר יש יותר מתכונה בכל קטגוריה נוצר בלגן וקשה לעקוב מה גדול ממה.

גרף חלופי – spider plot (שם נפוץ נוסף - radar plot):

• **Mark** - בגרף Spider Plot, ה-mark העיקרי הוא line. משתמשים גם ב point mark להצגת הנתונים עצמם על הצירים.

• **Channels**:

- Position (Theta): כל ציר בגרף מייצג תכונה שונה, המיקום של כל נקודה על כל ציר מציין את התכונה שהיא מייצגת.
- Position (Radius): המרחק של כל נקודה מהמרכז של הגרף (radius) מייצג את הערך של התכונה. ככל שהנקודה רחוקה יותר מהמרכז, הערך של התכונה גבוה יותר.
- Color: מאפשר הבחנה בין השחקנים השונים.

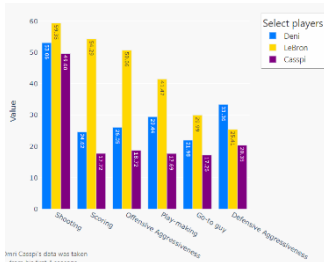
• **יתרונות:**

- גם גרף זה מאפשר להציג ולהשוות נתונים במגוון תכונות בצורה ברורה על אותו גרף.

• **חסרונות**

- קשה יותר לתפוס בעין שינויים קטנים ביחס בין 2 קטגוריות עם ערכים קרובים.
- ה-channel של זווית הוא מטבעו מחזורי (נקודת ההתחלה והסיום זהות) ופחות מתאים לנתונים לא מחזוריים.

סקיצה של הגרף:



סקיצה של הגרף:



אפקטיביות – ה-bars plot מאפשר להבחין בצורה טובה יותר מאשר ה-Spider Plot מהו היחס בין תכונה מסוימת לבין תכונות אחרות, מפני שיותר קל לבצע השוואה בין הנתונים ביחס למערכת צירים rectilinear (x,y) מאשר למערכת צירים radial.

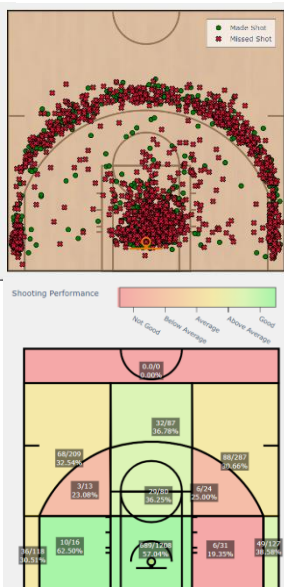
אקספרסיביות – ה-channel של זווית בגרף מסוג Spider Plot הוא מטבעו מחזורי (נקודת ההתחלה והסיום זהות) ופחות מתאים לנתונים לא מחזוריים כמו שאנחנו מנסים להציג. גרף מסוג bars plot מסביר בצורה יותר ברורה נתונים קטגוריאליים (במקרה שלנו, 7 מדדים שמהווים 7 קטגוריות).

החלטה – העדפנו להציג וויזואליזציה מסוג bar plot מכיוון שאופי הנתונים הוא קטגוריאלי ולא מחזורי. בעזרת bar plot אנחנו יכולים לענות על כלל תתי המטלות תחת מטלה 1 בצורה שמציגה ויזואלית את הנתונים באופן פשוט להבנה למשתמש.

גרף 2: גרף שנועד להסביר את מטלה 2 - ניתוח ביצועי הזריקות (shooting):

גרף נבחר - Scatter plot (on basketball court) + heatmap

סקיצה של הגרף:



• Mark –

Points :Scatter plot

Heatmap: ה-marks שמתווספים אם עוברים בתוך הגרף להצגת "אזורים חמים":

○ Rectangles(zones): חלוקה לאזורי זריקה

○ Text: מלל שמסביר על אחוזי הקליעה מאותו אזור

• Channels:

:Scatter plot

○ Horiz + verc Position (on court) – מציגים את מיקום הזריקה ביחס למגרש.

○ Shape – צורות שונות מגדירות האם זריקה נכנסה או שהחטיאו אותה.

○ Color – מימד נוסף שעוזר להבחין בין זריקה שנכנסה לזריקה שלא.

○ Distance – מימד המאפשר סיווג של הנקודות לפי המרחק שלהן מהסל.

Heatmap: ה-channel שמתווסף הוא Color – מימד שעוזר להבחין את ביצועי הזריקות מאזור מסוים.

• יתרונות:

○ ההקשר הגאוגרפי של מיקום הזריקה ביחס לסל מפשט את ההבנה של ניתוח נתונים מסוג זה (זריקות לסל).

○ מאפשר להציג נתונים רבים (אלפים), המאופיינים בריבוי תכונות בצורה ברורה על אותו גרף.

• **חסרונות** – לא ניתן להציג את כלל המידע האפשרי על תמונה סטטית אחת (גם אזורים חמים וגם התפלגות זריקות) מפני שהצגה זו יכולה להיות עמוסה ויזואלית, לכן נדרשת הבנה בתפעול הגרף במעבר בין מצב רגיל למצב "אזורים חמים".

גרף חלופי - Regular Scatter plot

• **Mark:** Points

• **Channels:**

○ **Horiz + vertc Position (x, y)** - מציגים את מיקום הזריקה ביחס לציר x, y.

○ **Shape** - צורות שונות מגדירות האם זריקה נכנסה או שהחטיאו אותה.

○ **Color** - עוד מימד שעוזר להבחין בין זריקה שנכנסה לזריקה שלא.

• **יתרונות:**

○ מאפשר למצוא מגמות, זריקות חריגות, והתפלגות של זריקות.

○ מאפשר סקלאביליות - הצגה של אלפי זריקות על גרף אחד.

• **חסרונות:** לא מאפשר להציג את הנתונים ביחס למגרש והסל ובפרט יותר קשה לענות על תת מטלה 2.2. (מאיזה אזורים הוא הכי אפקטיבי (אפקטיבי = קולע באחוזים הכי גבוהים)) ללא הצגת "אזורים חמים".

אפקטיביות - הגרף heatmap + Scatter plot (on basketball court) יותר אפקטיבי מפני שהוא מאפשר לנו לענות בצורה יחסית נוחה על כל תתי המטלות שהגדרנו תחת מטלה 2. הוא מאפשר לנו להגיע לתובנות גם ברמת הזריקות הבודדות (התפלגות זריקות) גם ברמת השוואת אזורים (בעזרת מפת האזורים החמים) וגם מאפשר לנו לעמוד מגמות בעזרת הסנן הראשי שמאפשר השוואה של נתונים בין שנים.

אקספרסיביות - בגרף Regular Scatter plot לא היינו מצליחים להציג את כלל המידע לעומת גרף heatmap + Scatter plot (on basketball court) שמאפשר זאת, אם כי באופן קצת מורכב.

החלטה - העדפנו לספק וויזואליזציה טיפה יותר מורכבת שמאפשרת לענות על כלל תתי המטלות ומספקת יכולת ניתוח גם ברמה פרטנית (התפלגות זריקות) וגם ברמת ניתוח מגמות, על פני גרף פשטני שלא מאפשר להציג את כלל המידע. בנוסף, העדפנו לקשר מבחינה גאוגרפית את הנתונים להקשר של מגרש וסל בכדי להקל על המשתמש להפיק תובנות.

גרף 3: גרף שנועד להסביר את מטלה 3 - ניתוח מגמת שינוי משחק הגנה/התקפה:

גרף נבחר - index line charts

• **Mark:** lines (נקודות שמחוברות בקווים ויוצרים קו מגמה) - כל קו מייצג מגמה של מדד ביצועים לאורך 82 משחקים (עונה ב-NBA) שמבוסס על ממוצע נע של הביצועים באותו מדד לאורך 82 משחקים.

• **Channels:**

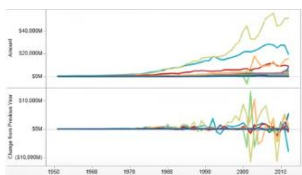
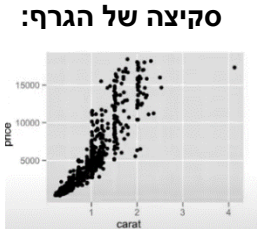
○ **Position (x, y)** - מיקום ציר ה-x מייצג את מספר המשחק (Index).

מיקום ציר ה-y מייצג את הערך של מדד הביצועים.

○ **Color** - משמש להבחנה בין המדדים השונים.

• **יתרונות:** גרף מסוג זה מותאם להציג שינוי/מגמה לאורך זמן למול מידע ממזין (לפי שדה מס' משחק).

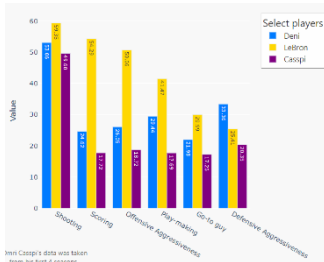
• **חסרונות:** תרשימי קו לא מספקים הרבה מידע באופן יחסי.



גרף חלופי - bars plot

- **Mark** – lines (או Bars) - מייצגים את הערך של כל מדד ביצועים לכל משחק.
- **Channels**:

סקיצה של הגרף:



- **Position** (x, y): מיקום ציר ה-x מייצג את מספר המשחק (Index).
- מיקום ציר ה-y מייצג את הערך של מדד הביצועים.
- **length**: מייצג את הערך של אותו מדד הביצועים. (אורך יותר-ערך גדול יותר)
- **Color**: מאפשר הבחנה בין השחקנים השונים.

• יתרונות:

- מאפשר להציג ולהשוות נתונים במגוון תכונות בצורה ברורה על אותו גרף.
- ניתן לראות את ההבדלים בערכים בצורה יחסית ברורה. יותר קל למשתמש לתפוס את היחס גם אם הוא קטן.
- **חסרונות**: גרף שיותר מתאים להצגת נתונים קטגוריאליים, אך הנתונים במטלה זו הם ממוינים לפי מס' משחק.

אפקטיביות – הפתרון index line charts יעיל יותר מפני שאנחנו מציגים מידע שמבוסס על שדה ממין (מס' המשחק שמהווה Index) ו-bars plot הוא דרך פחות נוחה להציג נתונים מסוג זה.

אקספרסיביות – גרף bars plot יהיה עמוס מאוד כי נצטרך לייצר עבור כל שבוע (יש 82 שבועות) bar עבור כל מדד ביצועים. באמצעות index line charts נוכל להציג את הנתונים בצורה יותר נוחה ונקייה לעין.

החלטה – העדפנו לספק וויזואליזציה פשוטה ונקייה לעין, שמתאימה לסוג הנתונים ושעונה על כל תת המטלות שהגדרנו במטלה 3. בעזרת index line charts אנחנו יכולים לתאר מגמות של שינוי משחק ההגנה וההתקפה של דני לאורך השנים, תוך התחשבות במספר מדדי ביצועים.

4. הסבר על יישום הוויזואליזציה:

א. עיבוד מקדים:

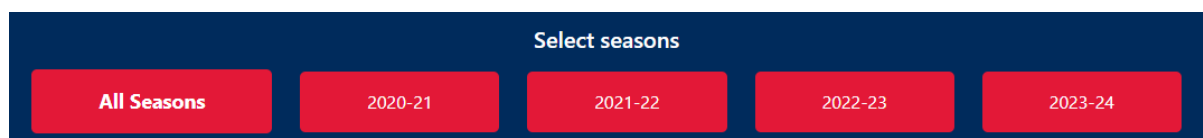
עבור טבלת - DA_shooting_stats, הורדנו את הטבלה shotchartdetail מחבילת nba_api המכילה מידע על כל הזריקות של דני (בצענו סינון רק לטבלה של דני). מחקנו עמודות לא רלוונטיות כגון מספר ID של השחקן, שם הקבוצה, שם השחקן ועוד.

עבור טבלת - DA_defence_offence_df, מיזגנו בין 3 טבלאות מחבילת nba_api הנקראות: TeamGameLog, PlayerGameLog, BoxScoreAdvancedV3 ובהן נתונים סטטיסטיים של כל משחק. מחקנו מהטבלה המאוחדת עמודות לא רלוונטיות כגון מספר מזהה המשחק, האם ניצחון/הפסד, נגד מי היה המשחק ועוד. *נדגיש כי לא הצלחנו לטעון את טבלאות גלם אלו באמצעות colab, לכן טענו אותם דרך visual code ולאחר מכן ייבאנו אותם כ-CSV ל-colab.

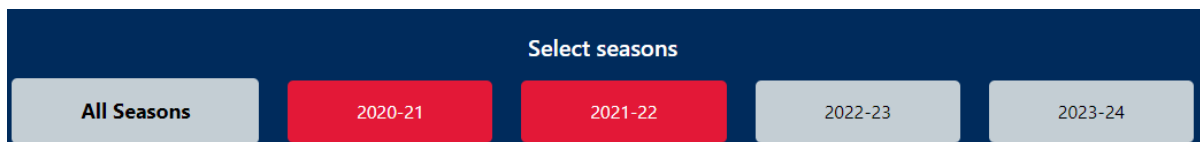
עבור טבלת הסטטיסטיקה של כל שחקן - הורדנו את הטבלה "playerdashboardbyyearoveryear" מחבילת nba_api המכילה נתונים סטטיסטיים עונתיים על השחקן. מחקנו עמודות לא רלוונטיות כגון מספר מזהה הקבוצה, תאריך למשחק האחרון, מספר ניצחונות, מספר הפסדים ועוד. בסיום תהליך העיבוד, הוספנו עוד 7 עמודות שמייצגות את תוצאות חישוב ערכי 7 מדדי הביצועים שהגדרנו.

ב. **שינוי חלופה נבחרת:** במסגרת תהליך החשיבה שקדם לבניית הוויזואליזציה הסופית, סקרנו דרכים נפוצות לניתוח והצגה ויזואלית של נתוני ספורט. אחת הדרכים הנפוצות להצגת ביצועים של שחקני כדורסל היא באמצעות spider plot. בהתאם לכך, תחילה מימשנו את גרף 1 שעונה על מטלה 1 כ-spider plot. עם זאת, לאחר שנלמד בהרצאה כי יש להימנע משימוש בגרף זה, אלא אם מדובר בהצגה של נתונים מחזוריים (על בסיס מימד זמן), בחרנו לשנות את העיצוב לגרף מתאים יותר (bar plot) שעונה על דרישות המטלה.

ג. **מדריך להפעלת הוויזואליזציה (כולל תצלומי מסך והסברים):**
בשלב הראשון **ניתן לבחור עונה (season)** לפיה יישתנו כלל הנתונים. ניתן לבחור את אחת מהעונות, קומבינציה של עונות, או את כל העונות. הבחירה תשפיע על הנתונים שיוצגו בכל הגרפים בוויזואליזציה.



לדוגמה, בחירה של נתונים שרלוונטיים לעונות 2020-21, 2021-22:



כאשר בוחרים בעונה/עונות מסוימות הלחצן הופך לאדום. עונה שלא נבחרה תהיה בצבע אפור. ברירת המחדל היא בחירה של כלל העונות (All Seasons).



*נגדיש כי הצבעים עבור הוויזואליזציה (הרקע, הכותרות, הפקדים והנתונים) נבחרו בהשראה לצבעי קבוצת ה-NBA שדני שיחק בה ב-4 העונות האחרונות Washington wizards.

בנוסף לכך, היה לנו חשוב לשלב בוויזואליזציה **הצגה של נתונים בסיסים** על דני אבדיה (כמו פרופיל השחקן והסטטיסטיקות הבסיסיות ביותר – ממוצעים עונתיים) בכדי לייצר רקע למשתמש שלא בהכרח מכיר את דני.

נדגיש כי הסטטיסטיקות של הממוצעים העונתיים גם משתנה בהתאם לעונה/עונות שבחרים:

Deni Avdija NBA Stats Dashboard

Select seasons

All Seasons

2020-21

2021-22

2022-23

2023-24

Season Averages

Player Profile

PPG	RPG	APG	PIE	HEIGHT 6'9" (2.06m)	WEIGHT 210lb (95kg)	COUNTRY Israel	LAST ATTENDED Maccabi Tel Aviv
9.6	5.9	2.4	8.5	AGE 23 years	BIRTHDATE January 3, 2001	DRAFT 2020 R1 Pick 9	EXPERIENCE 4 Years

עבור בחירה של נתונים מכל ארבעת העונות (All Seasons):

Deni Avdija NBA Stats Dashboard

Select seasons

All Seasons

2020-21

2021-22

2022-23

2023-24

Season Averages

PPG	RPG	APG	PIE
14.7	7.2	3.8	10.9

Player Profile

HEIGHT 6'9" (2.06m)	WEIGHT 210lb (95kg)	COUNTRY Israel	LAST ATTENDED Maccabi Tel Aviv
AGE 23 years	BIRTHDATE January 3, 2001	DRAFT 2020 R1 Pick 9	EXPERIENCE 4 Years

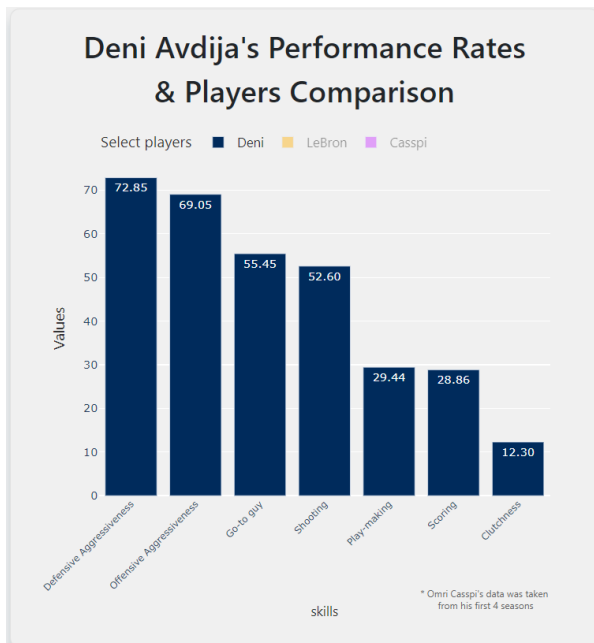
עבור בחירה של נתונים מהעונה האחרונה (2023-24):

ניתן לראות כי הסטטיסטיקות של season averages מתעדכנות בהתאם.

הגרף הראשון בוויזואליזציה הוא גרף שמתאר את "זהותו" של דני כשחקן, ומשווה אותו ביחס לשחקנים אחרים. גרף זה נועד לענות על הדרישות שהוגדרו במטלה 1. גרף זה הוא מסוג bar plot.

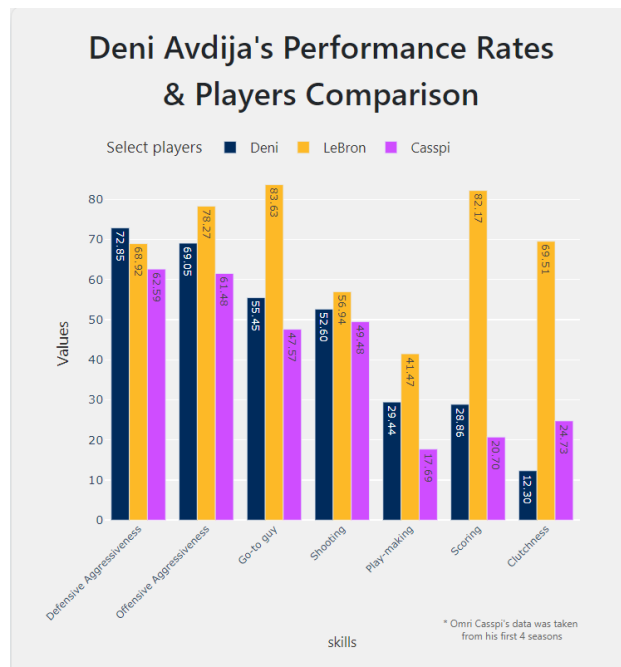
הנתונים בגרף זה מושפעים מהסנן הראשי של עונה/עונות. הגרף ממזין (orded) לפי המדד שדני הכי טוב בו. מיון זה יישאר עקבי תמיד, גם כשתבצע השוואה לשחקנים אחרים. הערכים מנורמלים לפי הציון המקסימלי שאחד השחקנים קיבל בכל אחד מהתחומים. גרף ברירת המחדל הוא גרף עמודות שמציג את הנתונים של דני בלבד למול 7 מדדי ביצועים¹, שנוגעים לתחומים שונים במשחק הכדורסל, ומאפשרים לייצר הבנה שלמה על "זהות" השחקן.

להלן דוגמה לגרף ברירת מחדל שמציג את הנתונים של דני על פני כל העונות (פקד בחירת העונה הוא all season):



בנוסף, ניתן לבצע השוואה בין הביצועים של דני אבדיה למול שחקנים אחרים:

בגרף זה מוצגת השוואה בין דני אבדיה לשחקן לברון ג'יימס ולעומרי כספי (פקד בחירת העונה הוא all season):



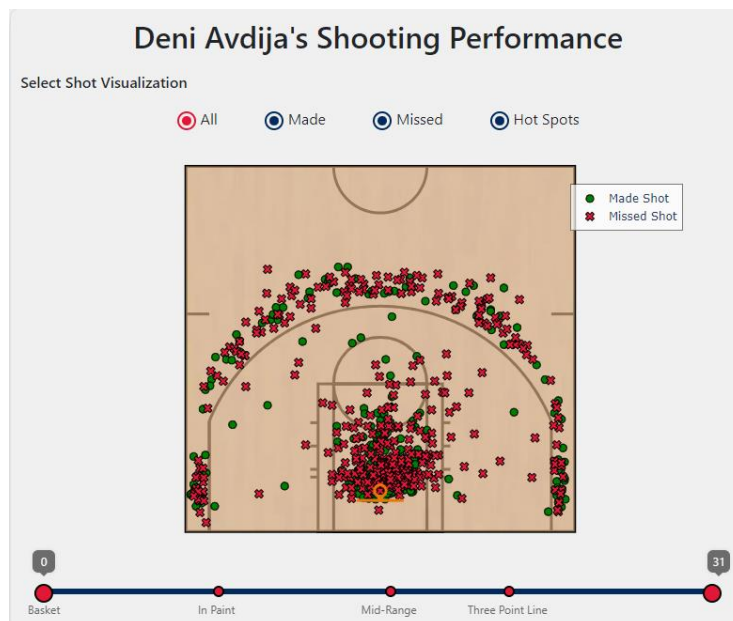
ניתן לראות שגם במצב בו אנחנו מוסיפים להשוואה שחקן שבמדדים מסויימים מקבל ציון גבוה מדני, הגרף נשאר ממזין לפי המדד בו דני מקבל את הציון הגבוה ביותר.

בחירת הצבעים עבור כל שחקן מושפעת מהקבוצה שהוא הכי מזוהה איתה (דני – וושינגטון, לברון – לייקרס, עומרי כספי – סקרמנטו).

כמו כן, ציינו בתחתית הגרף כי ההשוואה למול עומרי כספי היא השוואה לתקופה של ארבעת השנים הראשונות בהן עמרי כספי התחיל את קריירת ה-NBA שלו. (2009-10 הומר ל 2020-21 וכן הלאה).

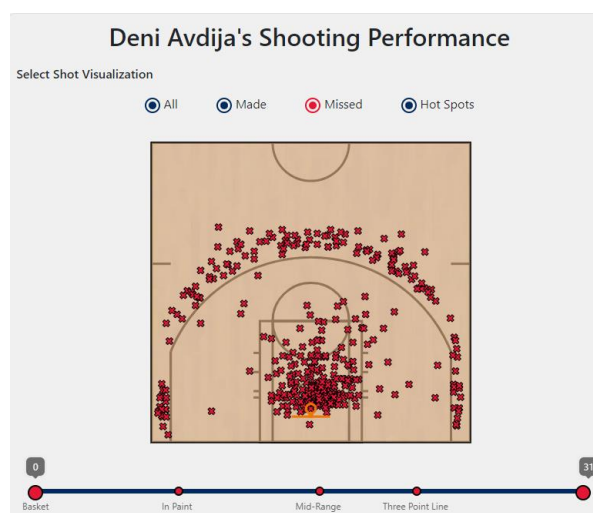
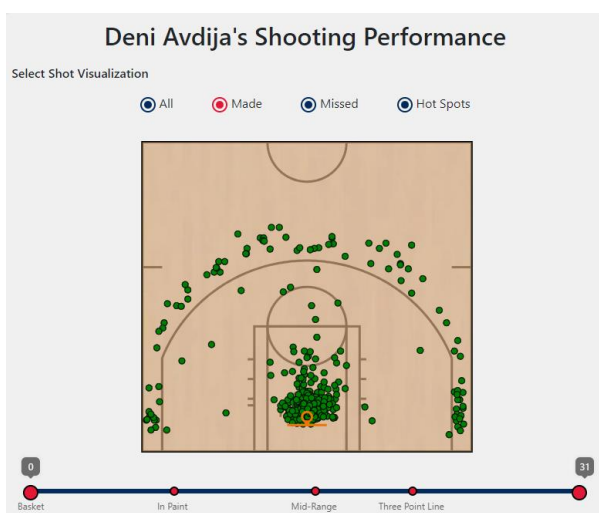
¹ הסבר על אופן חישוב מדדי הביצוע מפורט בנספח 1.

הגרף השני בוויזואליזציה הוא גרף הזריקות שנועד לענות על הדרישות שהוגדרו במטלה 2. גרף זה הוא גרף מסוג Scatter plot (on basketball court) שמתאפשר בו מעבר למצב הצגת הנתונים כ-heatmap.

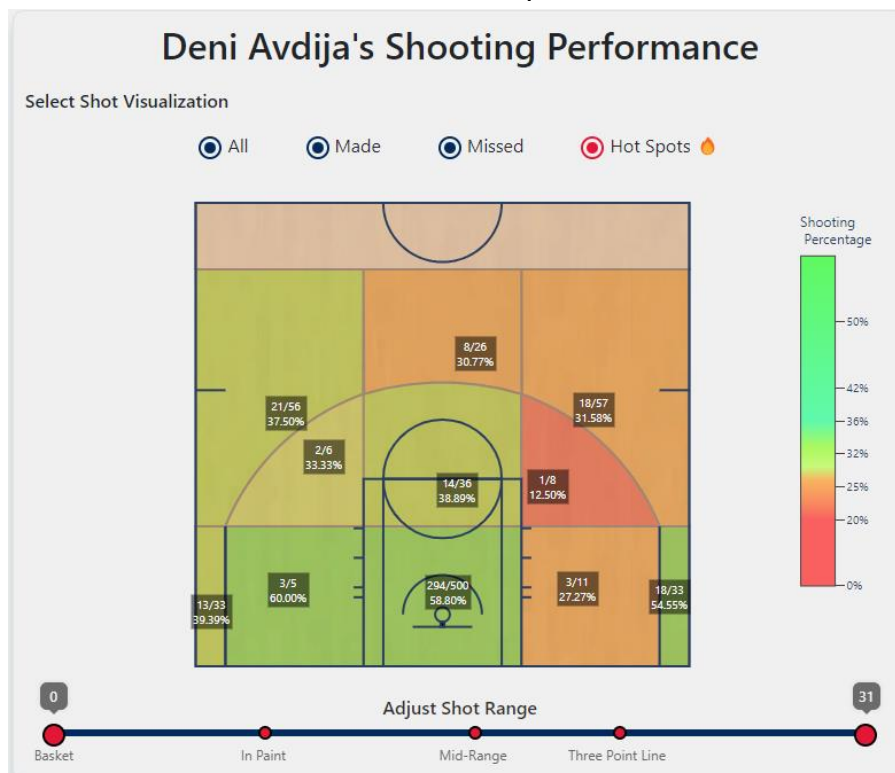


גרף זה מציג את הנתונים (בהתאם לבחירת העונה/עונות) לפי ארבעה מצבים (select shot visualization):

- 1 **All** - מצב המציג את כלל הזריקות שדני לקח, תוך הבדלה בין זריקות שהוא קלע (עיגול ירוק) לבין החטאות (איקס אדום). כפי שמוצג בתמונה למעלה. לגרף זה מצורף מקרא של סוגי הזריקות.
- 2 **Made** - מצב המציג רק את הזריקות שדני קלע.
- 3 **Missed** - מצב שמציג רק את הזריקות דני החטיא.



Hot Spot (4) - מצב שמאפשר מעבר להצגה לפי מפת חום של האזורים במגרש מהם דני קלע באחוזים הכי גבוהים. הנתונים בגרף זה מבוססים על אותם נתוני זריקות (קליעות והחטאות) שמוצגים במצבים הקודמים, אך במצב זה בחרנו לא להציג גם את הזריקות וויזואלית (את הנקודות עצמן), בכדי למנוע עומס מידע על הגרף.



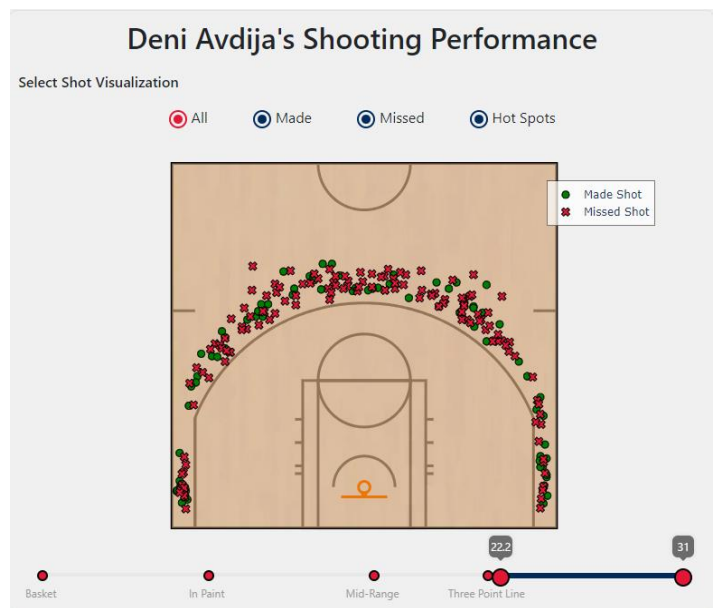
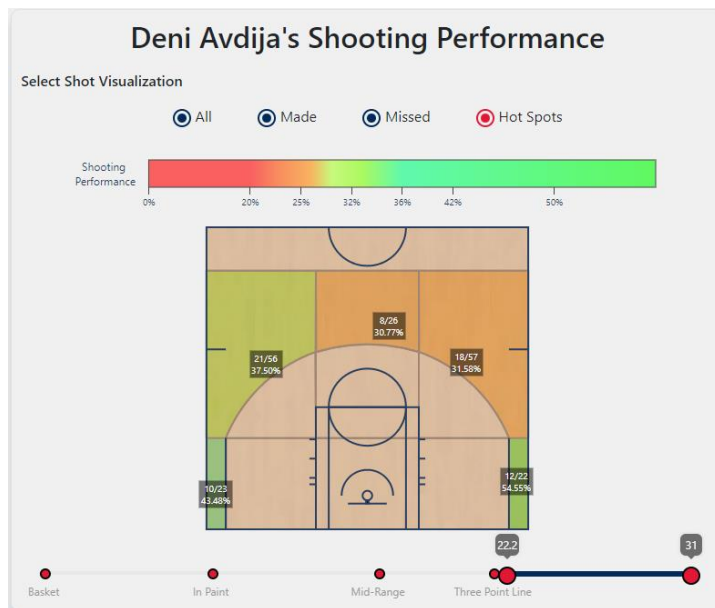
לגרף זה מצורף מקרא של סקאלת צבעים (color scale) שנהגה מגוון **אדום** לגוון **ירוק** במטרה להבחין בין אזורים טובים לאזורים טובים פחות. **אדום** מייצג אזור עם אחוזי קליעה נמוכים, **ירוק** מייצג אזור עם אחוזי קליעה גבוהים. ההנחה שעמדה מאחורי בחירת צבעים אלו היא **שאדום** הוא צבע שמתקשר עם כישלון - החטאה, **ירוק** הוא צבע שמקשר עם הצלחה - קליעה. כמו כן, מוצגים בסקאלה טווחים לפיהם נקבע הצבע (לדוגמה, אזור יצבע בצבע **כתום** אם הוא בטווח של 30%).

בנוסף, הוספנו תווית טקסט עבור כל אזור שמציגה כמה זריקות דני קלע מתוך כלל הזריקות שהוא זרק מאותו אזור, ואת היחס כאחוזים. מפת החום, צביעת האזורים והתווית מתעדכנים בהתאם לסינון הראשי של העונות.

בחרנו להוסיף לגרף השני סנן מסוג סליידר שמוצג בכל ארבעת המצבים. סליידר זה מאפשר לנו לסנן את הזריקות לפי טווחים שונים (זריקות מעבר לקשת ה-3, זריקות מטווח בינוני או זריקות מתחת לסל). אספקט זה הינו חשוב ביותר מפני שזריקות מאזורים שונים במגרש דורשות יכולות שונות. השימוש בסליידר מאפשר למשתמש לבחור בצורה נוחה את טווח המרחק שלפיו הוא רוצה לראות את הזריקות שדני זרק.

בנוסף, הוספנו לסליידר נקודות ציון משמעותיות שיאפשרו למשתמש להגדיר בקלות טווחי זריקה, בהתאם למרחקים פופולאריים (זריקות ב"צבע", שלשות וכו').

ניתן לראות כי הטווח שמוגדר בסליידר משפיע על הנתונים שמוצגים בכלל המצבים (גם במצב של הצגת פיזור הזריקות וגם במפת האזורים החמים). לדוגמה, אם נגדיר רק טווח מרחק של זריקות מעבר לקשת ה-3:



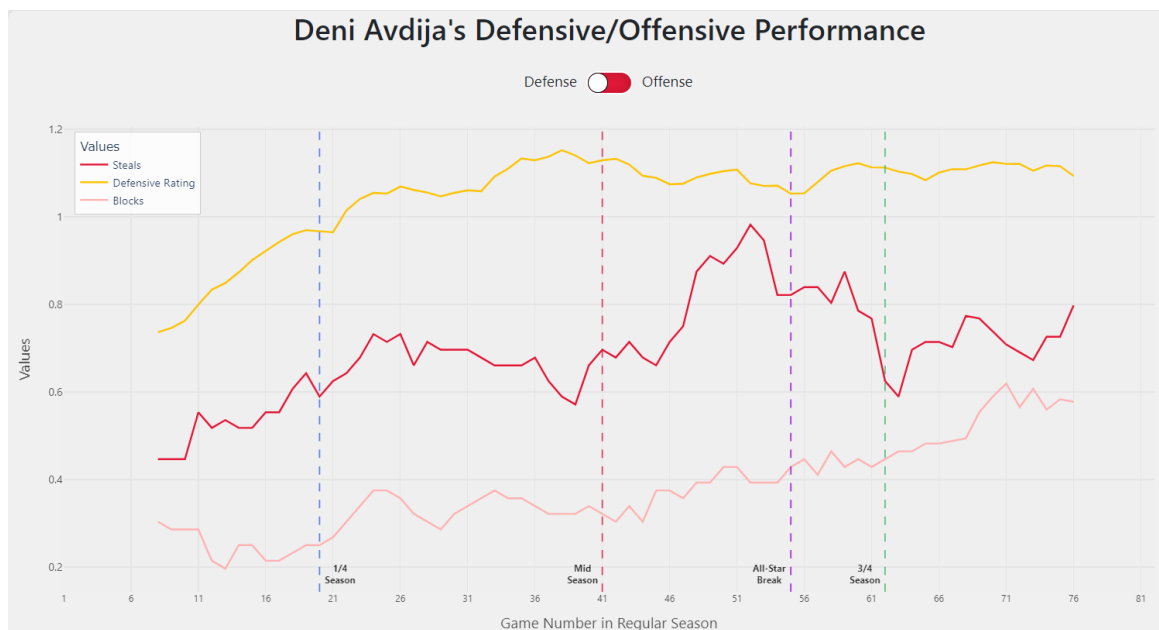
הגרף השלישי בוויזואליזציה הוא גרף שמתאר את מגמות השינוי במשחק ההגנה וההתקפה של דני. גרף זה בא לענות על תתי המטלות שהוגדרו תחת מטלה 3. גרף זה הוא גרף מסוג index line charts. בגרף זה מוגדרים 2 מצבים – מצב הגנה, המבוסס על הצגת נתונים של מדדי הגנה, וגרף התקפה שמבוסס על הצגת נתונים של מדדי התקפה. את המעבר בין שני המצבים ניתן לבצע באמצעות לחיצה על המתג (switch):



מעבר לשינוי בצבע, מיקום העיגול הלבן ב-switch משתנה לכיוון המצב שרוצים. הקווים שיוצגו בגרפים הם קווי מגמה שמורכבים ממוצע נע (לפי 14 משחקים) של ציוני המדדים. המגמה מוצגת על פני 82 משחקים (מספר המשחקים בעונה הסדירה), במטרה להצביע על מגמה עונתית (שנתית). ציר ה-X בגרף מתאר את מספר המשחקים בעונה הסדירה (מ-1 עד 82), וציר ה-Y מתאר את ערך המדד.

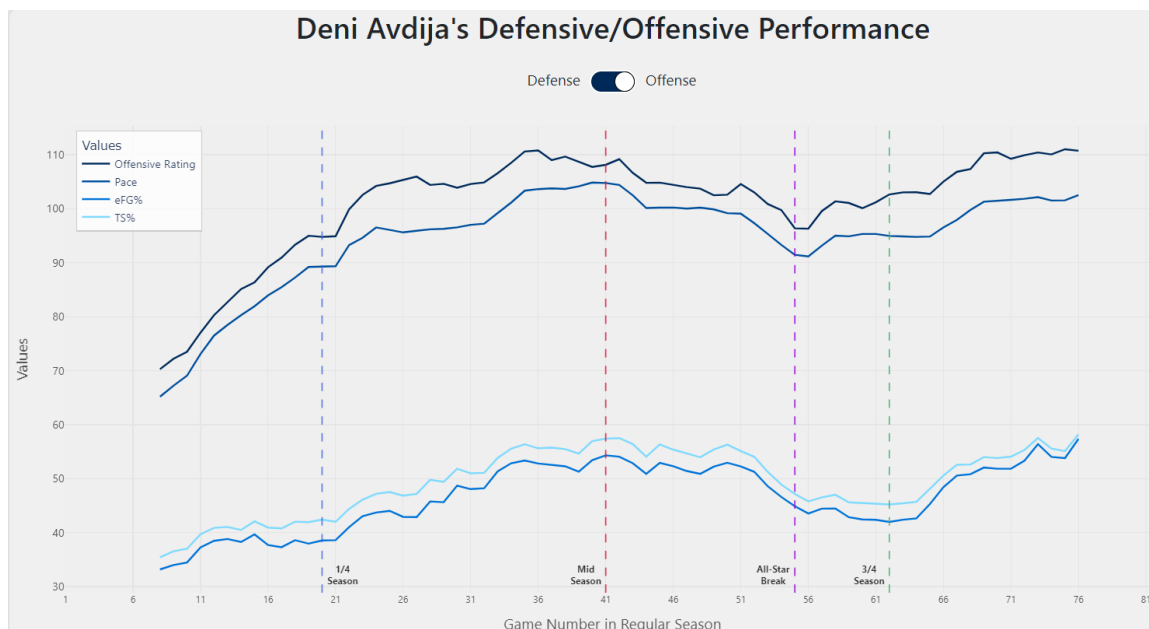
באמצעות הצגת מספר קווי מגמה של מדדים שונים, ניתן להגיע להבנה יותר עמוקה על שינויי המגמה הכללית של השחקן לאורך העונה. גרף זה מושפע גם הוא מסנן העונות הראשי, כלומר ניתן לראות את מגמת שינוי המשחק בהגנה/התקפה גם בחתך של קומבינציה בין עונות או למול 4 העונות ביחד. בנוסף, הוספנו ייצוג של נקודות ציון משמעותיות בעונה סדירה (1/4, אמצע העונה, שבוע האולסטאר, 3/4 עונה) כקווים מקווקוים קבועים בצבעים שונים.

מצב הגנה:



ניתן לראות בגרף זה קווי מגמה של שלושה מדדים שמקושרים עם משחק ההגנה (חטיפות, דירוג הגנה, חסימות). בחרנו צבעים "חמים" (גווני אדום – צהוב) לייצר עוד סממן שיגרום למשתמש להבחין שמדובר בנתונים שקשורים למשחק ההגנה.

מצב התקפה:



ניתן לראות בגרף זה קווי מגמה של ארבעה מדדים שמקושרים עם משחק ההתקפה (דירוג התקפה, כמה דני משפיע במשחק שלו על קצב המשחק הכללי, אחוז זריקות אמיתיות, אחוז שימוש בשחקן כשהוא על המגרש). בחרנו צבעים "קרים" (גווני כחול) לייצר לנו עוד סממן שיגרום למשתמש להבחין שמדובר בנתונים שקשורים למשחק ההתקפה.

ד. הסבר על כתיבת הקוד

השתמשנו ב Python בכדי לכתוב את קוד הוויזואליזציה.

בנוסף השתמשנו בספריות הבאות:

- Dash - לצורך בניית האפליקציה והעיצוב.
- Plotly - לצורך בניית הגרפים.
- NBA API – בכדי לגשת לנתונים הרלוונטיים.
- Pandas, Numpy, Regex – לטובת עיבוד הנתונים

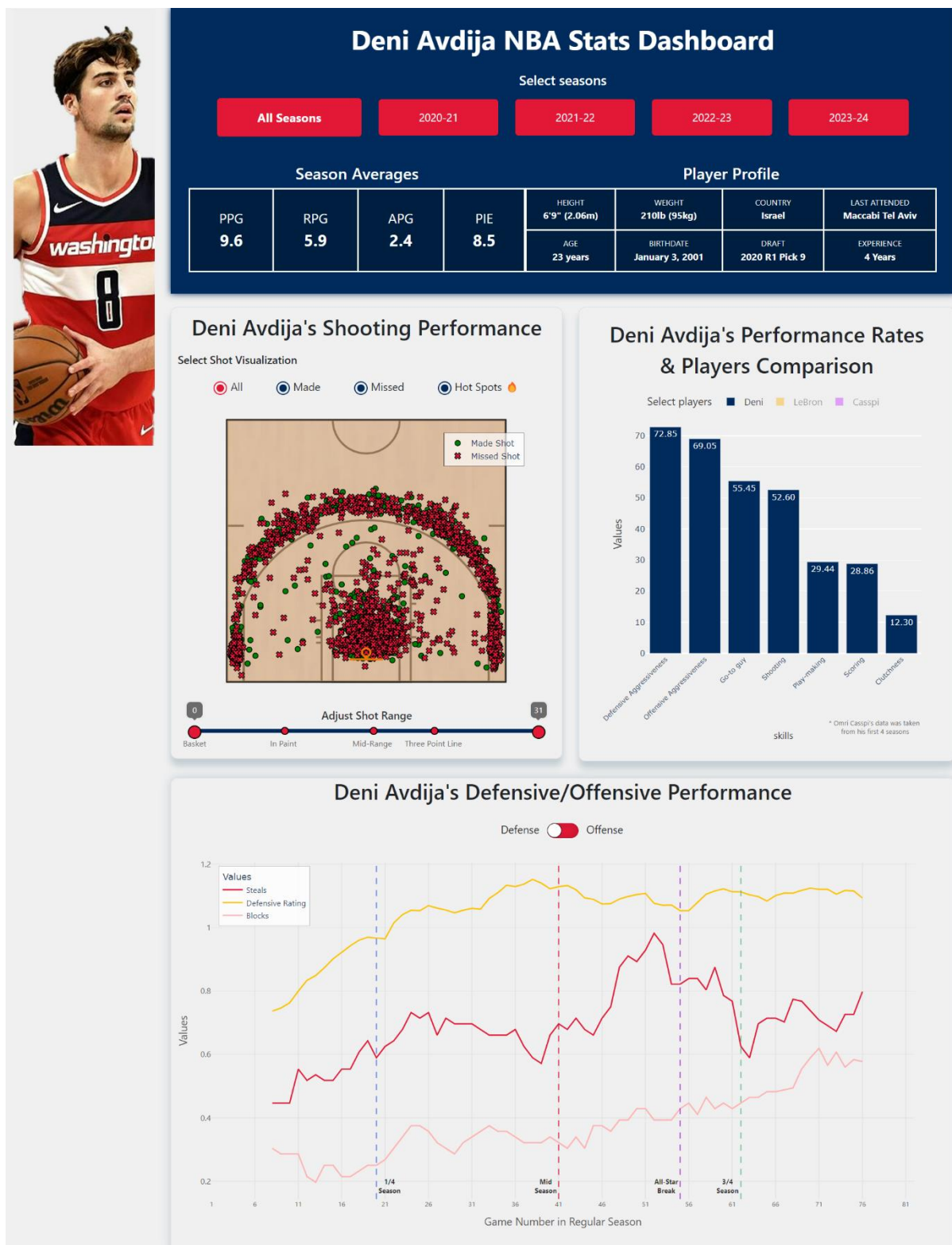
קישורים לדוגמאות של קוד ורעיונות עיצוביים שהסתמכנו עליהם:

- קוד לשרטוט מפת הזריקות - [שרטוט מגרש הכדורסל](#)
- רעיונות לעיצוב מפת הזריקות - [רעיון למגרש כדורסל](#)
- מילון מושגי כדורסל - <https://viziball.app/glossary/nba/en>
- השוואה בין שחקני כדורסל - <https://viziball.app/player/nba/en/290/48>
- שרטוט טבלת סטטיסטיקה - <https://www.nba.com/player/1630166/deni-avdija>

קישור לוויזואליזציה שלנו - <https://deniavdijadash-55aa58548f5b.herokuapp.com/>

נספחים:

1. צילום מסך של כל הדאשבורד:



2. צילום של טבלת - DA shooting stats

	SEASON	SHOT_DISTANCE	SHOT_MADE_FLAG	LOC_X	LOC_Y	SHOT_TYPE	SHOT_ZONE_BASIC	ACTION_TYPE
0	2020-21	24	1	-206	134	3PT Field Goal	Above the Break 3	Jump Shot
1	2020-21	25	1	-46	255	3PT Field Goal	Above the Break 3	Jump Shot
2	2020-21	5	1	20	47	2PT Field Goal	In The Paint (Non-RA)	Driving Layup Shot
3	2020-21	26	1	-171	204	3PT Field Goal	Above the Break 3	Jump Shot
4	2020-21	28	1	-151	236	3PT Field Goal	Above the Break 3	Jump Shot
...
2268	2023-24	23	1	228	34	3PT Field Goal	Right Corner 3	Jump Shot
2269	2023-24	5	1	-9	59	2PT Field Goal	In The Paint (Non-RA)	Driving Layup Shot
2270	2023-24	6	1	6	63	2PT Field Goal	In The Paint (Non-RA)	Driving Floating Jump Shot
2271	2023-24	25	0	-179	176	3PT Field Goal	Above the Break 3	Step Back Jump shot
2272	2023-24	0	1	7	5	2PT Field Goal	Restricted Area	Driving Layup Shot

2267 rows × 8 columns

3. צילום של טבלת - DA defence offence df

	GAME_NUMBER	SEASON	STL	BLK	offensiveRating	defensiveRating	effectiveFieldGoalPercentage	trueShootingPercentage	pace
0	1	2020-21	0	0	0.0	0.000	0.0	0.0	0.00
1	2	2020-21	0	0	0.0	0.000	0.0	0.0	0.00
2	3	2020-21	0	0	0.0	0.000	0.0	0.0	0.00
3	4	2020-21	0	0	0.0	0.000	0.0	0.0	0.00
4	5	2020-21	0	0	0.0	0.000	0.0	0.0	0.00
...
313	78	2023-24	0	0	103.9	1.314	81.3	78.8	103.00
314	79	2023-24	2	1	119.3	1.204	68.8	66.9	112.99
315	80	2023-24	1	0	80.8	1.360	45.0	50.6	110.17
316	81	2023-24	2	0	118.0	1.050	62.5	62.5	105.81
317	82	2023-24	1	0	112.2	1.400	50.0	50.7	114.97

318 rows × 9 columns

4. טבלאות של הסטטיסטיקות לכל שחקן *טבלה זו זהה בעמודות בין דני, לברון וכספי.

המשך פירוט העמודות של פורטו בתחילת העבודה:

- עמודת GROUP_VALUE_x מסמלת את שנת העונה.
- עמודת GP מסמלת את מספר המשחקים בעונה ששיחק.
- עמודת FGM מסמלת את מספר הקליעות שדה שקלע בעונה.
- עמודת FGA מסמלת את מספר ניסיונות הזריקות שדה בעונה.
- עמודת LOC_X מסמלת את ציר ה - x של הזריקה.
- עמודת FG3_PCT מסמלת את אחוז הקליעות לשלשה.
- עמודת FTM מסמלת את מספק קליעות העונשין שנקלעו בעונה.
- עמודת FTA מסמלת את מספר ניסיונות הזריקות עונשין בעונה.
- עמודת FT_PCT מסמלת את אחוז קליעות עונשין בעונה.
- עמודת OREB מסמלת את מספר ריבאונד התקפה בעונה.
- עמודת DREB מסמלת את מספר ריבאונד ההגנה בעונה.
- עמודת AST מסמלת את מספר האסיסטים בעונה.
- עמודת TOV מסמלת את מספר איבודי הכדור בעונה.
- עמודת STL מסמלת את מספר החטיפות בעונה.
- עמודת BLK מסמלת את מספר החסימות בעונה.
- עמודת PF המסמלת את מספר העבירו בעונה.
- עמודת PFD המסמלת את מספר העבירות שסחט בעונה.
- עמודת PTS המסמלת את מספר הנקודות שקלע בעונה.
- עמודת Season המסמלת את שנת העונה.
- עמודת GROUP_VALUE_y מסמלת את שנת העונה.
- עמודת OFF_RATING מסמלת את דירוג ההתקפה באותו משחק.
- עמודת DEF_RATING מסמלת את דירוג ההגנה באותו משחק.
- עמודת AST_PCT המסמלת אחוז הזריקות שנקלעו בקבוצה נמסרו מאסיסט של השחקן.
- עמודת EFG_PCT מסמלת את אחוז זריקות השדה מותאם לסוג זריקה.
- עמודת TS_PCT מסמלת את יעילות הזריקה.
- עמודת PACE מסמלת את הערכת מספר הפוזישנים במשחק.

	GROUP_VALUE_x	GP	FGM	FGA	FG3_PCT	FTM	FTA	FT_PCT	OREB	DREB	REB	AST	TOV	STL	BLK	PF
0	2023-24	75	406	803	0.374	202	273	0.740	84	456	540	286	160	59	36	185
1	2022-23	76	253	579	0.297	122	165	0.739	74	413	487	211	123	65	29	209
2	2021-22	82	250	579	0.317	103	136	0.757	52	373	425	167	87	59	44	190
3	2020-21	54	130	312	0.315	29	45	0.644	22	240	262	63	33	32	15	140

	PFD	PTS	Season	GROUP_VALUE_y	MIN	OFF_RATING	DEF_RATING	AST_PCT	AST_TO	EFG_PCT	TS_PCT	USG_PCT	PACE	PIE	POSS	PTS_RANK	FG_PCT_RANK
0	251	1102	2023-24	2023-24	30.1	109.7	116.2	0.174	1.79	0.560	0.597	0.196	105.06	0.109	4936	226	172
1	136	697	2022-23	2022-23	26.6	113.3	114.1	0.142	1.72	0.497	0.535	0.165	99.76	0.084	4192	197	201
2	113	685	2021-22	2021-22	24.2	109.2	110.1	0.117	1.92	0.503	0.536	0.158	99.88	0.086	4118	203	211
3	44	342	2020-21	2020-21	23.3	107.8	109.9	0.064	1.91	0.502	0.515	0.118	106.28	0.060	2773	261	176

5. הסבר על 7 מדדי הביצוע מגרף 1:

- (1) **Defensive Aggressiveness** - מדד שמחשב את האגרסיביות ההגנתית של השחקן תוך שילוב של דירוג הגנה (DEF_RATING), חטיפות למשחק (STL), חסימות למשחק (BLK), ועבירות למשחק (PF).
- (2) **Offensive Aggressiveness** - מדד שמחשב את האגרסיביות ההתקפית של השחקן על ידי שילוב של עבירות שסוחר השחקן (PFD), דירוג התקפה (OFF_RATING), וקצב משחק (PACE).
- (3) **Go-to guy (שחקן מרכזי)** - מדד שמחשב כמה השחקן נתפס כשחקן מרכזי בקבוצה על ידי שילוב של דקות משחק ממוצעות למשחק (MIN), אחוז שימוש (USG_PCT), מספר פוזשנים (הזדמנויות קליעה שהוא מייצר לעצמו ולאחרים) (POSS), ומדד (PIE Player Impact Estimate).
- (4) **Shooting** - מדד שמחשב את יכולת הקליעה של השחקן על ידי שילוב אפקטיביות (EFG_PCT), אחוז קליעות אמיתי (TS_PCT), אחוז קליעות עונשין (FT_PCT), ואחוז קליעות לשלוש נקודות (FG3_PCT).
- (5) **Play-making** - המדד הזה מחשב עד כמה השחקן תורם להצלחת חברי הקבוצה, תוך שילוב של היחס בין אסיסטים לאיבודים (AST_TO), אחוז האסיסטים של השחקן (AST_PCT).
- (6) **Scoring** - מדד שמחשב את יכולת הקליעה של השחקן על ידי שילוב של הנקודות למשחק (PTS) ביחס למשחקים ששיחק (GP), והמדד GMSCR למשחק.
- (7) **Clutchness** - מדד המחשב את יכולת קבלת ההחלטות של השחקן ברגעים קריטיים על ידי שילוב של דירוג הנקודות (PTS_RANK), ודירוג אחוז הקליעות מהשדה (FG_PCT_RANK).