

פרויקט ויזואליזציה :

" אי שם מעבר לקשת 🌈 🏀 "

שם הקורס: ויזואליזציה של מידע

שם ות"ז המגיש: איתי אסרף, 322601907

תאריך ההגשה: 27.06.22



1. מבוא :

1.1 מוטיבציה :

במהלך 14 שנים מחיי שיחקתי כדורסל באופן מקצועי. מליגות הקטסל והילדים עם אליפויות בגילאים הנמוכים, המשיך לאימונים בנבחרת ישראל עד גיל 16 (נבחרת הקדטים) ונעצר (זמנית) בגילאי הבוגרים בליגה א'.

מאז ומתמיד קיבלתי הערות ותגובות על הגובה שלי (הגובה שלי כיום הינו 1.74 מטר לאחר שבגילאים הצעירים יותר הגובה שלי היה באזור 1.60 מטר) כגון: "כדורסל זה משחק של גבוהים", "אתה גמד בארץ של ענקים", "איך תסתדר עם אנשים שאתה מגיע להם לאזור הבטן?" ועוד...

ההערות היו קשות, אך גרמו לי לפתוח את המחשבה בסגנון המשחק שלי.

בגילאים הנמוכים יותר, מכיוון שאני נמוך ויותר זריז מאחרים, הייתי מנצל את הצעד הראשון המהיר שלי ובעזרת מהירות מתפרצת עוקף את שחקני היריבה ומגיע אל הסל.

לצערי, בגילאים היותר מתקדמים (נערים) נתקלתי בקושי שלמרות שהייתי מצליח לעקוף את שחקני היריבה, הזריקות שלי היו נחסמות/ לא מספיק איכותיות. ככל שהייתי נע קרוב יותר לכיוון הסל, ליריבי היה קל יותר לחסום אותי מכיוון הגובה היותר נמוך שהיה לי.

כתוצאה מכך, השקעתי רבות בצפייה של משחקי כדורסל רבים בליגת הכדורסל האמריקאית הטובה בעולם (NBA). בעקבות הצפייה המרובה במשחקים, החלטתי להתאמן על זריקות ממרחקים שונים ובמיוחד מעבר לקשת ה-3. לאחר זמן מועט, הייתי קולע כמות נקודות גבוהה יותר ודבר זה פיצה על חיסרון הגובה שלי בצורה משמעותית ובעקבות זאת זכיתי לזימון לנבחרת ישראל.

הודות לכך ששיניתי את סגנון המשחק שלי אל סגנון של פחות חדירות ויותר זריקות מכיוון השלוש, פרצתי בעולם הכדורסל. בנוסף לכך, עוד רבים וטובים השתמשו בסגנון משחק זה כדי לפרוץ את דרכם בעולם הכדורסל.

כיום, עולם הכדורסל כולו השתנה, אנחנו זוכים לראות את "עידן השלשות", עידן שבו זריקות ל-3 נקודות הן חלק עיקרי ממשחק הכדורסל.

המטרה שלי בפרויקט זה, לבדוק האם אכן המהפכה המדוברת העצומה שקורית היא נכונה ואמיתית ואם כן להראות את התהליך והשינוי בהדרגה שזה קרה במהלך השנים ולספק תובנות לשחקני כדורסל.

בהמשך הדו"ח אנחנו נראה תובנות משמעותיות שלא חשבתי מלכתחילה שהן קורות ובנוסף נוכל לשים לב לתופעה מעניינת.

1.2 נושא הויזואליזציה:

הוויזואליזציה עוסקת בתחום הכדורסל ומתמקדת בבדיקת המהפכה שעובר ה NBA בין השנים 2000-2020, מהפכת השלשות- כמות הזריקות אל מעבר קשת השלוש הולכת ועולה במהלך השנים הללו עד כדי מרכיב עיקרי ומשמעותי בכדורסל ב NBA. ביחד עם זאת, אני מניח שתהיה ירידה בכמות הזריקות ככל שנתקרב לסל.

מכיוון שיש 4 מיליון נתונים שעלינו לעבד ולהראות כמות נתונים עצומה תגרום לצופה הוויזואליזציה חוסר הבנה בתהליכים ובמגמות בגרף. לכן, בחרתי לקחת את 100 המקומות הכי נפוצים שנזרקו בהם הכי הרבה פעמים במהלך עונת כדורסל ולראות אם יש שינוי במקומות הכי נפוצים, כמות הזריקות שנזרקו באותם מקומות וכו....

לאחר שנמחיש את הוויזואליזציה הבודקת אם המהפכה וההשערה שלנו, נוכל לספק תובנות משמעותיות לשחקני כדורסל.

1.3 נתונים- מקור והיקף:

הנתונים נלקחו מ nba_api המחזיק נתונים על עונות הכדורסל במהלך השנים שבכל עונה נקבל מידע על משחקים, מאמנים, שחקנים, זריקות, סטטיסטיקות ועוד..

לצורך פרויקט זה, הנתונים שנלקחו מן ה api הם נתונים העוסקים ב shotChart Details, כל הזריקות שנזרקו (בין אם נקלעו/הוחטאו) במהלך עונה מסוימת.

מכיוון שאנו מתמקדים בשנים 2000-2020 אנחנו לקחנו את נתוני כל הזריקות שנזרקו במהלך כל העונות הללו.

כתוצאה מכך אנחנו מתעסקים בסדר גודל של 4 מיליון נתונים. הסיבה לכך היא שבכל עונה נזרקות כ 200,000 זריקות. אנו חוקרים 21 עונות לכן נקבל סדר גודל עצום של נתונים.

מתוך ה 4 מיליון הנתונים הללו, אנחנו לקחנו את 100 המקומות הכי נפוצים שנזרקו בהם זריקות בכל עונה לאחר סינון הנתונים (יובא בהמשך הדו"ח הסבר לכך).

2. נתונים:

הדאטה סט שהשתמשתי יש לנו 27 פיצרים ועוד 7 פיצרים (attributes) שהוספתי על ידי חישוב בעזרת הפיצרים הנתונים שהיו בדאטה סט (לצורך pre-process של הוויזואליזציה). כל שורה (item) בדאטה סט היא זריקה שנזרקה במשחק והעמודות הן הפיצרים של אותה זריקה. גודל הדאטה סט הוא כמות הזריקות שנזרקו באותה שנה

אופי דאטה סט זה הוא מתאר כל זריקה שנזרקה, יותר מפורט. פחות נתונים כמותיים, יותר תיאורים של מקום הזריקה, תאריך הזריקה, מתי, באיזה משחק ועוד...

Datasets Types, Data types, Attributes 2.1**Dataset Types**

Table	Data set type
-------	---------------

Data types, Attributes

Description	Data types	Attributes
מספר מזהה של משחק ב NBA	categorical	Game id
מספר מזהה של אירוע במשחק ב NBA . כגון : זריקה, עבירה, איבוד כדור וכו.. כל מספר ייחודי בפעם זה מתאר אירוע של זריקה.	categorical	Game event id
מספר מזהה של שחקן שביצע זריקה.	categorical	Player id
שם השחקן שביצע זריקה	categorical	Player name
מספר המזהה של קבוצה שבה שיחק השחקן שביצע זריקה.	categorical	Team id
שם הקבוצה שבה שיחק השחקן שביצע זריקה	categorical	Team name
מזהה הרבע (יש 4 רבעים) שבו נזרקה הזריקה במהלך משחק ב NBA.	categorical	Period
מספר הדקות שנשארו לסיום הרבע מהרגע שנזרקה הזריקה. לדוג': הזריקה נזרקה כאשר נשארו 11 ו 48 שניות לסיום הרבע. בפעם זאת הערך 11.	quantitative	Minutes remaining
השנייה (מתוך דקה-60 שניות) המדויקת שנזרקה בה הזריקה . לדוג': הזריקה	quantitative	Seconds remaining

נזרקה כאשר נשאר 11 ו 48 שניות לסיום הרבע. בפיצ'ר זה נראה את הערך 48.		
סוג אירוע מתאר אם הזריקה נכנסה/הזריקה הוחמצה.	ordinal	Event type
סוג הזריקה שנזרקה. לדוג': שחקן הטביע, לכן בערך זה נראה dunk. שחקן שקלע בצעד וחצי, לכן בערך זה נראה lay-up.	categorical	Action type
אם הזריקה שנזרקה היא ל 2 נקודות או 3 נקודות.	ordinal	Shot type
האיזור שבו נזרקה הזריקה. יתכנו 7 אפשרויות לאזורים. לדוגמה: באזור הצבע, חצי מרחק, שלשה בפינה ימנית וכו..	categorical	Shot zone basic
הכיוון שממנו נזרקה הזריקה. יתכנו 6 אפשרויות של כיוונים. לדוג': מרכז, פינה שמאלית, ימנית וכו..	categorical	Shot zone area
המרחק שבו נזרקה הזריקה. יש 4 קטגוריות למרחק שממנו נזרקה הזריקה. לדוג': מתחת ל 8ft, בין 8-16 ft, בין 16-24 ft, מעל 24 ft.	ordinal	Shot zone range
המרחק המדויק שבו נזרקה הזריקה, שמשתייך לקטגוריית המרחק (תא אחד למעלה). לדוג': במרחק 3ft מן הסל זרק השחקן. (ערך זה ישתייך ל מתחת ל 8ft)	quantitative	Shot distance

המרחק האופקי שבו נזרקה הזריקה מן הסל, כאשר ראשית הצירים זה הסל עצמו (הטבעת)	quantitative	Loc x
המרחק האנכי שבו נזרקה הזריקה מן הסל, כאשר ראשית הצירים זה הסל עצמו (הטבעת)	quantitative	Loc y
אם הזריקה שנזרקה נכנסה או לא. 1- נכנס, 0- הוחמץ.	ordinal	Shot made flag
תאריך המשחק שבו נזרקה הזריקה. הערך יראה בצורה של: yyyymmdd	quantitative	Game date
הקבוצה המארחת של המשחק שבו התרחשה הזריקה	categorical	HTM
הקבוצה האורחת של המשחק שבו התרחשה הזריקה	categorical	VTM

3. מטלות:

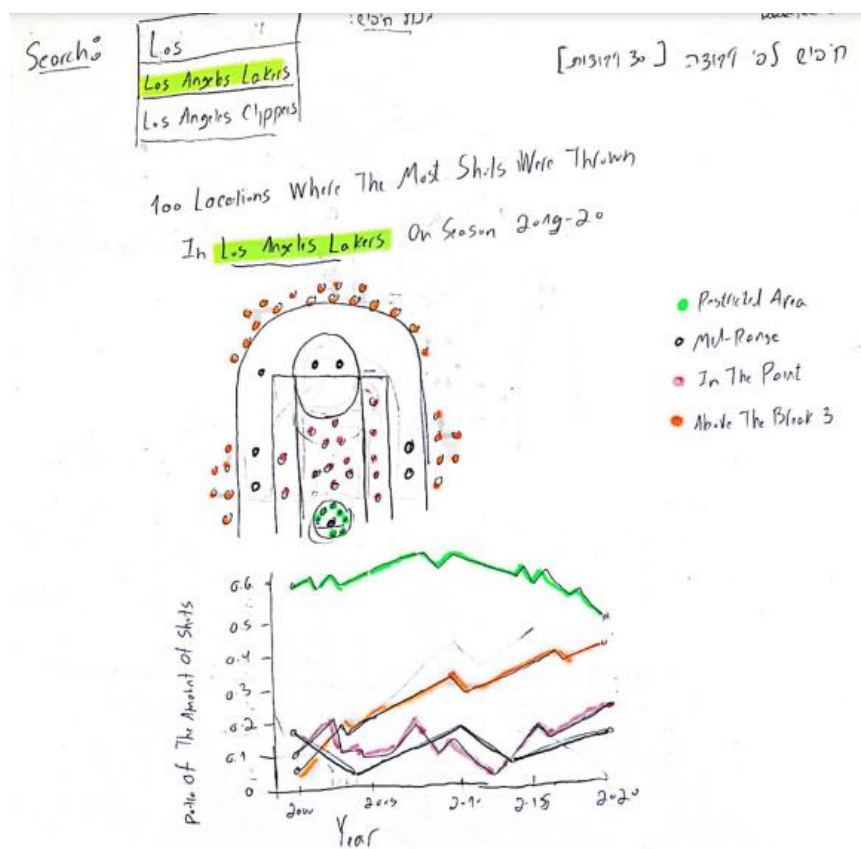
1. האם ניתן לראות מגמה שהולכת ומתפתחת עם השנים בעניין מיקומי זריקות לסל?
2. האם ככל שנתקדם עם השנים אז נראה יותר זריקות ל 3?
3. מה השינוי היחסי שעבר כל איזור במהלך השנים?
4. מי השחקן שזרק הכי הרבה במיקום מסוים?
5. באיזו שנה לא היה כלל חצי מרחק ? ("מותו של החצי מרחק").
6. האם יש קשר בין המשתנים : עונת הכדורסל ל אחוז היחסי של הזריקות בכל איזור?

מספר שאלה	Action	Target
1	השוואה (compare)	מגמות של פופולריות הזריקות בכל איזור.
2	מציאת	מגמה של זריקות ל 3 כתלות בשנים. (שנים יותר מאוחרות יותר זריקות)
3	השוואה (compare)	מגמות של סך הזריקות

היחסי בכל אזור במהלך השנים.		
השחקן שזרק הכי הרבה במיקום מסוים.	זיהוי (identity)	4
השנה שלא היה בה כלל חצי מרחק	זיהוי (identity)	5
קורולציה בין אחוז היחסי של זריקות לאזור.	מציאת	6

4.1. רעיונות אלטרנטיביים:

רעיון 1: מנוע חיפוש שבו נחפש עבור הקבוצה את 100 המיקומים הכי נפוצים שבהם נזרקו בהם באותה עונה, ובכך להראות עבור כל קבוצה את המגמה. המטרה של רעיון זה היא לבצע "זום אין" ממגמה של ליגת ה NBA לבדיקת מגמה לפי כל קבוצה.



תמונה 1- סקיצה של יישום מנוע החיפוש בגרף. חיפוש של קבוצה ועבורה נקבל את מיקומי הזריקות הנפוצים ביותר ואת אחוז הזריקות היחסי בכל אזור.

לא ביצעתי את הרעיון הנ"ל היא בגלל ממספר סיבות:

- אנחנו מתעסקים על 4 מיליון נתונים שמתפרשים על 20 שנים, דבר הלוקח המון זמן (בין שעה וחצי לשעתיים). הנוסחה של מיקומי המגרש (כפי שיובא בהמשך) לקח המון מן זמן תכנון הפרויקט) הרבה פעמים ביצעתי ניסוי ותהייה כדי לשרטט את הנקודות שיהיו מותאמות לגודל של המגרש שלי (!).
- הדאטה סט שאני טענתי לא מצוין בו מאיזו קבוצה נזרקה הזריקה, אלא רק שם השחקן. ביכולתי למצוא את שם הקבוצה לפי שם השחקן (לשמור את כל השחקנים שזרקו ואז למצוא את הקבוצות של השחקנים שהן שיחקו בהם). הייתי צריך לטעון עוד דאטה סט מתוך ה NBA API ולמצוא את כל השחקנים מרשימת השחקנים של אותה שנה ולמצוא את הקבוצה שלהם. מכיוון שישנם טריידים במהלך אותה שנה (שחקן ששיחק בקבוצה א עבר לשחק בקבוצה ב) הייתי צריך למצוא עוד דאטה סט מן NBA API שימצא את השחקנים שעברו בטרייד באותה שנה וכך לעדכן אותם בקבוצותיהם. בנוסף לזמן הריצה הגדול שלוקח טעינת הנתונים , יש לסנכרן בין 3 דאטה סטים שונים שרק יגבירו את זמן הריצה ומגדיל את הקושי.
- כפי שלמדנו בהרצאה ובכיתה, לא קורים מצבים שבהם אנחנו רואים את היוזאליזציה שברצוננו ליצור מתחילתה ועד סופה. תמיד נוספים עוד פרטים, עוד תיקונים, שינויים בגלל דברים מסוימים. כתוצאה מכך, היוזאליזציה עוברת תהליך ושינוי מהרעיון הבסיסי שעליו חשבתי. רעיון זה הגיע בשלב יותר מאוחר- בשלב הסיום של היוזאליזציה של הרעיון שרציתי (שהתפתח עם הדרך כאשר התחלתי אותו בשלב מאוד מוקדם) ובעקבות זאת הזמן הנותר לי לבצע זאת בנוסף לדרגת הקושי הגבוהה הייתה לא מספיקה עבורי .

רעיון 2:

הוספה של לחיצת כפתור שבה רציתי לעשות אחת מ 2 האופציות הבאות:

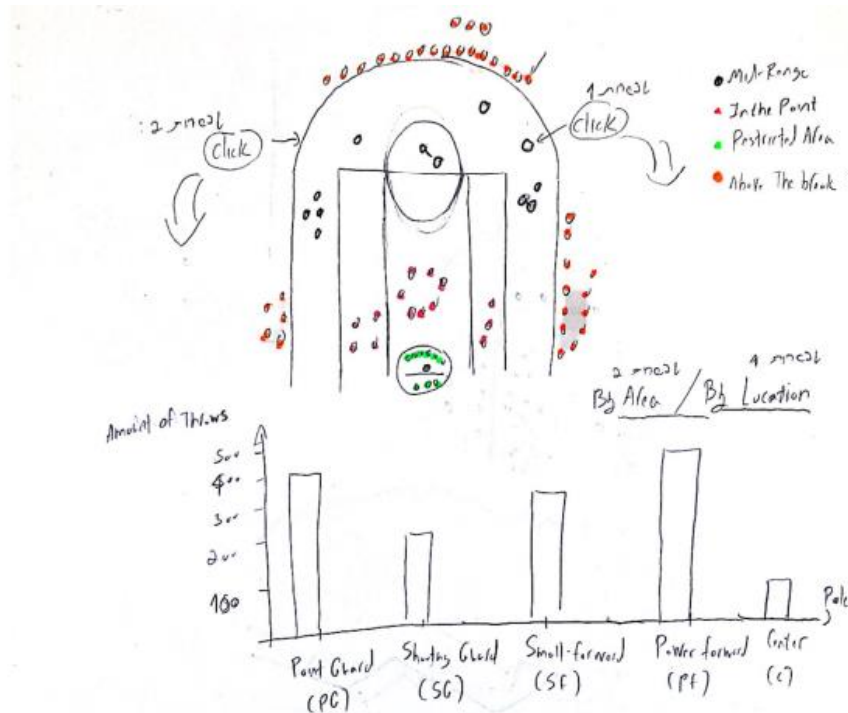
1. עבור כל מיקום- כאשר נלחץ על נקודה מסוימת (מיקום) אנחנו נקבל גרף שבו נראה את התפלגות

הזריקות לפי כל עמדה

כאשר העמדות האפשריות הן :

- PG-Point Guard
- SG-Shooting Guard
- SF-Small Forward
- PF-Power Forward
- C-Center

2. נגדיר אזורים על המגרש. כאשר נלחץ על **אזור** אנחנו נראה את התפלגות הזריקות לפי כל עמדה. בסקיצה ניתן לראות שלחיצה על קשת ה-3 תביא אותנו לראות את הגרף המדובר.



תמונה 2- סקיצה של 2 אפשרויות ליישום של קישור בין גרף של התפלגות הזריקות לפי עמדה לגרף המגרש. אפשרות 1- לקשר לפי מיקום. אפשרות 2- לקשר לפי אזור.

לא ביצעתי את הרעיון הנ"ל היא בגלל ממספר סיבות:

- האפשרות השנייה של הוספה של לחיצת כפתור על אזור של מגרש, מימושה איננו קל. הסיבה לכך היא שצריך להגדיר את הטווח שבהם ניתן לקבל את הלחיצה (בטווח שמקואורדינטה א לקואורדינטה ב אז נקבל את הגרף עבור אזור X). יש מעוט של דברים שנעשו באינטרנט וכתוצאה מכך ניסיתי לבצע זאת לבד. לצערי, לא הצלחתי בגלל הקושי להגדיר טווח ללחיצה מן העכבר
- האפשרות הראשונה של לבצע לפי מיקום לא בוצעה מכיוון שלדעתי היא לא מספרת את סיפור הוויזואליזציה. אני מתמקד במיקום מאוד ספציפי של הזריקה ומחפש את ההתפלגות הזריקה לפי עמדות. מיקומים זה דבר מאוד דינמי ולכן אנחנו נראה הרבה דינמיות בהתפלגות הזריקות לפי כל עמדה (בשנה אחת הרכז יותר, בשנה שנייה הסנטר יותר). המטרה שלי היא להראות תופעה לפי אזורים ולא לפי מיקומים, לכן רעיון זה בחרתי שלא לבצע אותו.
- סיבה נוספת היא אותה סיבה שציינתי עבור רעיון 1. רעיון זה הגיע בשלב יותר מאוחר- בשלב הסיום של הוויזואליזציה של הרעיון שרציתי (שהתפתח עם הדרך כאשר התחלתי אותו בשלב

מאוד מוקדם) ובעקבות זאת הזמן הנותר לי לבצע זאת בנוסף לדרגת הקושי הגבוהה הייתה לא מספיקה עבורי.

4.2. עיצובים אלטרנטיביים:

מכיוון שיצרתי dashboard עם שני גרפים, אציע 2 אלטרנטיבות עבור כל גרף:

1. עבור גרף ה dot / line chart (נקודות שמחוברות על ידי קו) המבטא את המגמה עבור היחס של כמות הזריקות באזור לזמן (מראה שינוי לאורך זמן) ובנוסף מבטא יחס בין משתנים !

אלטרנטיבה 1 – Stacked Bar

הגרף שלנו מבטא יחס בין "חלק לשלם" - החלק היחסי של הזריקות בכל אזור.

Stacked Bar עונה בצורה טובה על הגדרה. הוא מייצג את החלק היחסי של משתנים מהסך הכולל. כאשר נשים את ה bars על ציר העונות, נוכל לראות מגמה של החלק היחסי של כל אזור עם הזמן (אם הוא גדל או קטן). כתוצאה מכך אנחנו עונים על 2 המטרות שלנו בייצוג גרף זה !

Channels:

- length - אורך ה bar
- Color Hue - לכל אזור יש צבע (לכן יש שימוש בצבע)
- spatial regions: one per glyph - כל bar נמצא במקום מופרד, באזור משלו (בערך האיקס שלו) לכן אין ערבוב ביניהם.

Mark:

- vertical stack of line marks
- glyph: composite object, internal structure from multiple marks – זה אובייקט מורכב. כל bar יש בתוכו מבנה של מספר צבעים - כאשר כל צבע מייצג ערך קטגוריאלי אחר בתוך משתנה (בדוגמה שלנו : לאזור יש 4 ערכים קטגוריאליים).

אלטרנטיבה 2 - scatter plot.

אנחנו מחפשים להראות מגמה אם לאורך השנים אנחנו נראה באזור מסוים עלייה או ירידה של יחס הזריקות שנזרקו ממנו לכמות הזריקות הכוללת. בנוסף לכך ניתן להסיק קורולציה בין משתנים לדוגמה: עבור אזור זריקה מ 3 ככל שיעברו השנים היחס הנ"ל יגדל.

Scatter Plot עונה על הגדרה זו ולכן נוכל להחליף את הגרף שהצגנו בגרף זה.

Channels:

- Vertical position - מיקום אנכי
- Horizontal position - מיקום אופקי
- Color Hue - לכל אזור יש צבע (לכן יש שימוש בצבע)

Mark:

- Point

2. עבור גרף המגרש: שרטטתי מגרש כדי להראות את מיקומי 100 הזריקות שבהם נזרקו מהם הכי הרבה זריקות.

מכיוון שזו לא תבנית טריוויאלית, ישנם מספר שינויים שיצטרכו שינוי מהותי במבנה הגרף (ויזואליזציה מאוד שונה מן הויזואליזציה שביצעתי) אך עדיין תוכל לענות באופן מספק את שאלת המחקר שלי.

אלטרנטיבה 1 - Scatter Plot

המטרה שלנו זה להראות את כיצד מגמת המיקומים הולכת ומשתנה עם השנים. ז"א שכלל שיעברו שנים אנחנו נראה יותר זריקות מרחוק ופחות זריקות מאזור האמצע (או בפנים או בחוץ). לכן אנחנו נמיר את הגרף לציר איקס וציר y כאשר כל אחד מן הצירים מבטא את המיקום של הזריקה במגרש (לדוגמה: עבור (-2125,48)). כל מיקום יצבע בצבע האזור שלו בהתאם. ככל שנראה יותר נקודות באזור מסוים נוכל להבין שיש עלייה בתדירות הזריקות מאותו אזור. אנחנו מתייחסים למיקומים כאל ערכים כמותיים ואל האזורים כקטגוריליים.

Channels:

- Vertical position - מיקום אנכי
- Horizontal position - מיקום אופקי
- Color Hue - לכל אזור יש צבע (לכן יש שימוש בצבע)

Mark:

- Point

אלטרנטיבה 2 - Bar Chart (Race)

יהיו לנו 4 עמודות כאשר כל עמודה היא אזור. (Bars). ציר ה X יהיו הערכים הקטגורילים של האזורים וציר ה Y יהיה כמות המקומות השונים שנזרקו בהם הכי הרבה זריקות באותה שנה (הסכום של כולם יהיה 100, מכיוון שאני מתעסק ב 100 המיקומים הפופולריים ביותר). ה bars יזוזו בהתאם לתדירות שלהם (אם בשנת 2010 מספר המקומות השונים הנפוצים ביותר מחוץ לקשת ה 3 היה 30 וזריקות מחצי מרחק היה 31 זריקות. מבחינת סדר העמודות העמודה לחצי מרחק יהיה בצד הימני יותר מן העמודה של קשת (יותר קרובה לראשית הצירים) ואם בשנת 2011 זריקות מחוץ לקשת ה 3- 38 זריקות וזריקות לחצי מרחק- 29 זריקות, הסדר יהיה אחרת, כעת העמודה של מחוץ קשת ה 3 תהיה ימנית יותר מן עמודת חצי המרחק).

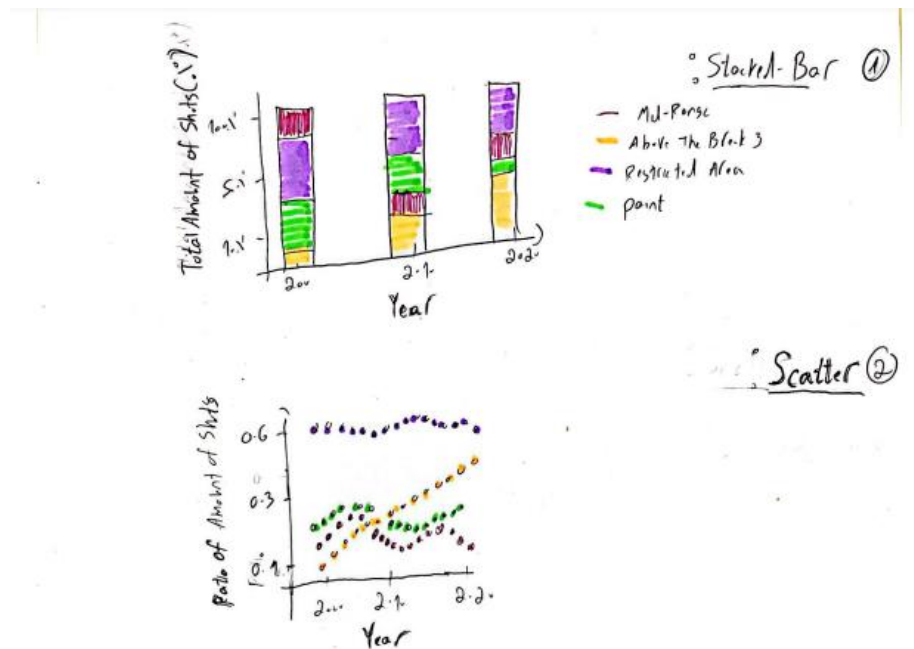
- length- אורך ה bar
 - Color Hue- לכל אזור יש צבע (לכן יש שימוש בצבע)
 - spatial regions: one mark glyph- כל bar נמצא במקום מופרד, באזור משלו (בערך האינסולו) לכן אין ערבוב ביניהם.
- בנוסף לכך, אנחנו ממיינים לפי סדר כמות המקומות השונים שבהם נזרקו באותו אזור (מהגדול לקטן).

Mark:

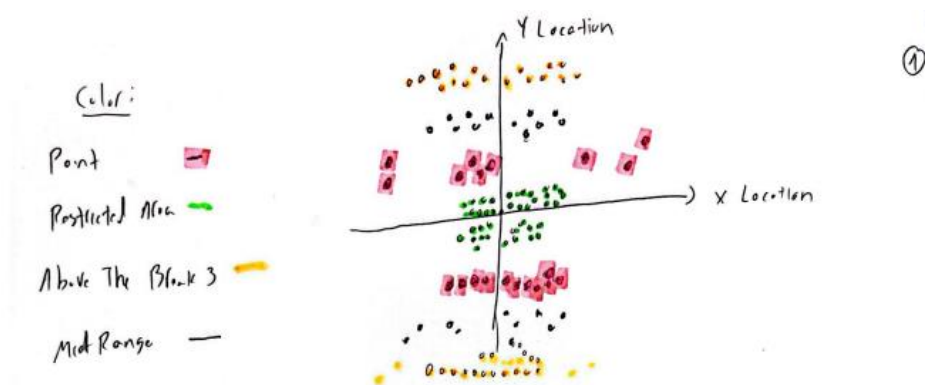
- vertical stack of line marks

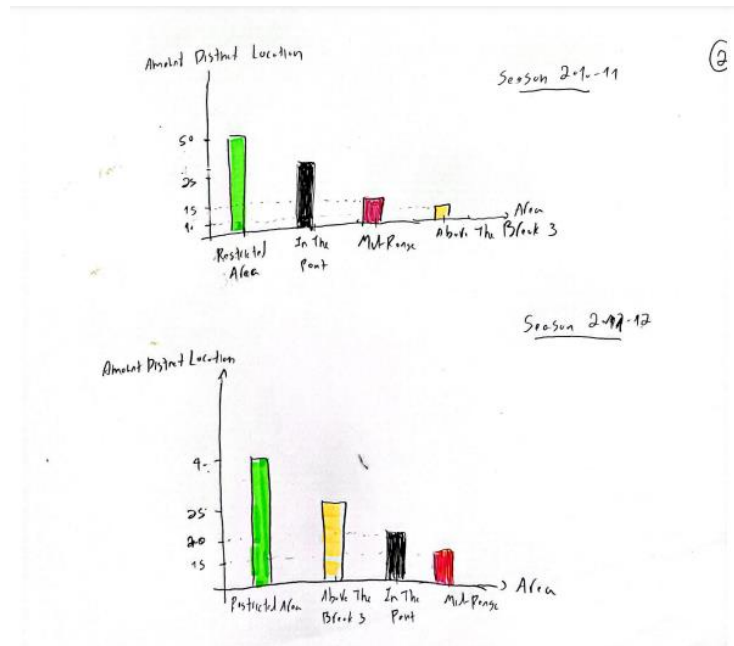
סקיצות על כל אחד מן האלטרנטיבות של כל אחד מן היוזאליזציות:

- סקיצות אלטרנטיבות עבור ויזואליזציות ה line chart.



- סקיצות של אלטרנטיבות עבור ויזואליזציות המגרש (court chart)





תמונה 3- סקיצה של אלטרנטיבות עבור הוויזואליזציה court chart .

5. הסבר על היישום:

Preprocessing: 5.1

במהלך תהליך הוויזואליזציה נתקלתי במספר קשיים:

1. עיבוד של 4 מיליון של נתוני הזריקות במהלך 21 שנים.
2. נתוני מיקום הזריקות (בציר האנכי והאופקי) אינם רלוונטיים מכיוון שהם מתייחסים לקואורדינטות של מגרש אמיתי.
3. חסרים נתונים סיכומים עבור כל מרחק (יש עבור כל אזור וכיוון, לדוגמה: מעבר לקשת השלוש באמצע, נזרקו סך הכל 4000 זריקות בשנת 2019).

כתוצאה מכך החלטתי לפתור את הקשיים הללו בדרכים יצירתיות אחרות:

1. לקחתי את 100 המיקומים הכי נפוצים שבהם נזרקו הכי הרבה זריקות בכל עונה. הסיבה לכך היא שיש מגוון רחב מאוד של זריקות. תצוגה של מגוון רחב זה תוריד משמעותית את מטרת הוויזואליזציה – לענות אם אכן היה מהפכה אל מעבר קשת השלוש.
- מעבר להראות את מיקומי הזריקות, הינני רוצה להראות את פופולריות המיקומים, שכן להראות את המיקומים בלבד לא תוביל אותי למסקנות משמעותיות.

2. יצרתי 2 פיצ'רים חדשים (הסבר יובא בטבלה הפיצ'רים שהוספתי) המתאימים למיקומים למגרש שיצרתי. השתמשתי בנוסחה שתנרמל את המיקומים האנכיים והאופקיים באופן שיתאים לגודל המגרש ובאופן יחסי לסל (ראשית הצירים).
2 הנוסחאות הללו מטרתן להביא את ראשית הצירים ל (0,0) לאחר ניסוי ותהייה על ערכים המצויינים בנוסחה, כאשר המשתנים הינם:

- Len_x_bound - אורך קו החוץ האופקי של המגרש.
- Len_y_bound - אורך קו החוץ האנכי של המגרש.
- Location_x - מיקום האופקי הנתון בדאטה סט
- Width - רוחב המגרש
- Height - גובה המגרש

$$normalized\ x = \left(\frac{location_x + 250}{width} \right) * len_x_bound$$

$$normalized\ y = \left(\frac{location_y + 50}{height} \right) * len_y_bound$$

לאחר שהתאמנו את הנקודות בהתאם לראשית הצירים, נמיר כל נקודה בהתאם לגודל המגרש.

$$normalized\ x = \left(\frac{\frac{normalized\ x * width}{len_x_bound + (normalized\ x + 1) * width}}{len_x_bound} \right) * len_x_bound$$

$$normalized\ y = \left(\frac{\frac{normalized\ y * height}{len_y_bound + (normalized\ y + 1) * height}}{len_y_bound} \right) * len_y_bound$$

כתוצאה מכך, הנקודות פרופורציונאליות לגודל המגרש ובנראות טובה יותר מן המיקומים הנתונים מן הדאטה סט המקורי.

3. לאחר שיצרתי קובץ חדש שבו יש את 100 המיקומים הכי נפוצים שנזרקו בהם הכי הרבה זריקות בכל עונה (2100 רשומות, $21 * 100 = 2100$). עבור כל מיקום (מיקום אנכי, מיקום אופקי)) סכמתי את כמות הזריקות שנקלעו, כמות הזריקות שנזרקו, אחוזי הזריקות שנכנסו.

על כן הוספתי את הפיצ'רים הבאים בשלב ה preprocess הנגזרים מן הפיצ'רים של הדאטה סט כדי למקם את מיקומי הזריקות שקיבלתי מן הדאטה סט אל המגרש שיצרתי. בנוסף, הוספה נתונים לאנשים שירצו להתעמק ולעשות zoom in לנתונים.

Data types, Attributes

Description	Data types	Attributes
תיקון למרחק האופקי של הזריקה שנזרקה לסל בדאטה סט של NBA	quantitative	On court loc x

הקואורדינטות של המרחק האופקי והאנכי מתייחסות לגודל מגרש אמיתי. מכיוון שהוויזואליזציה שייצרתי בגודל שונה ממגרש אמיתי, הייתי צריך לעשות תיקון לקואורדינטות הללו.		
תיקון למרחק האנכי של הזריקה שנזרקה לסל בדאטה סט של NBA הקואורדינטות של המרחק האופקי והאנכי מתייחסות לגודל מגרש אמיתי. מכיוון שהוויזואליזציה שייצרתי בגודל שונה ממגרש אמיתי, הייתי צריך לעשות תיקון לקואורדינטות הללו	quantitative	On court loc y
כמות הזריקות שנכנסו לפי מיקום האנכי והאופקי שנזרקה הזריקה. לדוג': עבור מיקום (-45, 89) כמות הזריקות שנכנסו הינה 1570.	quantitative	Count throws loc
האחוז הזריקות שנכנסו מסך כל הזריקות שנזרקו לפי המיקום האנכי והאופקי המתוקן שנזרקה הזריקה. לדוג': עבור מיקום (-45, 89) כמות הזריקות שנכנסו הינה 1570.	quantitative	Fg (field goal) location
שם השחקן שזרק הכי הרבה זריקות מן המיקום האנכי והאופקי המתוקן שנזרקה בו הזריקה.	categorical	Player most frequency loc
אחוז הזריקות שנכנסו מסך הזריקות של השחקן שזרק הכי הרבה זריקות מן המיקום	quantitative	Fg player most frequency loc

האנכי והאופקי המתוקן שנזרקה בו הזריקה.		
כמות הזריקות של השחקן שזרק הכי הרבה זריקות מן המיקום האנכי והאופקי המתוקן שנזרקה בו הזריקה	quantitative	Amount throws of player most frequency loc
העונה שבה נזרקו הזריקות לפי המיקום המתוקן	quantitative	season

יצרתי דאטה סט נוסף עבור הגרף השני ב dashboard שעמודותיו נגזרות מן הדאטה סט המקורי. הסיבה לכך היא כדי להראות את הדינאמיות והשינוי בכמות הזריקות לפי איזורים: מעבר לקשת השלוש, באזור הצבע, חצי מרחק ובעיגול שמתחת לסל (restricted area) ביחס אל הכמות הזריקות הכוללת.

Table	Data set type
-------	---------------

Data types, Attributes

Description	Data types	Attributes
העונה שבה נזרקו הזריקות.	quantitative	Season
כל השנים שעברו משנת 2000. לדוגמה: עבור עונת 2006, יהיה לנו את השנים: 2000-2006.	quantitative	Year
יחס של כמות הזריקות לשלוש נקודות ביחס לכמות כל הזריקות	quantitative	3Points_Shots_To_All_Shots
יחס של כמות הזריקות לחצי מרחק ביחס לכמות כל הזריקות	quantitative	MID_Points_Shots_To_All_Shots
יחס של כמות הזריקות בתוך הצבע ביחס לכמות כל הזריקות	quantitative	PAINT_Points_Shots_To_All_Shots
יחס של כמות הזריקות בעיגול שמתחת לסל (restricted area) ביחס	quantitative	RESTRICT_Points_Shots_To_All_Shots

לכמות כל הזריקות		
------------------	--	--

5.2 תצלומי מסך והסבר על הויזואליזציה:

אני מראה בויזואליזציה האם קיימת מגמה של סגנון כדורסל שונה- סגנון כדורסל שבו אנחנו רואים הרבה יותר זריקות מעבר לקשת, הרבה פחות באזור ה 2 נקודות הרחוק (mid-range).

לצורך הויזואליזציה הקמתי dashboard שבו אני מראה 2 ויזואליזציות- האחת חצי מגרש כדורסל שבו ניתן לראות את 100 המיקומים שנזרקו בהם הכי הרבה זריקות באותה עונה. הויזואליזציה השניה היא להראות עבור כל אזור (חצי מרחק, צבע, שלשה, restricted area) מתוך 100 המיקומים שנזרקו בהם זריקות, את היחס בין כמות הזריקות שזורקים בו לכמות הזריקות הכוללת! עבור גרף זה בחרתי את indexed-line-chart מכיוון שהיא מאוד טובה עבור משתנים שמשתנים לאורך זמן. כך אנו יכולים לראות מגמה שמשתנית לאורך העונות.

ראשית, כותרת הויזואליזציה (שם הפרויקט והסבר קצר וכולל על הויזואליזציה)

Somewhere Over the Rainbow 🌈🏀

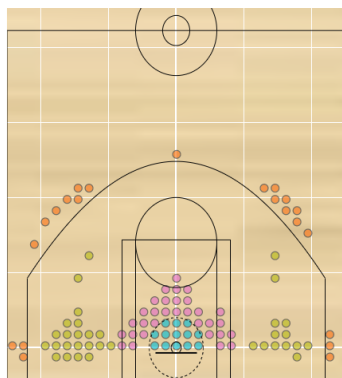
The NBA Revolution

Analysis of the 100 Locations where the most shots were thrown between the years 2000 and 2020

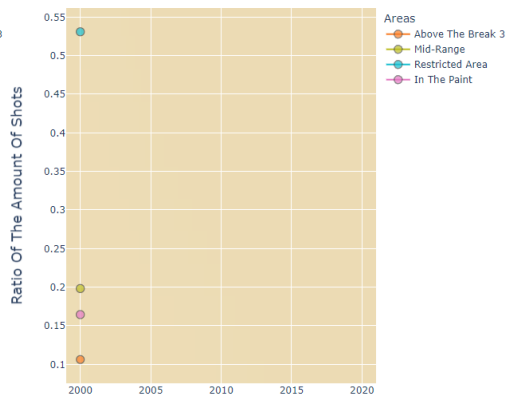
תמונה 4 – כותרת הפרויקט ותיאור מתומצת על הפרויקט

שנית, הויזואליזציות עצמן- נקודת ההתחלה.

100 Locations Where The Most Shots Were Thrown
On 1999-2000 Season



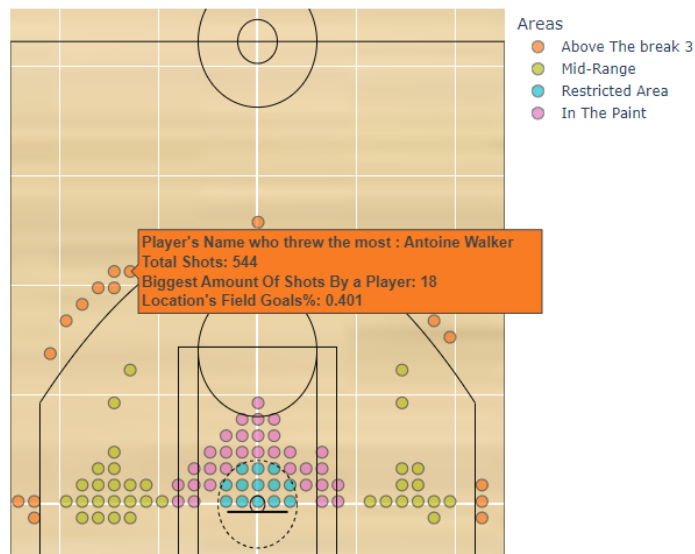
The Change In The Ratio Of Shots From Each Area
Ratio Of Shots Thrown From Area To The Total Amount Of Shots



תמונה 5 – נקודת ההתחלה של הגרפים

"זום אין"

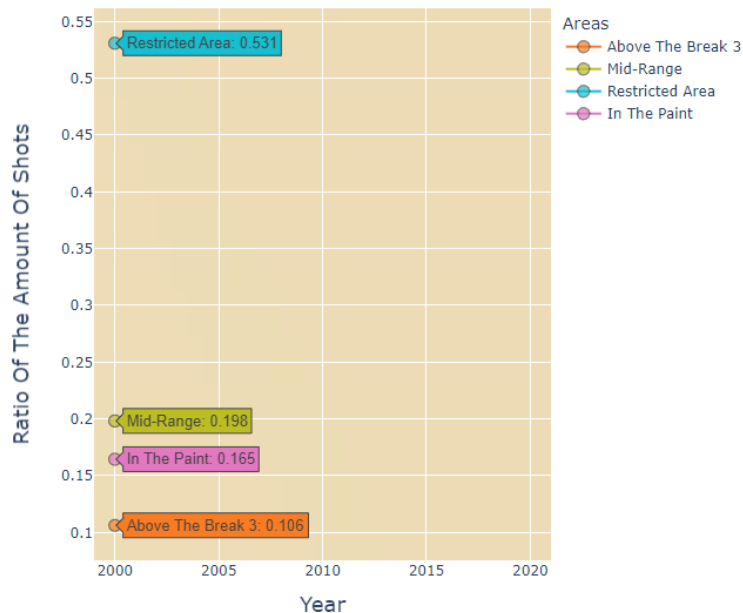
יתר פירוט עבור המשתמש – מעט סטטיסטיקות, יותר אינפורמציה.



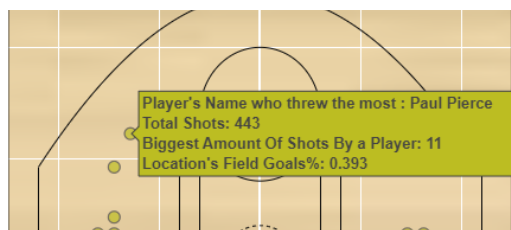
תמונה 6- סטטיסטיקות שניתן לראות על המגרש: כמות הזריקות, אחוז הזריקות שנקלעו, מי השחקן שקלע הכי הרבה באותה עמדה

וגם- אם המשתמש ירצה להתמקד רק עבור אזור אחד של המגרש- אז יש לו אפשרות:

כנ"ל – מבצע זאת עבור הויזואליזציה השנייה – מעניק את המידע של האחוז היחסי של כמות הזריקות של אותו אזור.



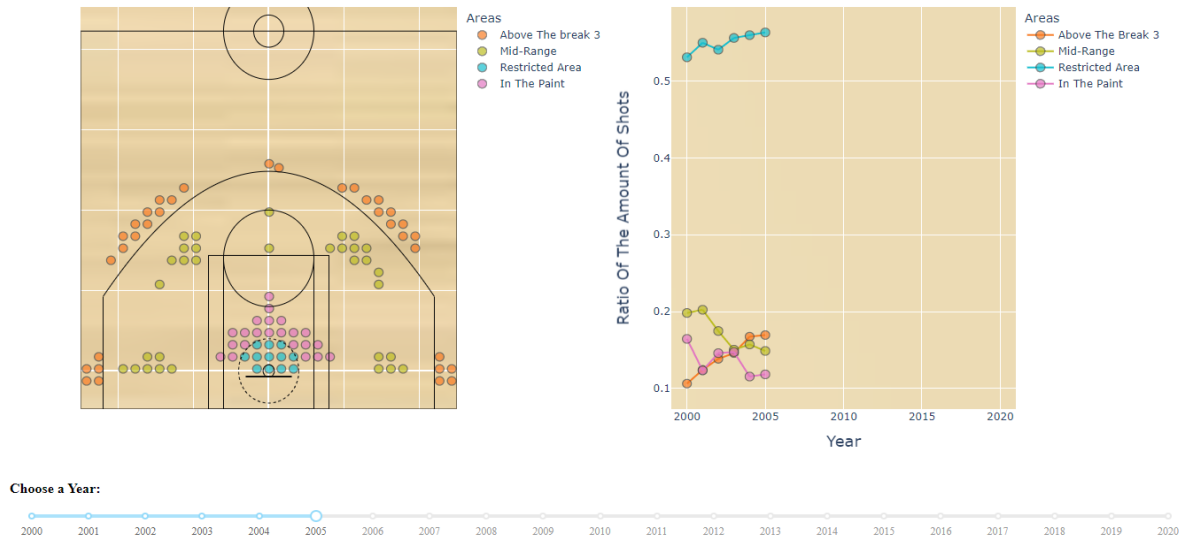
תמונה 7 – האחוז היחסי של כמות הזריקות בכל אזור מתוך סך הזריקות הכולל באותה עונה



תמונה 8 – התמקדות על אזור אחד בלבד ופירוט על הסטטיסטיקות הנכונות לאותו מיקום

מכיוון שאני מראה שינוי לאורך זמן, הוספתי ציר של עונות ה NBA (מ 2000 עד 2020) ולכן אני מעניק למשתמש אפשרות לגלול עם השנים.

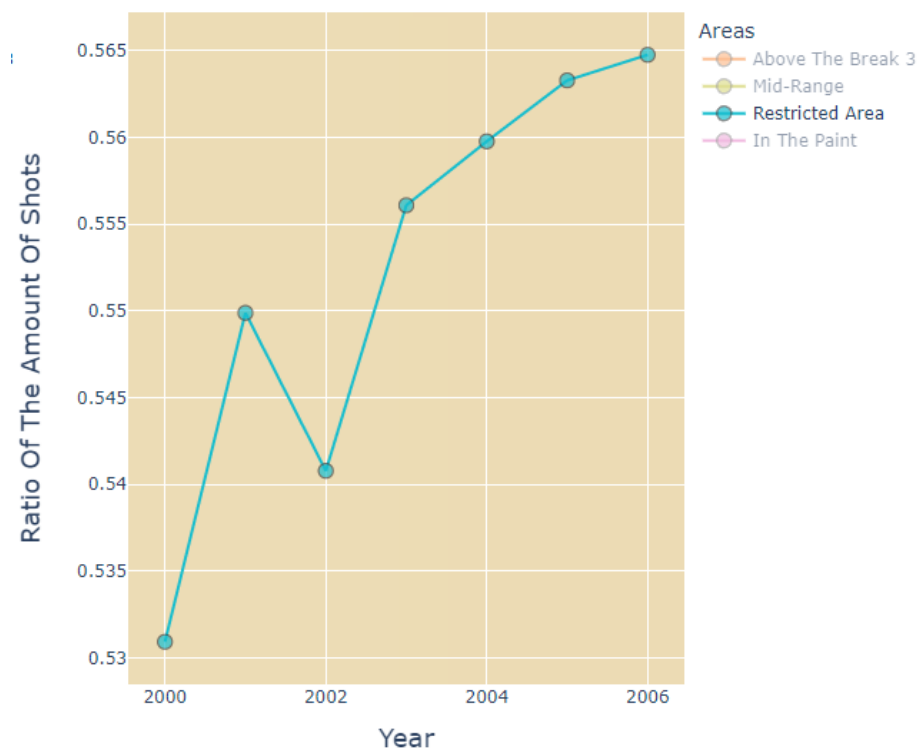
דוג' - עבור עונת 2005:



תמונה 9 – שימוש ב year-slider עבורו נראה את השינוי לאורך העונות ב NBA .

גם כאן, אם ארצה להתמקד עבור אזור מסוים- כתוצאה מכך ערכי הצירים משתנים (שנוכחים רק לאותו אזור) כדי שהגרף ישקף מציאות נכונה ולא מציאות שקרית.

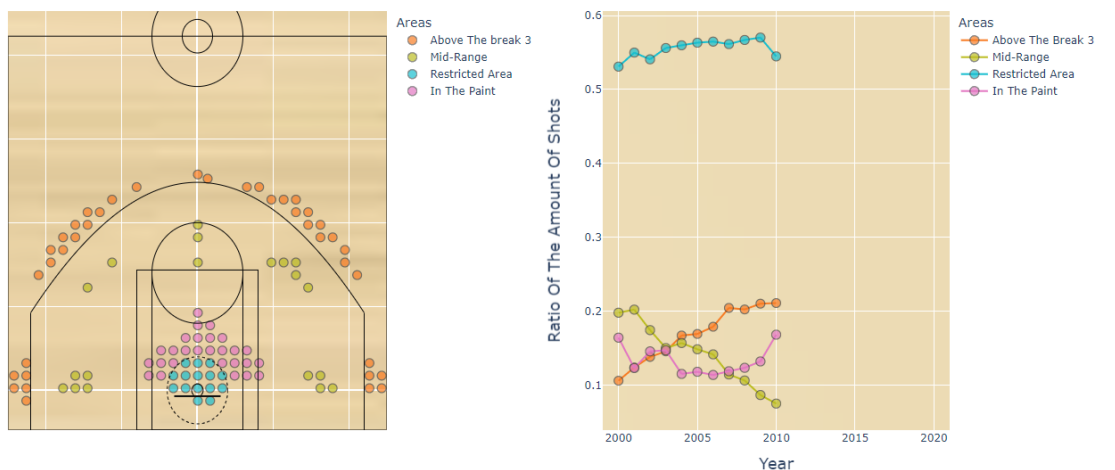
דוג' עבור Restricted Area:



תמונה 10 – שינוי ערכי הצירים בהתאם להתמקדות באזור מסוים

תובנות ומסקנות:

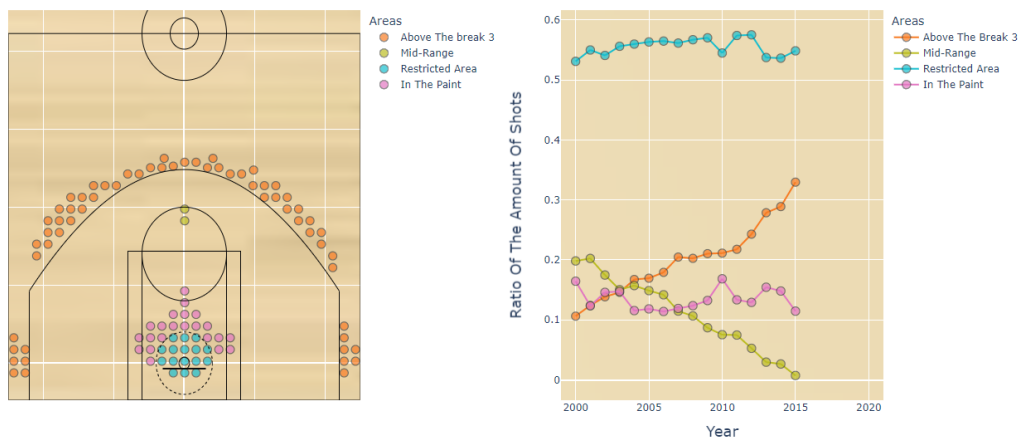
עונת 2010:



Choose a Year:



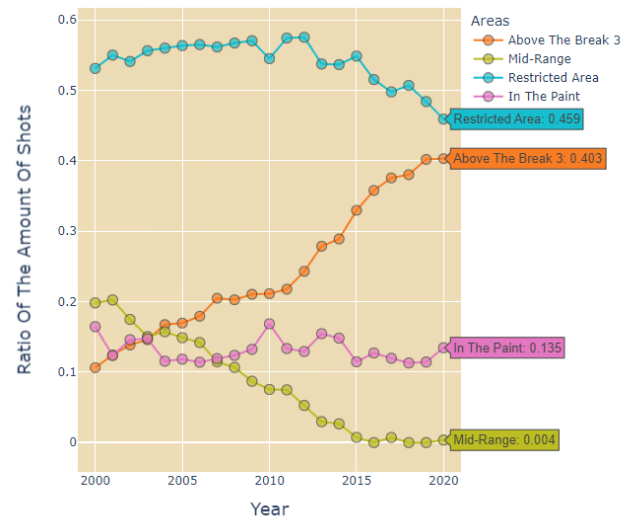
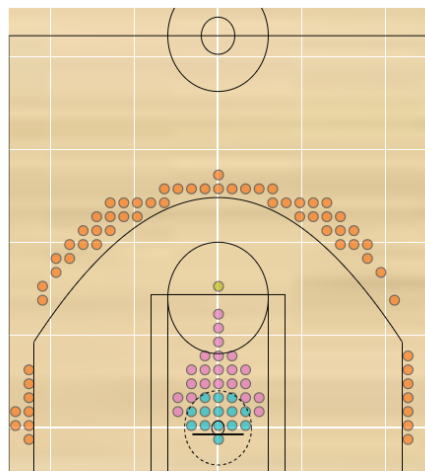
עונת 2015:



Choose a Year:



עונת 2020:



Choose a Year:



תמונות 11, 12, 13 – עונות 2010, 2015, 2020 בהתאמה.

ניתן להסיק מספר נקודות:

- כפי שניתן לראות יש שינוי עצום בין השנים. ככל שאנו מתקדמים עם השנים, אנחנו רואים עליה של הרבה יותר זריקות מעבר לקשת ה-3 באופן משמעותי יותר מהאחרים (מבחינת המגמה). בשעה שבאזורים אחרים אנחנו רואים מגמה של ירידה, בזריקות מעבר לקשת ה-3 אנחנו רואים מגמה של עליה.
- בנוסף לכך, אנחנו רואים במהלך השנים שינוי בסגנון המשחק – ממשחק די מפוזר, של הרבה יותר בפנים וזריקות ל-2 נקודות באופן מרוקח יותר (חצי מרחק), כעת אנחנו רואים משחק שונה לחלוטין! משחק של "בפנים או בחוץ", או באזור שקרוב לסל או באזור שרחוק מאוד מן הסל, באזור ה-3.
- ניתן גם לראות שהכדורסל יותר מתמקד לכיוון המרכז-חצי פינה – פחות לאזור הפינות.
- תובנות אלו יכולות לעזור מאוד לשחקנים צעירים, לשחקנים שברצונם להתפתח יותר באופן מקצועי בכדורסל. שחקנים צעירים יכולים להבין מכך שכיום כלי מרכזי ועיקרי זו זריקה מחוץ לקשת ה-3 שבלי כלי זה יהיה קשה יותר להתמקצע בכדורסל של ימינו. בנוסף לכך, ניתן לראות שכלי עיקרי שהיה בעבר-ה"חצי מרחק"- כמעט ואין צורך בו, לכן שחקנים צעירים יעבדו בדגשים של דברים אחרים ויעבדו פחות על דגשים הקשורים לחצי מרחק.

- ,תובנות אלו יכולות לעזור גם למאמנים, מאמנים יכולים להתאים את שיטת המשחק שלהם למגמה שהולכת כיום/ לעשות ההפך ואולי להצליח.
- המסקנה שניתן להסיק מן הגרף היא "למה צריך לזרוק רחוק מן הסל שאותו מרחק מביא לי רק 2 נקודות?". עדיף שאני אתרחק עוד טיפה כדי להשיג יותר נקודות. "3 יותר שווה מ 2". לכן ניתן לראות קיצוניות כזאת. אם לזרוק מ 2 אז לזרוק מקרוב ואם לזרוק מרחוק אז לזרוק מ 3.

5.3 קוד וחבילות:

לצורך כתיבת הקוד השתמשתי בחבילות למען שלבים שונים ביצירת הוויזואליזציה.

לצורך שלב הכנת הנתונים:

- nba_api – מחבילה זו לקחתי נתוני הזריקות במהלך העונות 2000-2020 (השתמשתי בחבילה של nba_api.stats.endpoints ובתוכה ניגשתי לחבילה ב ShotChartDetail).
- Pandas - מכיוון שהנתונים מיובאים ב data frames השתמשתי ב pandas, חבילה המוכרת לעבודה עבור data frames.
- time - הסיבה לכך היא ש shotchartDeatail יש עבור כל עונה בנפרד, כתוצאה מכך הייתי צריך לגשת אל נתוני הזריקות בכל עונה באופן נפרד (לדוג': פעם אחת עבור עונת 2000 ופעם אחרת עבור עונת 2001). nba_api לא מקבל בקשות רצופות אחת אחרי השנייה בזמן קצר מאוד. על כן, לאחר כל בקשה, אני מחכה שנייה לפני הבקשה הבאה של הנתונים לעונה הבאה.

לצורך הקמת dashboard, ליצירת הוויזואליזציה-

- Plotly - חבילה בפיתוח המייצרת אינטראקציה עם הגרף. (כגון views , mouse hover וכו..)
- Dash – חבילה המייצרת dashboard של ויזואליזציות ומשתמשת ב ברכיבים של HTML כדי לעצב את ה dashboard. בנוסף לכך, ניתן בעזרתה לממש אינטראקציה עם הוויזואליזציה לדוג': הוספת כפתור שתשנה את רקע הגרף. בוויזואליזציה אצלי, השתמשתי ב year-slider שבו אני גולל את השנה שבה אני ארצה לראות את נתוני 100 הזריקות הכי פופולריים.
- PIL – מכיוון שרציתי לשים תמונת רקע לוויזואליזציה של מגרש הכדורסל, השתמשתי בחבילה שפותחת תמונות.
- JupyterLab - חבילה המקימה שרת מקומי שעליה ניתן להריץ את הוויזואליזציה (את ה dashboard)

הסבר כללי על הקוד:

לצורך שלב הכנת הנתונים:

מספר שלבים:

- עבור כל עונה אני עובר על כל הנתוני זריקות (על כל ה data frame) ומשנה את המיקומים האנכיים והאופקיים של כל זריקה. (כפי שהוסבר בסעיף 5.1) ועבור כל מיקום מתוקן, אני סופר את כמות הזריקות שנזרקו מאותו מיקום.
 - מיינתי את הדאטה סט לפי כמות הזריקות שנזרקו מן המיקומים המתוקנים ולקח את טופ 100 המיקומים הכי נפוצים
 - עבור 100 המיקומים הנפוצים יצרתי מילון של מילונים כאשר מפתח המילון הוא המיקום האנכי והאופקי המתוקן ומפתחות המילונים הם שמות השחקנים שערכיהם זו רשימה של 2 ערכים: 1. כמות הזריקות שנזרקו 2. כמות הזריקות שהם קלעו. לאחר מכן מיינתי את המילונים הללו לפי כמות הזריקות שזרק כל שחקן כך שהערך הראשון יהיה שם השחקן שזרק הכי הרבה. כתוצאה מכך אני יכול גם לחשב את אחוזי הדיוק של השחקן באותו מיקום (מכיוון שיש לנו את כמות הזריקות שקלע וכמות הזריקות שזרק)
 - בנוסף, עבור כל אזור (יש 4 איזורים: 1. מעבר לקשת השלוש, 2. חצי-מרחק, 3. בצבע, 4. בעיגול שמתחת לסל- restricted area) חישבתי את כמות הזריקות באותו אזור ביחס לכמות הזריקות הכוללת, בשביל להראות את השינוי שקורה במהלך השנים ביחס לכל אזור.
- הסיבה לכך שחישבתי יחס ולא להראות כמות זריקות היא מכיוון שברצוני להראות עד כמה עיקרי ומרכזי נהיו זריקות מעבר לקשת ה 3 וכמה יותר זניח נהיו הזריקות מאזורים אחרים. לכן יחס במקרה הזה נותן מענה טוב לבעיה.

לצורך הקמת dashboard, ליצירת הוויזואליזציה-

מספר שלבים:

- יצירת מגרש הכדורסל לצורך ה shot chart. יצרתי את קו המגרש: החוץ, השלושה, העונשין, ה restricted area , חצי המגרש.
- יצירת ה dashboard: 1. יצירת כותרות וכותרות משנה של הוויזואליזציה.
2. יצירת ציר זמן אינטראקטיבי של שנים שהמשתמש יכול להזיז ולראות איך הוויזואליזציה משתנה בהתאם. (year slider).
- 3. יצירת הפונקציות של הוויזואליזציות- מכיוון שאני מציג 2 גרפים הייתי צריך לכתוב 2 פונקציות, אחת לכל גרף.
- 4. עבור כל אזור עשיתי סינון לפי שם האזור ובהתאם ייצגתי את הנתונים הנכונים לאותו אזור. (אני משתמש ב legend, לכן כדי להראות עבור כל ערך ב legend את הנתונים הנכונים לו, הייתי צריך לעשות סינון לפי ערך ה legend).

5. עיצוב ה dashboard: גודל ורוחב הגרף, רקע (תמונת רקע לגרף), שמות הצירים, כותרת הגרפים, עיצוב ה dashboard כך ששני הגרפים יהיו אחד מול השני, צבעי התצפיות לפי כל אזור.

6. יצירת mouse hover- כל פעם שנתקרב לתצפית נוכל לראות נתונים נוספים (כפי שמוצג בחלק 5.2).

- הקמת השרת באמצעות ngrok שנוכל להריץ את הוויזואליזציה וליצור עימה את האינטראקציות.

6. פירוט התהליך: