

UNIVERSITÁ DEGLI STUDI DI MILANO-BICOCCA

Scuola di Economia e Statistica

Corso di laurea in

SCIENZE STATISTICHE ED ECONOMICHE



UN'ANALISI SULL'IMPATTO DELLA STATISTICA NEL  
MONDO DEL BASKET:  
IL CASO DELLA GERMANI BRESCIA

Relatore: Prof. Andrea Marletta

Tesi di Laurea di:

Francesco Riboni

Matricola 889568

Anno Accademico 2023/2024

# Sommario

<b>INTRODUZIONE .....</b>	<b>3</b>
<b>1 L'IMPATTO DELLA STATISTICA NELLO SPORT .....</b>	<b>5</b>
1.1 Lo sport e il basket.....	5
1.1.1 Il basket .....	6
1.1.2 Federazioni e competizioni .....	8
1.2 La statistica e i big data .....	10
1.2.1 Il concetto di statistica .....	10
1.2.2 I Big Data .....	11
1.3 Sport Analytics.....	14
1.4 Basketball Analytics.....	17
1.4.1 Principali statistiche nella misura delle performance .....	20
1.5 La basket analytics nella LBA .....	23
<b>2 IL TIROCINIO CON LA GERMANI BRESCIA .....</b>	<b>25</b>
2.1 BDsports.....	25
2.2 L'esperienza di tirocinio .....	25
2.3 L'estrazione dei dati .....	26
2.4 L'analisi dei dati raccolti .....	27
2.5 La costruzione dei poligoni .....	28
2.6 L'integrazione tra i poligoni e i dati raccolti .....	30
2.7 La rappresentazione grafica .....	31
<b>3 TECNICHE UTILIZZATE .....</b>	<b>36</b>
3.1 Python .....	36
3.1.1 Scraper.....	37
3.1.2 Creazione poligoni .....	38
3.1.3 Associazione tra dati e settori .....	39
3.2 R ed RStudio .....	39
3.2.1 Script realizzati in R.....	40
3.3 Tableau.....	43
3.3.1 La costruzione delle dashboard .....	43

<b>CONCLUSIONI.....</b>	<b>48</b>
<b>BIBLIOGRAFIA .....</b>	<b>50</b>
<b>SITOGRAFIA.....</b>	<b>51</b>
<b>RINGRAZIAMENTI .....</b>	<b>52</b>

# Introduzione

Lo sport è un mondo straordinario, contraddistinto da una costante ricerca dell'eccellenza e da una competizione serrata, in cui ogni piccolo dettaglio può fare la differenza ed essere il punto di svolta tra una sconfitta e il successo. Per questo motivo, la statistica e i big data hanno assunto un ruolo rilevante in questo contesto, passando da semplice strumento di supporto a elemento cruciale nell'ottimizzazione delle performance, nel selezionare le strategie e nelle scelte in altri ambiti come in quello medico, per ridurre le probabilità di infortuni o nel campo del marketing, per garantire maggiori ritorni economici a società e atleti. La capacità di analizzare e interpretare una quantità sempre maggiore di dati ha permesso agli allenatori, ai dirigenti e agli atleti stessi di spingere i limiti dei vari sport sempre più in alto, cambiando radicalmente il modo in cui questi vengono vissuti e gestiti.

Nel basket, l'uso della statistica su grosse quantità di dati ha rivoluzionato molti aspetti del gioco, influenzando tattiche e strategie di quest'ultimo, ma anche decisioni manageriali e di scouting nella ricerca di nuovi talenti, condizionando in modo significativo i risultati sul campo e la gestione delle squadre. La statistica si presenta quindi come una potente arma che permette di migliorare ogni aspetto del gioco.

Questo lavoro di tesi si propone di esaminare l'impatto della statistica nel contesto del basket professionistico, focalizzandosi sul caso della Germani Brescia, squadra che nella stagione 2023/24 ha partecipato alla Lega Basket Serie A, informalmente nota anche solo come Serie A, la massima divisione professionistica del campionato italiano di pallacanestro. Durante questo periodo, più precisamente da gennaio a maggio del 2024, ho svolto un tirocinio presso questa squadra grazie ad una collaborazione con l'Università di Brescia, in cui ho avuto l'opportunità di applicare concretamente le competenze maturate durante il percorso di studi. In questa esperienza, svolta insieme ad un altro studente dell'Università di Milano Bicocca, Fabio Galante, il mio compito era quello di pubblicare regolarmente dei report statistici con cui lo staff tecnico avrebbe potuto preparare al meglio le partite e valutare le performance individuali dei membri della squadra e degli avversari.

Prima di affrontare nel dettaglio il lavoro svolto durante il tirocinio, la tesi presenterà un primo capitolo in cui verrà approfondita l'incidenza che questa scienza ha avuto nel corso degli anni nell'ambiente sportivo, affrontando i casi più lampanti per dimostrare come la

statistica sia diventata uno strumento fondamentale per le squadre professionistiche di alto livello.

In seguito, sarà quindi trattato il lavoro svolto per la squadra, mostrando le fasi del processo che ha portato alla realizzazione delle dashboard rappresentanti le analisi effettuate, facendo il possibile per renderle quanto più intuitive e chiare agli occhi dello staff.

Nel terzo ed ultimo capitolo, la tesi si concentrerà sui linguaggi di programmazione, i software e le tecniche impiegate nella realizzazione dei report, soffermandosi sulla loro importanza e sul modo in cui sono stati utilizzati.

Attraverso l'analisi del caso Germani Brescia, questo elaborato cercherà di mostrare l'efficacia della statistica come mezzo di supporto ed il suo potenziale nel migliorare le strategie e la gestione delle squadre, nel breve ma, soprattutto, nel lungo periodo, in modo tale da costruire un vantaggio competitivo duraturo.

# 1 L'impatto della statistica nello sport

Il capitolo iniziale della tesi si aprirà dunque con una breve descrizione dei benefici dello sport per la salute fisica e mentale di ognuno di noi. Successivamente, l'attenzione si concentrerà sulle due aree d'interesse della tesi: il basket e l'analisi statistica. Verrà introdotto questo gioco con le sue regole principali e verranno definiti i concetti di big data e statistica. Infine, andremo ad analizzare come questi due mondi si incontrano, mostrando le modalità in cui questa scienza riesce a migliorare le prestazioni degli atleti e citando alcuni casi che hanno cambiato notevolmente il suo ruolo nello sport e, più precisamente, nella pallacanestro.

## 1.1 Lo sport e il basket

“Lo sport ha il potere di cambiare il mondo. Ha il potere di ispirare. Ha il potere di unire le persone in un modo che poche altre cose fanno.” Con queste parole, il premio Nobel per la pace Nelson Mandela cattura l'essenza profonda dello sport nella vita quotidiana. Esso non costituisce solo un'attività fisica, ma è anche un mezzo per connettere le persone e distruggere barriere.

Praticare sport permette di condurre una vita sana, essere più resistenti, forti e flessibili, ma la sua importanza va ben oltre il benessere fisico. È una scuola di vita dove si allenano valori essenziali come la costanza e la determinazione, il rispetto delle regole e il lavoro di squadra.

Tutto ciò è alimentato dalla passione per lo sport che nasce dal piacere del movimento e dalla volontà di sfidare sé stessi e gli altri per tentare di superare i propri limiti. Proprio questa passione è fondamentale per spingerci a cercare di migliorare con costanza, a dedicare tempo ed energie per raggiungere nuovi obiettivi.

Il desiderio di raggiungere il “limite” permette a ognuno di noi di conoscere il suo vero potenziale e a misurare la propria forza di volontà attraverso il coraggio e la determinazione.

Lo sport rappresenta perciò una metafora della vita stessa: insegna a compiere sacrifici, a combattere per qualcosa, a rispettare delle regole e a rialzarsi dopo una caduta, allenando ad affrontare le sfide quotidiane con uno spirito combattivo e positivo.

### 1.1.1 Il basket

La pallacanestro (detta più comunemente basket, abbreviazione di basketball) è uno sport nato alla fine del diciannovesimo secolo negli Stati Uniti d'America grazie all'idea di un insegnante di educazione fisica canadese, James Naismith. Negli anni si è diffuso enormemente diventando ad oggi uno degli sport più diffusi e praticati al mondo.

Le regole di questo gioco sono numerose e alcune anche complesse, per questo motivo saranno trattate solo le regole principali, sufficienti a poter affrontare i temi centrali della tesi.

La pallacanestro è uno sport di squadra dove ad affrontarsi sono due squadre composte da cinque giocatori ciascuna. Il campo da gioco è un rettangolo lungo 28 metri e largo 15 metri, il campo è diviso in due parti dalla linea di centrocampo in modo da ottenere due rettangoli 14x15 metri. Ogni metà possiede un canestro e lo scopo del gioco è quello di segnare punti facendo entrare un pallone nel canestro. Esso è situato, per regolamento, ad una altezza di 3,05 metri, con una circonferenza di 45 centimetri ed il pallone ha caratteristiche diverse a seconda delle competizioni a cui sono destinati. Nelle competizioni femminili, la sfera deve avere una circonferenza tra i 72,4 e i 73,7 centimetri con un peso compreso tra i 510 grammi e i 567 grammi; nelle competizioni maschili, invece, la circonferenza ha una circonferenza tra 74,9 e i 78 centimetri con un peso tra i 567 e 650 grammi.

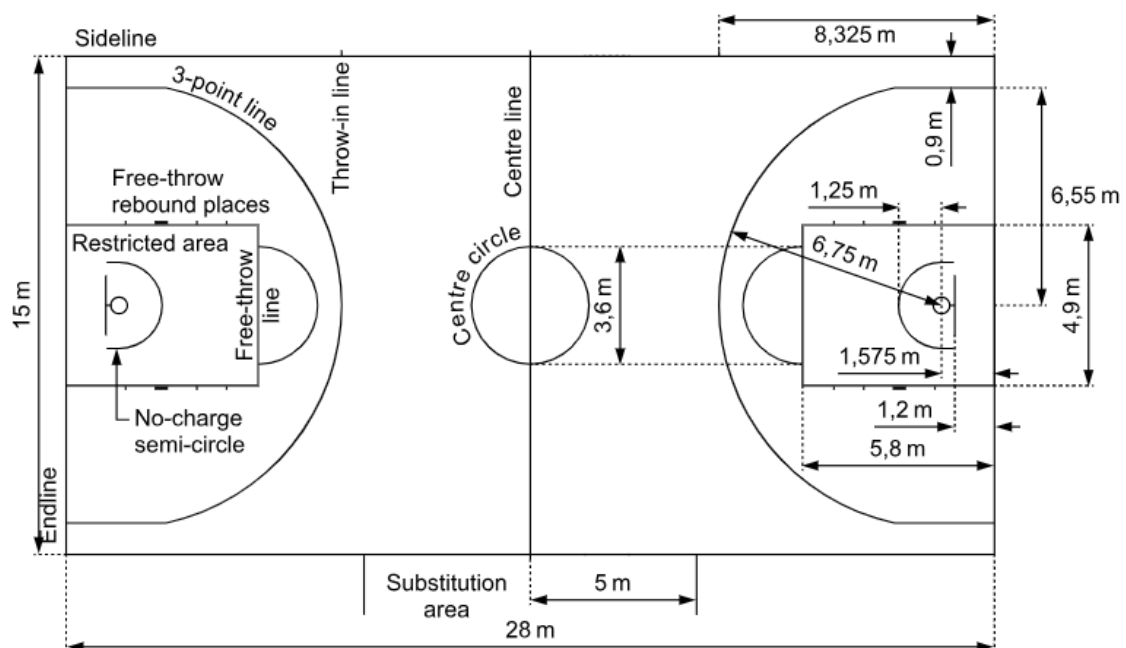
Una partita è divisa in quattro quarti, ognuno della durata di 10 minuti, per un totale di 40 minuti, anche se in alcune competizioni la durata è di 12 minuti per quarto. Al termine della partita la squadra che ha segnato più punti vince la partita; se finisce in pareggio, vengono istituiti uno o più tempi supplementari di 5 minuti ciascuno, fino a quando al termine di uno di essi una delle due squadre ha segnato più punti. La partita inizia con una palla a due al centro del campo. L'arbitro lancia la palla in aria tra due giocatori e uno di loro cerca di deviarla verso un compagno.

Ogni squadra solitamente ha a disposizione 12 giocatori ed il numero di sostituzioni possibili è illimitato; un giocatore che viene sostituito può persino rientrare in campo in seguito.

I punteggi che vengono assegnati dopo un canestro variano a seconda della tipologia di tiro:

- un punto: viene assegnato dopo aver effettuato un tiro libero, ovvero un tiro che viene concesso a un giocatore dopo che ha subito un determinato tipo di fallo, senza la pressione degli avversari ad una distanza di 5,8 metri dal canestro;
- due punti: vengono assegnati quando un giocatore segna un canestro all'interno dell'arco da tre punti, una linea solitamente bianca o nera visibile sul terreno di gioco;
- tre punti: quando il tiro segnato è partito al di fuori dell'arco da tre punti.

Questa linea, secondo il regolamento FIBA, è situata ad una distanza di 6,75 metri dal canestro, anche se in altri contesti queste misure possono variare.



*Figura 1. Campo di pallacanestro FIBA.*

Ogni squadra, dal momento in cui entra in possesso del pallone di gioco, ha a disposizione 24 secondi per tentare una conclusione, mentre, se una squadra dopo un tentativo di tiro sbagliato raccoglie il pallone e resta in attacco, ha a disposizione 14 secondi per tirare un'altra volta. Inoltre, all'inizio del possesso, ogni squadra deve superare la linea di metacampo entro 8 secondi e una volta superata non è possibile tornare indietro. Queste regole di tempo danno una maggiore fluidità al gioco, rendendolo più dinamico e impedisce alla squadra che si trova in vantaggio di far trascorrere il tempo a lungo con la palla nelle proprie mani.



Per quanto riguarda il movimento dei giocatori, è importante ricordare che ogni giocatore, per potersi muovere con il pallone, deve palleggiare il pallone a terra mentre si muove. Prima e dopo la fase di palleggio, il giocatore non può fare più di due passi con la palla tra le mani, mentre incorre in un'infrazione di "doppio palleggio" se dopo aver arrestato il palleggio bloccando la palla, egli ricomincia a palleggiare.

Un altro aspetto che caratterizza notevolmente questo gioco sono le regole sui falli. Ogni giocatore durante una partita, può "spendere" fino a 5 falli (alcune competizioni ne prevedono 6), al termine dei quali viene espulso e non può più rientrare in campo per il resto della partita, mentre la squadra del giocatore può inserirne un altro dalla panchina. I falli non sono classificati tutti allo stesso modo; se un giocatore commette fallo ad un avversario che sta cercando di fare canestro, quest'ultimo ha diritto a dei tiri liberi. Se il giocatore che viene ostacolato riesce anche a fare canestro, questo risulta valido e il giocatore ha a disposizione un tiro supplementare. Altrimenti, il giocatore ha a disposizione 2 o 3 tiri liberi a seconda della zona in cui tira. Se il fallo non avviene durante un tentativo di tiro, allora la squadra che l'ha subito riparte con una semplice rimessa laterale.

Inoltre, durante ogni quarto, dopo che una squadra ha subito 5 falli entra in "bonus", ovvero a seguito di ogni fallo subito corrispondono 2 tiri liberi. Questo conteggio si azzerava nuovamente all'inizio di ogni quarto di gioco [FIBA, 2024].

Queste sono le regole principali di questo sport che, come detto all'inizio del paragrafo, ne prevede molte altre più complesse e tecniche, che non sono fondamentali ai fini della tesi.

### 1.1.2 Federazioni e competizioni

La federazione che regola il basket a livello mondiale è la FIBA (federazione internazionale pallacanestro), la quale gestisce il regolamento di gioco, organizza i campionati mondiali e continentali e coordina quelli nazionali. Nasce nel 1932 a Ginevra e, ad oggi, è riconosciuta come la sola organizzazione di pallacanestro mondiale dal Comitato Olimpico. Questa federazione organizza la coppa del mondo FIBA, i campionati continentali (tra i quali il più prestigioso è sicuramente l'EuroBasket) e supervisiona il torneo olimpico di basket.

Nonostante la sua grande importanza, la FIBA non ha nessuna influenza sulla lega professionistica di basket più competitiva e famosa al mondo, la National Basketball Association. La NBA nasce nel 1946 sotto l'acronimo di BAA ed ha sede negli Stati Uniti. Questa lega si distingue dalle altre competizioni per diversi motivi:

- Un calendario fittissimo: la lega è composta da 30 squadre o franchigie, ognuna di esse durante la cosiddetta Regular Season gioca 82 partite ed al termine della stagione regolare, le migliori 16 squadre si sfidano ai playoff. Le squadre che raggiungono le prestigiose NBA Finals, perciò, tendono a superare le 100 partite disputate in una stagione.
- Un regolamento diverso: a differenza del regolamento FIBA che è stato descritto ad inizio capitolo, la NBA presenta alcune variazioni. La durata totale di gioco è di 48 minuti (12 minuti per quarto), il campo è leggermente più grande e l'arco dei tre punti è posizionato ad una distanza di 7,25 metri dal canestro, anche se ai lati del campo tale linea si trova a 6,72 metri dal canestro. Inoltre, un giocatore viene espulso dopo aver commesso 6 falli anziché 5. Vi sono poi una serie di regole più tecniche che hanno delle piccole differenze rispetto al regolamento internazionale.
- La maggior parte dei migliori giocatori al mondo gioca in questa lega, che perciò è considerata da tutti la più competitiva al mondo, tanto che, spesso negli Stati Uniti i vincitori del campionato vengono definiti “campioni del mondo”.

A livello europeo, la competizione di spicco è l'Eurolega, una lega semichiusa di cui fanno parte ogni anno le migliori squadre dei campionati europei che fanno parte della FIBA Europe. L'organismo che organizza questa competizione è l'Euroleague Basketball. Il campionato è composto da 18 squadre, 11 delle quali sono fisse, mentre le restanti 7 possono cambiare annualmente in base alle prestazioni delle squadre. Ogni squadra affronta due volte le altre 17 nel corso del campionato per un totale di 34 partite. A seguito di esse, le migliori otto accedono ai playoff e successivamente la competizione termina con le Final Four per decretare il vincitore. Oltre a questa competizione, la federazione organizza anche l'EuroCup, un torneo anch'esso prestigioso, ma sicuramente di livello inferiore.

Inoltre, ogni paese organizza i campionati nazionali. I più prestigiosi sono la Liga ACB in Spagna, la Basketball Bundesliga (BBL) in Germania e la Lega Serie A (LBA), di cui fa parte la Germani Brescia.

La LBA è la massima divisione del campionato professionistico italiano di pallacanestro, che vide i primi tornei nazionali giocarsi nel 1920 ed è organizzato dalla Federazione Italiana di Pallacanestro (FIP) dal 1972. La lega è composta da 16 squadre che si affrontano due volte ciascuna durante la stagione regolare per un totale di 30 partite, dopodiché, le prime 8 si affrontano nei playoff in sfide al meglio delle 5 partite fino alla finale che decreta la società vincitrice dello scudetto [<https://www.legabasket.it/la-lega/storia>].

## 1.2 La statistica e i big data

### 1.2.1 Il concetto di statistica

La statistica è una disciplina che studia dei fenomeni con lo scopo di mettere in evidenza gli aspetti essenziali, cercando di risalire eventualmente alle leggi che li regolano. Essa ci permetta di trasformare dati grezzi in informazioni utili, attraverso la raccolta, l'interpretazione, l'analisi e la presentazione dei dati. In molti casi la statistica rappresenta il mezzo più efficace per ridurre il margine di incertezza delle nostre scelte.

La statistica può essere sostanzialmente suddivisa in due tipologie a seconda dell'analisi che si vuole eseguire:

- Statistica descrittiva: ha il compito di presentare e descrivere un certo fenomeno in maniera chiara e sintetica. L'analisi è limitata ai dati raccolti e non cerca di andare oltre a ciò che è stato osservato.

Le principali tecniche di questo tipo di analisi riguardano l'uso di indicatori di posizione (ad esempio la media e la mediana), misure di dispersione (come la varianza) e distribuzioni di frequenza. In questo ambito, assumono molta importanza grafici e tabelle che sono strumenti fondamentali nella divulgazione delle analisi effettuate, in modo da rappresentarle in maniera più chiara e immediata a chi non ha le conoscenze e il tempo per leggere tutte le rilevazioni effettuate. Gli strumenti grafici più frequenti sono gli istogrammi, i diagrammi a torta e i box plot, dei grafici che mostrano la distribuzione dei dati individuando eventuali valori anomali.

- Statistica inferenziale: a differenza della statistica descrittiva, partendo da un numero limitato di osservazioni, si occupa di fare previsioni su una popolazione più ampia. Lo scopo di queste analisi è quello di generalizzare i risultati ottenuti dal campione selezionato su una popolazione a cui appartengono gli elementi del campione. Nella statistica inferenziale è molto importante valutare l'incertezza e la variabilità dei risultati ottenuti. Questo tipo di analisi è fondamentale quando la popolazione oggetto di indagine è troppo elevata o indefinita, ed effettuare rilevazioni su tutta la popolazione sarebbe molto costoso e richiederebbe troppo tempo, mentre in alcuni casi sarebbe addirittura impossibile.

Le tecniche utilizzate in questi casi sono più complesse rispetto a quelle descrittive e si concentrano sulla produzione di stime e sulla verifica della validità di esse. Nella produzione di stime vengono impiegati intervalli di confidenza o stime puntuali; mentre per verificare la validità di esse vengono impiegati i test di ipotesi.

Vi sono inoltre una serie di tecniche di analisi multivariata utilizzate per studiare e comprendere le relazioni tra più variabili in complessi insiemi di dati, al fine di classificare dati, sintetizzarli e stimare dei modelli per effettuare previsioni. Le tecniche principali impiegate nell'area multivariata sono le regressioni, la cluster analysis e l'analisi delle componenti principali (PCA) [Guseo, 2006].

Oggi vi sono numerosi organismi pubblici che si occupano di raccogliere e diffondere informazioni statistiche. A livello locale, gli enti territoriali, come comuni e province sono incaricati soprattutto di raccogliere dati di base, ossia le informazioni principali che verranno poi trasmesse agli enti centrali, i quali si occupano della verifica, dell'elaborazione e dell'aggregazione di tali dati. In Italia, l'ente centrale che si occupa di questo ruolo è l'ISTAT, mentre a livello europeo questo compito spetta all'EUROSTAT.

### 1.2.2 I Big Data

I big data, come lascia intendere il concetto stesso, sono insieme di dati molto vasti e complessi, che i software e gli strumenti tradizionali per gestire informazioni non sono in grado di gestire. Al giorno d'oggi, molte delle più grandi aziende al mondo investe molto tempo e denaro per raccogliere e analizzare questi dati, in modo tale da ottenere un vantaggio competitivo sulla concorrenza.

Questi dati sono molto utilizzati in tutti i settori dell'economia e della società, come il marketing, la sanità, l'ambiente e la ricerca scientifica. Il loro impiego, infatti, può garantire numerosi vantaggi, come ad esempio la riduzione dei costi e delle tempistiche in fase di produzione, lo sviluppo di nuovi prodotti e servizi che soddisfino le richieste del mercato o l'ottimizzazione di offerte e coupon per attirare i consumatori.

Il concetto di big data è relativamente nuovo, ma i primi sistemi di gestione dei database, risalgono agli anni '60 e '70 grazie ad aziende come IBM e Oracle. Successivamente, la creazione di internet ha dato una grossa mano allo sviluppo di questi database. Grazie alla nascita dei siti web, i dati a disposizione acquisirono sempre più varietà, divennero sempre più vasti e semplici da raggiungere.

Intorno ai primi anni 2000 prese piede il concetto di "Big Data", coniato negli anni '90 da John Mashey, un informatico che ha lavorato presso la Silicon Graphics, Inc., dove ha ricoperto il ruolo di chief scientist. Poco dopo, Doug Laney presentò per la prima volta il concetto delle "3V" dei big data in un rapporto di ricerca presentato da Meta Group. Le "3V" rimangono ancora oggi il modo più efficace per definire il concetto di big data e hanno contribuito alla diffusione del termine stesso, questo modello viene affrontato in modo più dettagliato in seguito nello stesso paragrafo.

Negli anni successivi, l'analisi dei big data divenne una vera e propria disciplina scientifica, la "Data Science". L'integrazione dell'intelligenza artificiale (IA) con i big data ha permesso lo sviluppo delle tecniche di machine learning, una branca dell'IA che si occupa dello sviluppo di algoritmi che permettono ai computer di analizzare dati e fare previsioni senza che essi vengano esplicitamente programmati per tali compiti [Tiao, 2024].

L'esplosione mediatica dei social network ha contribuito a una raccolta sempre più vasta di dati da parte delle aziende. Al giorno d'oggi, le grandi aziende gestiscono volumi impressionanti di dati nell'ordine dei petabyte, che corrispondono ad un milione di gigabyte, o, addirittura, in exabyte, equivalenti a mille petabyte. In futuro, la quantità di dati disponibili è destinata ad aumentare ancora notevolmente e la nascita di nuove tecniche di elaborazioni e analisi dei dati porterà a nuove scoperte e innovazioni in molti campi.

In un celebre report del 2001, Doug Laney, allora vicepresidente e Service Director dell'azienda Meta Group, definiva il modello delle 3V. Un modello semplice e sintetico

con cui Laney descrive le caratteristiche dei big data, che ancora oggi è alla base di questa tecnologia. Le 3V corrispondono a tre proprietà fondamentali che caratterizzano questi dati e che verranno affrontati nel dettaglio: volume, varietà e velocità.

- **Volume:** la quantità di dati a disposizione è molto grande, tale da rendere impossibile l'uso di tecnologie classiche per la raccolta. Il volume di dati oggi registra continui incrementi; ciò a causa della crescita di internet, dei social media e di altre fonti di dati.

Questo implica, una costante ricerca di nuove tecnologie di archiviazione e sistemi di gestione dei database in grado di organizzare ed elaborare petabyte di dati.

- **Varietà:** si riferisce ai molti tipi di dati disponibili. I dati tradizionali erano strutturati in modo da adattarsi a dei database relazionali. Con l'avvento dei big data, i dati possono presentarsi sottoforma di dati strutturati (che seguono un formato predefinito e organizzato), semi-strutturati (non seguono uno schema rigido, ma le informazioni sono parzialmente organizzate) e non strutturati (non seguono un formato predefinito e non sono organizzati). La varietà delle tipologie di dati richiede metodi di analisi avanzati e flessibili che siano in grado di gestirli tutti.
- **Velocità:** si riferisce alla rapidità con cui questi dati vengono generati, elaborati e analizzati. La diffusione di device dotati di sensori dedicati alla raccolta di dati in tempo reale ha contribuito a dare una forte spinta alla velocità con cui questi vengono gestiti.

Oggi le aziende devono affrontare una nuova sfida: oltre a raccogliere dati in modo rapido ed efficiente, devono analizzarli in tempo reale (real-time analytics). In questo modo le imprese possono rendere il processo decisionale il più tempestivo ed efficace possibile.

Con il passare degli anni, alcuni esperti si sono chiesti se fosse necessario espandere il modello di Laney ad altre dimensioni, in modo da mettere in risalto altre caratteristiche. Ad oggi, quindi, si parla del modello delle 5V dei big data, poiché sono state aggiunte altre due proprietà:

- **Valore:** si riferisce al potenziale valore che può essere estratto dai dati. Raccogliere grandi quantità di dati diventa superfluo se da essi non si riesce ad estrapolare informazioni utili; perciò, la quantità e la qualità dei dati assumono la stessa

importanza. Il valore dei dati, dunque, si può misurare in conoscenza, che emerge quando le informazioni estratte aiutano le imprese durante il processo decisionale, prendendo delle scelte che si concretizzeranno in azioni.

- Veridicità: spesso viene affermato che “Bad data is worse than no data”, ciò per sottolineare che dati di scarsa affidabilità possono essere pericolosi e causare più danni rispetto alla totale assenza di essi. I dati devono cercare di rappresentare il più possibile la realtà. Le aziende, in questo caso, devono impegnarsi nel prevenire la raccolta di informazioni inaffidabili e ordinare quelle disponibili cercando di risolvere eventuali incongruenze.

In alcuni casi viene considerata addirittura una sesta “V”, la variabilità. Essa si riferisce alla possibilità dei dati di assumere significati diversi a seconda del tipo di interpretazione. Ciò è dovuto alla rilevazione dei dati da diverse fonti e contesti, oltre che dalla loro disponibilità in diversi formati. Questa problematica tende ad accentuarsi quando diversi membri dell’azienda, con ruoli e conoscenze diverse, leggono ed interpretano questi dati [Rezzani, 2017].

Conoscere il modello delle 5V (o 6V) è importante, in quanto aiuta a comprendere le caratteristiche principali dei big data e permette di cogliere gli aspetti da tenere in considerazione durante il processo di trasformazione dei dati in informazioni da impiegare nei processi decisionali delle aziende, per garantire loro delle migliori prestazioni.

Infatti, i big data analytics consentono alle aziende di prendere decisioni consapevoli basate su dati concreti e su analisi oggettive piuttosto che su intuizioni o sensazioni e per fare ciò è imprescindibile conoscere le 5V e padroneggiarle.

### 1.3 Sport Analytics

Quando l’analisi di grandi quantità di dati entra in contatto con il mondo dello sport, parliamo allora di “Sport Analytics”. Questo campo è caratterizzato da una costante ricerca e raccolta di dati sulle prestazioni di giocatori e squadre sportive professionistiche, ma anche tutto ciò che riguarda le attività finanziarie e di marketing delle società, poiché ormai, alcune di esse costituiscono delle vere e proprie aziende.

A livello professionistico, ogni minimo dettaglio assume grande importanza, perciò l’utilizzo di nuove tecnologie, per raccogliere nuove informazioni può fare la differenza

per riuscire ad eccellere. In generale, le principali aree di applicazione in cui statistica e big data hanno maggiore influenza sono:

- **Analisi delle prestazioni:** con lo sviluppo di sensori, dispositivi GPS e videocamere, è possibile monitorare le prestazioni degli atleti anche in tempo reale. Ciò aiuta lo staff tecnico nella personalizzazione degli allenamenti, in modo da permettere agli atleti di esprimere tutto il loro potenziale.
- **Ottimizzazione delle strategie di gioco:** analizzando nel dettaglio i dati delle squadre avversarie, gli allenatori sono in grado di conoscere quelli che sono i punti di forza e di debolezza delle squadre che dovranno affrontare. Possono perciò sviluppare delle strategie a doc per affrontare le partite nel modo più efficace possibile.
- **Gestione della squadra:** analizzare le prestazioni collettive della squadra e le interazioni tra i giocatori aiuta ad ottimizzare le prestazioni. Per esempio, se due o più giocatori hanno una buona intesa e rendono al meglio quando si trovano in campo contemporaneamente, lo staff tecnico tenderà a inserirli in campo insieme per avere un vantaggio competitivo.
- **Prevenzione e recupero degli infortuni:** negli ultimi anni la salute fisica e mentale degli atleti è stato uno dei punti di maggiore interesse della sport analytics. Migliorare la tecnica degli atleti attraverso l'analisi biomeccanica e una maggiore attenzione ai carichi di lavoro imposti ha aiutato a prevenire gli infortuni. Inoltre, altrettanto cruciali sono diventate le analisi tecniche durante il percorso di recupero da un infortunio; confrontare le prestazioni precedenti ad un trauma con quelle più recenti è importante per capire se l'atleta sia tornato sui propri standard per evitare di sovraccaricarlo durante gli allenamenti, rischiando una ricaduta.
- **Osservazione e scouting:** anche la ricerca di nuovi talenti è diventata sempre più influenzata dall'analisi statistica. Grazie alla possibilità di raccogliere dati anche su sportivi giovani o stranieri, gli osservatori delle società più importanti e ricche riescono a confrontare ed analizzare le prestazioni e le potenzialità di centinaia di atleti emergenti, riuscendo ad ingaggiare i migliori talenti già a livello giovanile. Tali dati sono impiegati anche per determinare il valore economico dei giocatori.
- **Interazioni con i fan e marketing:** l'analisi dei dati influisce anche sul modo di coinvolgere i fan da parte di società e federazioni. Attraverso i social media, le



vendite di biglietti e merchandising delle squadre, queste sono in grado di individuare le modalità con cui attirare il maggior numero di tifosi.

- Automatizzazione delle regole: i grandi passi in avanti della tecnologia e la grande quantità di dati a disposizione hanno contribuito a rendere più automatiche e oggettive alcune decisioni arbitrali. Ad esempio, l'utilizzo del cosiddetto "occhio di falco" nel tennis per determinare il punto preciso in cui la pallina rimbalza, oppure il caso del fuorigioco semiautomatico nel calcio [Vermeulen, 2018].

L'uso di dati e statistiche per prendere decisioni più informate in ambito sportivo risale agli anni '50 e '60. In questo periodo, seppur con mezzi molto limitati, le società più all'avanguardia di baseball, calcio e basket iniziarono ad utilizzare questi strumenti per riuscire ad eccellere. Negli anni queste analisi subirono forti sviluppi e, si iniziò a riconoscere la loro importanza per migliorare le performance.

Per primo, lo sport che più di tutti fu influenzato dalla statistica fu il baseball. Negli anni '70 e '80, Bill James, un appassionato di baseball e statistico autodidatta, rivoluzionò questo sport attraverso la creazione e lo sviluppo della "Sabermetrica" ("Sabermetrics" in inglese), un'organizzazione dedicata allo studio e all'analisi del baseball. James iniziò a scrivere alcuni articoli in cui criticava le statistiche utilizzate per valutare le prestazioni dei giocatori, sostenendo che molte di esse erano fuorvianti e limitate, in quanto non riuscivano a spiegare il vero contributo degli atleti alla squadra. Propose invece una serie di nuove statistiche più complesse, ma che avrebbero potuto considerare aspetti più rilevanti del gioco. Introdusse i concetti di "on base percentage" e "slugging percentage" in cui tentò di misurare la frequenza con cui un giocatore raggiungeva la base e la potenza con cui ogni battitore riusciva a colpire. Introdusse anche lo "win shares", ovvero un indicatore che cercava di quantificare l'effettivo contributo di un giocatore alla vittoria della squadra, statistica che, col tempo, venne introdotta anche in altri sport di squadra.

La Sabermetrica guadagnò popolarità tra fan e studiosi di questo sport, ma venne inizialmente ignorata dalle squadre della Major League Baseball (MLB), la lega americana di baseball più prestigiosa al mondo.

Un punto di svolta si verificò nel 2001 a Oakland grazie a Billy Beane, il general manager (GM) della squadra della città, gli Oakland Athletics. In quell'anno, la società perse alcuni dei suoi giocatori più importanti e il budget a disposizione per realizzare una squadra competitiva era basso. Beane decise quindi di acquistare una serie di giocatori sottostimati

con delle ottime statistiche in merito alla capacità di conquistare basi. Inizialmente, però, l'allenatore della squadra non seguì la filosofia del GM e l'inizio di campionato non fu tra i migliori. Nella seconda parte di campionato, Beane convinse il coach a far giocare più spesso i nuovi giocatori e il risultato fu straordinario: la squadra ottenne 20 vittorie consecutive stabilendo un nuovo record e terminò la stagione al primo posto. Gli Oakland Athletics non riuscirono poi a vincere i playoff, ma il percorso della squadra colpì tutto il mondo del baseball, spingendo le altre società ad investire nelle analisi delle prestazioni dei giocatori.

La storia di Billy Beane e degli Oakland è stata resa celebre dal libro "Moneyball" di Michael Lewis [Lewis, 2004] che ebbe un successo enorme, tanto che nel 2011 venne realizzato l'omonimo film, in cui l'attore Brad Pitt interpretava la parte di Billy Beane.

Oggi la Sabermetrica ricopre un ruolo indispensabile nella scelta delle strategie ed ha influenzato anche altri sport, soprattutto nel basket, che verrà trattato nel paragrafo successivo; un caso simile ispirato proprio a Billy Beane è avvenuto negli ultimi anni nel calcio. Nel 2012 Matthew Benham ha acquistato il Brentford, squadra inglese che militava nella terza serie nazionale e due anni dopo anche il FC Midtjylland, una squadra di serie A danese. Benham decise di seguire l'esempio degli Oakland Athletics, utilizzando un approccio analitico per acquisire e far crescere i propri atleti. In questo modo, ha portato il Brentford a partecipare alla maggiore competizione di calcio in Inghilterra, la Premier League; mentre il Midtjylland negli ultimi anni è riuscito a conquistare quattro titoli del campionato danese, i primi della sua storia, facendosi conoscere anche a livello europeo [Zanni, 2022].

## 1.4 Basketball Analytics

La pallacanestro è sicuramente tra gli sport più influenzati dalla data analysis. La fluidità del gioco, in cui le squadre passano costantemente da attacco a difesa, la facilità nell'identificare ogni minimo aspetto del gioco, aiutato da un campo di gioco relativamente piccolo e da soli 5 giocatori per squadra, permette la rilevazione di una grande quantità di dati da analizzare a fini statistici. La forte diffusione di analisi statistiche è legata sicuramente al caso degli Oakland nel baseball, che ha condizionato anche la pallacanestro, e soprattutto, la lega americana NBA. Oggi, tutte le franchigie della lega dipendono fortemente da queste analisi e ogni aspetto delle strategie di gioco,

della gestione della squadra e dell'osservazione dei giovani talenti è regolato da statistiche e dati.

Allenatori e staff tecnici, tendono a impostare il proprio gioco in maniera tale da creare situazioni che sono statisticamente più efficienti per la squadra oppure provano a creare situazioni di gioco statisticamente sfavorevoli per gli avversari. Questo approccio ha rivoluzionato, negli ultimi anni, il modo in cui le squadre affrontano le partite [Sarlis, 2020].

Il pioniere di questa rivoluzione è stato Daryl Morey, general manager degli Houston Rockets dal 2007, squadra della NBA. Morey adottò un'impostazione sistematica e basata sull'analisi di dati per prendere decisioni strategiche. L'obiettivo del GM dei Rockets era quello di massimizzare l'efficienza offensiva della sua squadra.

Le analisi dei Rockets portarono a una conclusione drastica: tentare di eliminare totalmente il tiro dalla media distanza (mid-range shot) o "Long Two" come dicono negli Stati Uniti, ovvero il tiro da due punti effettuato tra l'area adiacente al canestro e l'arco da tre punti.

La filosofia della franchigia prese il nome di "Moreyball", ricordando il caso "Moneyball" degli Oakland Athletics; i punti chiave di questo gioco erano sostanzialmente tre:

1. Velocizzare il gioco: adottando uno stile di gioco più veloce e fluido, sfruttando il mal posizionamento degli avversari nelle transizioni da attacco a difesa, la squadra riusciva a trovare tiri meno contestati e più efficienti.
2. Aumentare i tiri da tre punti: Morey ed il suo team riuscirono a dimostrare che i tiri da tre punti offrono un valore atteso maggiore rispetto ai tiri dalla media distanza. Di conseguenza, la squadra aumentò notevolmente il volume dei tiri da tre punti a discapito dei Long Two, aumentando così l'efficienza.
3. Aumentare i tiri liberi: i tiri da tre punti non erano però i tiri più pregiati, bensì i tiri liberi. Poiché, una squadra ottiene in genere più tiri liberi se "attacca il ferro" con frequenza, ovvero se effettua molti tentativi di schiacciata o di tiro vicino al canestro. Morey incoraggiò i suoi giocatori a tentativi di tiro più vicino al canestro e alla ricerca di falli [Catalfano, 2015].

Il Moreyball diede i suoi frutti e la franchigia nei campionati seguenti conquistò i playoff con regolarità, arrivando al suo apice nella stagione 2017/18 quando la squadra ottenne il primo posto nel campionato per poi essere sconfitti dai Golden State Warriors ad un passo dalle NBA Finals.

Proprio i Golden State Warriors, guidati da Steve Kerr, hanno seguito la filosofia di Daryl Morey, costruendo quella che ad oggi è considerata da alcuni esperti la squadra più forte della storia dell'NBA. Lo staff della franchigia riuscì a combinare il talento dei suoi giocatori con l'uso dei big data.

I Warriors adottarono lo slogan "Strength in numbers" per sottolineare l'importanza dei numeri nel loro stile di gioco. Oltre a far loro i concetti del Moreyball, adottarono il cosiddetto "Small-ball", uno stile di gioco in cui viene eliminata la figura del pivot o centro, il giocatore solitamente più alto della squadra che ha soprattutto il compito di proteggere il canestro dai tentativi ravvicinati. Il pivot viene sostituito da giocatori più versatili e rapidi, in grado di difendere anche i giocatori avversari lontano dal canestro.

Grazie a questo approccio, i Golden State Warriors tra la stagione 2014/15 e la stagione 2021/22 hanno raggiunto per 6 volte le NBA Finals, vincendo il campionato per ben 4 volte.

Con gli anni, tutte le franchigie NBA hanno assunto data analysts e ciò ha contribuito a rivoluzionare le strategie di gioco. Per dare un'idea dell'impatto che la statistica ha avuto nelle scelte delle squadre è sufficiente analizzare le quantità di tiri da tre punti che vengono tentati mediamente in una partita NBA:

durante i primi anni 2000, ogni squadra tentava circa 15 tiri a partita da tre punti; col passare degli anni, questi tentativi sono aumentati ad un ritmo sempre più alto. Infatti, nel 2010/11 i tentativi erano circa 20, nel 2014/15 circa 22 e oggi il numero di tentativi è esploso numericamente, arrivando ai 35,1 tiri da tre tentati mediamente a partita della stagione 2023/2024. Nonostante l'aumento di tentativi anche le percentuali di realizzazione sono leggermente aumentate, passando dal 33% del 1998/1999 al 36% della stagione passata, dimostrando come l'aumento delle conclusioni non ha ridotto l'efficienza di questo tiro.

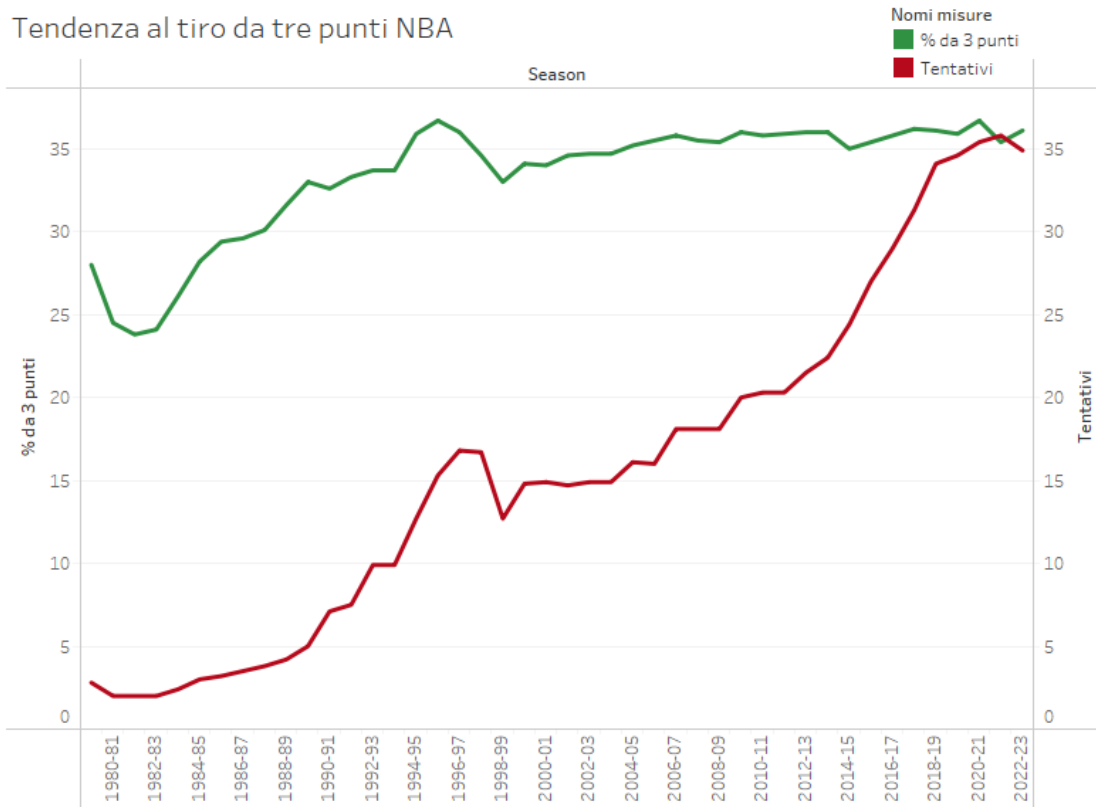


Figura 2. L'evoluzione del tiro da tre punti dalla stagione 1980/81 alla stagione 2022/23. Campionato NBA.

#### 1.4.1 Principali statistiche nella misura delle performance

La disponibilità sempre più vasta di dati e la costante ricerca nell'individuare dei parametri che riflettessero l'effettiva bontà delle prestazioni dei giocatori o delle squadre durante le partite, ha permesso lo sviluppo di una quantità ingente di tecniche e statistiche avanzate con cui le squadre possono analizzare le performance dei giocatori in campo. Molte di queste statistiche e il modo in cui esse vengono raccolte sono trattate nel libro *"Basketball data science: With applications in R"*, scritto dalle professoresse Zuccolotto e Manisera dell'Università di Brescia.

Tra i più diffusi al giorno d'oggi possiamo citare le statistiche di offensive e defensive ratings. Sono indicatori che misurano l'efficienza offensiva o difensiva delle squadre durante le partite o in porzioni di esse.

L'offensive rating risulta essere un indicatore più importante rispetto a quello dei punti segnati da una squadra, poiché calcola un nuovo punteggio basandosi su un numero di

possessi prefissato. In questo modo si può confrontare in modo più efficace la capacità delle squadre di segnare punti in partite diverse.

Per prima cosa è necessario definire il concetto di possesso, ovvero le azioni di attacco che una squadra compie durante una partita. Il numero di possesi che una squadra gioca può essere calcolato manualmente oppure più comunemente viene stimato attraverso la seguente formula:

$$Possessi = tiri\ tentati + 0.44 * tiri\ liberi + palle\ perse - rimbalzi\ offensivi$$

A questo punto, si può calcolare il valore dell'offensive rating in modo immediato:

$$ORtg = \frac{PuntiSegnati}{Possessi} * 100$$

Maggiore è il valore prodotto, maggiore è la capacità di una squadra di segnare punti durante un possesso. Dal semplice rapporto tra punti segnati e possesi si può individuare anche la stima dei punti segnati per possesso.

La stessa procedura viene effettuata per identificare la stima del defensive rating (DRtg), ovviamente utilizzando i dati offensivi prodotti dalla squadra avversaria.

Un'altra statistica che si può ricavare da questi dati è il net rating:

$$Net\ Rtg = ORtg - DRtg$$

Con questi dati, gli analisti possono ottenere informazioni sulla qualità del gioco e se il punteggio della partita sia stato influenzato da un basso o un alto ritmo.

A questo proposito, un altro indicatore spesso utilizzato per comprendere la fluidità e la velocità con cui una squadra gioca è il *pace*; può essere interpretato come il numero di possesi che, in media, una squadra gioca in una partita.

A livello individuale, uno strumento molto diffuso per valutare le prestazioni (oltre alle classiche rilevazioni su punti, assist, rimbalzi e palle rubate) è il plus/minus, che calcola semplicemente la differenza tra punti segnati e subiti da una squadra mentre è in campo uno specifico giocatore. Questa analisi a volte può risultare fine a sé stessa; perciò, negli anni ha preso sempre più piede il concetto di “Player efficiency rating” e di “Usage rate”.

Il player efficiency rating o PER è una statistica avanzata sviluppata da John Hollinger e molto utilizzata soprattutto nella NBA. Ha lo scopo di misurare il livello della prestazione di un giocatore tenendo in considerazione una serie di dati.

Il PER è un parametro molto interessante poiché presenta dei valori standardizzati intorno ad un valore fissato pari a 15; perciò i giocatori con un PER superiore a 15 sono superiori alla media dei giocatori NBA, mentre chi ha un PER inferiore a 15 sarà inferiore alla media. Questo dato tiene conto sia del ritmo di gioco della squadra, sia dei minuti in cui il giocatore si trova in campo.

Il calcolo del PER è abbastanza complesso, ma si può approssimare con la seguente formula:

$$PER = \frac{SommaStats}{MinutiGiocati} * \frac{PaceMedio}{PaceSquadra}$$

In cui per *SommaStats* si intende la somma delle statistiche positive di un giocatore (punti, assist, rimbalzi, ecc.) a cui si sottraggono le statistiche negative (palle perse, tiri sbagliati, ecc.). Il PER risulta un ottimo indicatore per confrontare cestisti di ruoli e squadre diverse e per analizzare il loro impatto nella squadra.

Lo usage rate, invece, è una statistica che misura quanto un giocatore è coinvolto nelle azioni offensive di una squadra e viene misurato come la percentuale di possesi offensivi che un giocatore utilizza durante la sua permanenza in campo. Si calcola rapportando la quantità di tiri (liberi o dal campo) e palle perse che il giocatore sotto indagine effettua con il totale di tiri e palle perse compiute dalla squadra nel periodo in cui si trova in campo; maggiore è lo usage rate, maggiore è l'importanza del giocatore nelle strategie offensive della squadra.

Altre analisi hanno il vantaggio di essere chiaramente intuibili graficamente, come nel caso delle shotchart. Sono delle mappe di tiro in cui vengono visualizzati i tentativi di tiro dei giocatori con il relativo risultato, in questo modo si possono conoscere le tendenze di tiro dei giocatori o delle squadre e la loro efficacia nelle varie zone del campo, ricavando anche il loro stile di gioco.

Alcuni modelli includono anche statistiche avanzate nelle mappe, visualizzando anche la percentuale reale di tiro, i tiri previsti o attesi oppure la qualità dei tiri in base alla difesa avversaria [Zuccolotto, 2020].

## 1.5 La basket analytics nella LBA

Anche se con un po' di ritardo rispetto agli USA, anche in Europa la big data analysis ha influenzato strategie e stili di gioco e, ad oggi, praticamente tutte le squadre di Eurolega hanno assunto data analysts con l'obiettivo di ottimizzare le performance attese. In Italia, le due squadre più blasonate, Olimpia Milano e Virtus Bologna, sono le società che più investono in questa tecnologia per migliorare le prestazioni sul campo e restare al passo con le migliori squadre europee. Le altre società del campionato italiano, pur avendo budget decisamente inferiori, cercano di stare al passo con le squadre di Bologna e Milano nell'utilizzo di statistiche avanzate e big data, spesso affidandosi ad aziende specializzate come TeamKeys, che elaborano statistiche e report appositi per le squadre interessate.

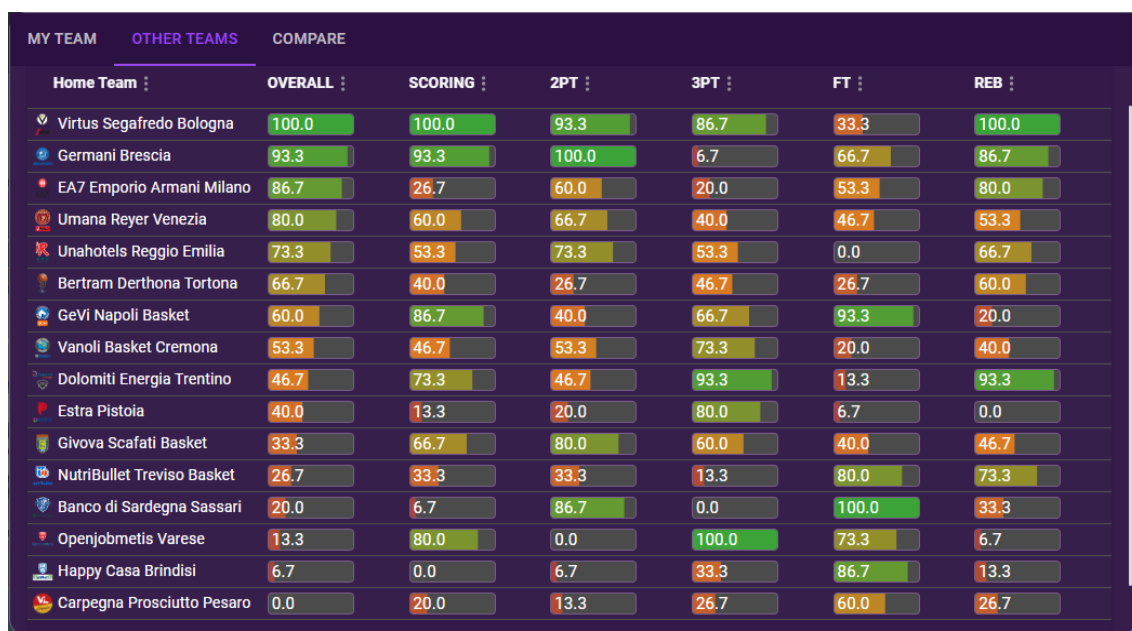


Figura 3. Un esempio di statistiche pubblicate da [TeamKeys](#).

In questo grafico riportato nel sito web di TeamKeys relativo alla stagione 2023/2024, sono mostrati una serie di indicatori con dei relativi punteggi. Questi valori indicano la percentuale di squadre che sono meno efficienti in una determinata caratteristica rispetto alla squadra indicata. In questo modo si può avere una panoramica di quelli che sono i punti di forza e le debolezze di ogni squadra comparati con le avversarie del campionato.





## 2 Il tirocinio con la Germani Brescia

### 2.1 BDsports

BDsports è un'iniziativa sviluppata dal "Big&Open Data Innovation Laboratory (BODaI-Lab)", dell'Università di Brescia. Questo progetto di ricerca è ideato per creare una collaborazione tra esperti dell'analisi sportiva sia dal punto di vista scientifico che pratico, in modo da creare una rete in grado di facilitare i contatti e le iniziative di ricerca congiunte.

BDsports è impegnata su quattro campi: ricerca scientifica, applicazione, didattica e divulgazione. Questi canali non sono indipendenti, ma entrano spesso in contatto fra di loro. Con questa iniziativa, il progetto si impegna a migliorare il contatto tra i diversi mondi interessati al tema della sport analytics, con una particolare attenzione ai giovani che, tramite progetti e stages, possono avvicinarsi a questo mondo che offre ormai sempre più sbocchi lavorativi [<https://bodai.unibs.it/bdsports/>].

Grazie a questa iniziativa dell'Università di Brescia, ho avuto la possibilità di entrare in stretto contatto con il mondo della sport analytics; in particolare, da gennaio a maggio del 2024, tramite un tirocinio presso questo laboratorio, ho potuto collaborare con la squadra di pallacanestro Germani Brescia, militante nella Lega Serie A di basket.

### 2.2 L'esperienza di tirocinio

Durante questa esperienza, svolta insieme ad un altro studente tirocinante, il nostro compito è stato quello di creare un report con analisi e statistiche a doc per la squadra che lo staff tecnico avrebbe poi sfruttato per preparare al meglio la partita successiva. Questi report venivano aggiornati settimanalmente, in modo tale da renderli disponibili alla squadra all'inizio di ogni settimana così da avere tempo per analizzarli in vista della partita in programma solitamente nel weekend. Essi comprendevano una serie di statistiche avanzate come net ratings, usage rate, shotcharts e altre statistiche offensive e difensive riguardanti la squadra e la relativa avversaria da affrontare.

Il mio lavoro ha riguardato nello specifico la creazione di shotcharts che rappresentavano graficamente le zone del campo in cui le squadre tendevano a tirare con più frequenza, comparando i dati con quelli dell'intera lega, con la possibilità di visualizzare le conclusioni filtrandole per giocatore o, nel caso dei playoff, per partita giocata.

## 2.3 L'estrazione dei dati

Per poter effettuare una qualsiasi analisi era necessario ricavare i dati del sito web della LBA. In particolare, era necessario raccogliere le informazioni derivanti dai “play by play” di tutte le partite; ovvero dei resoconti in cui vengono registrati in ordine cronologico tutti gli eventi rilevanti durante l'incontro. Nel play by play di ogni incontro, quindi, vengono segnalate tutte le conclusioni, gli assist, le palle perse e recuperate, i rimbalzi, le sostituzioni e altri dettagli; a ognuno di essi è associato il momento della partita in cui si è svolta l'azione ed il punteggio. L'accesso a questi dati avrebbe permesso di ricavare una serie di statistiche rilevanti per questo lavoro.

TABELLINI   RECAP   PHOTO GALLERY   BEST   **PLAY BY PLAY**   SHOT CHART   TEAM LEAD   MVP

GERMANI BRESCIA				ESTRA PISTOIA			
		35	2°Q 07:42	24	Fallo Subito	Varnado Jordon	23
		35	2°Q 07:39	26	2 punti segnato Giro e Tiro	Varnado Jordon	23
25	Cournooh David Reginald	Palla persa Passaggio sbagliato	35	2°Q 07:23	26		
			35	2°Q 07:23	26	Palla recuperata	Della Rosa Gianluca
			35	2°Q 07:16	26	3 punti sbagliato Tiro in sospensione	Varnado Jordon
4	Burnell Jason	Rimbalzo difensivo	35	2°Q 07:16	26		
5	Massinburg C.J.	2 punti segnato Tiro in corsa	37	2°Q 06:59	26		

Figura 5. Esempio di dati pubblicati nel play by play del sito web della [Lega Serie A](#).

Inizialmente, l'estrazione dei dati è stata effettuata utilizzando uno scraper in Python creato precedentemente da Francesco Olivo, data analyst per la Virtus Bologna, per la stagione 2021/2022 e riadattato per la stagione corrente. In seguito, a causa di alcune modifiche al sito web della lega, c'è stata la necessità di modificare questo codice, poiché presentava alcuni errori dovuti agli aggiornamenti del sito.

Il codice permetteva di raccogliere in un unico file CSV decine di migliaia di eventi descritti dai play by play della lega; in questo file, per ogni riga, oltre alla descrizione dell'evento sono riportati i seguenti dati:

- Il codice ID dell'incontro;
- Il punteggio della partita;
- Il tempo di gioco rimanente e trascorso, sia sul totale della partita che sullo specifico quarto;

- La lista dei dieci giocatori in campo;
- Il nome del giocatore che compie l'evento descritto e, in alcuni casi, chi lo subisce;
- Le coordinate in cui si svolge l'evento.

Il file ottenuto costituiva, perciò, il punto di partenza per effettuare ogni analisi specifica.

## 2.4 L'analisi dei dati raccolti

In questa fase l'obiettivo era di effettuare una serie di analisi che sarebbero state integrate graficamente con delle mappe appositamente create; ciò è stato fatto tramite il software R, un linguaggio di programmazione molto utilizzato da statistici e analisti.

Lo scopo era quello di ottenere tre tabelle distinte, una riguardante tutte le squadre del campionato, una nello specifico la Germani Brescia e l'altra il suo avversario della partita successiva. In queste tabelle vi erano tutti i dati relativi alle conclusioni tentate dai giocatori della relativa squadra.

Il codice R realizzato importava i dati raccolti dal sito web della lega e li filtrava, individuando tutte le conclusioni da 2 o 3 punti di una specifica squadra. Dopodiché, le coordinate relative alla posizione di tiro sono state convertite: inizialmente, i dati ottenuti dalla lega rappresentavano gli eventi in un campo di coordinate x, y con un range da 0 a 100 in entrambe le direzioni. L'esigenza era quella di ottenere i dati posizionati solo in un'unica metà del campo e coerenti con le reali misure del campo per semplificare l'integrazione grafica dei risultati. Perciò, i dati sono stati riconvertiti secondo le misure reali di una metacampo di basket in centimetri (1500x1400).

Successivamente, le tre tabelle sono state semplificate eliminando le variabili non rilevanti per il lavoro da svolgere:

- `game_id`: indicava il codice della partita, poi modificato durante la fase finale con il nome della partita dei playoff;
- `description`: una breve descrizione del tipo di tiro;
- `player`: il giocatore che ha tentato la conclusione;
- `assist`: in caso di realizzazione, viene indicato il giocatore che ha passato la palla a chi ha tirato;
- `result`: è una variabile dicotomica che assegna "missed" o "made" a seconda del risultato del tiro;

- `periodTime` e `totalTime`: indicano i relativi secondi trascorsi dall'inizio del quarto o dall'inizio della partita;
- `xx` e `yy`: sono le coordinate da cui è partito il tiro.

	game_id	description	assist	player	result	periodTime	totalTime	xx	yy
1298	Brescia Pistoia game 1	3 punti segnato	Semaj Christon	Jason Burnell	made	416	1016	1215	812
1305	Brescia Pistoia game 1	2 punti sbagliato		Nicola Akele	missed	452	1052	885	336
1309	Brescia Pistoia game 1	3 punti sbagliato		Amedeo Della Valle	missed	472	1072	285	728
1318	Brescia Pistoia game 1	2 punti sbagliato		Miro Bilan	missed	533	1133	795	168
1322	Brescia Pistoia game 1	3 punti segnato	Nicola Akele	C.J. Massinburg	made	553	1153	180	756
1327	Brescia Pistoia game 1	2 punti segnato		Jason Burnell	made	598	1198	780	224
1335	Brescia Pistoia game 1	3 punti sbagliato		Amedeo Della Valle	missed	28	1228	540	952
1339	Brescia Pistoia game 1	2 punti sbagliato		Semaj Christon	missed	62	1262	735	364
1343	Brescia Pistoia game 1	3 punti segnato	Amedeo Della Valle	Miro Bilan	made	85	1285	555	812
1346	Brescia Pistoia game 1	3 punti sbagliato		Kenny Gabriel	missed	113	1313	270	812
1354	Brescia Pistoia game 1	2 punti segnato	Amedeo Della Valle	Miro Bilan	made	134	1334	750	168
1359	Brescia Pistoia game 1	2 punti sbagliato		Kenny Gabriel	missed	185	1385	630	168
1372	Brescia Pistoia game 1	3 punti sbagliato		Kenny Gabriel	missed	215	1415	1155	840
1378	Brescia Pistoia game 1	3 punti sbagliato		Nicola Akele	missed	252	1452	1470	224

Showing 29 to 43 of 308 entries, 9 total columns

Figura 6. Alcuni dati relativi alle conclusioni della Germani Brescia nei playoff.

In seguito, dopo aver creato una mappa con dei poligoni opportunamente disegnati per dividere il campo in zone definite, queste osservazioni sarebbero state visualizzate nella mappa per avere una chiara idea delle tendenze di tiro di squadre e giocatori.

## 2.5 La costruzione dei poligoni

Questa parte del lavoro consisteva nel ricreare una mappa di un campo da basket che in seguito sarebbe stata utilizzata in due tipi di rappresentazioni:

1. Una semplice shotchart in cui sono indicati tutti i tiri tentati (da filtrare per squadra, giocatori, partite, ecc.)
2. Una mappa in cui le conclusioni vengono raggruppate in varie zone in modo da poter calcolare statistiche e fare confronti tra squadre diverse.

Per effettuare una scelta riguardo al numero di poligoni e a come posizionarli nel campo, sono state analizzate alcune shotcharts esistenti cercando di capire quali fossero le caratteristiche generali da rispettare. Sul sito web della LBA, le shotcharts risultavano un po' confusionarie, con zone di tiro di forme e grandezze molto diverse tra loro, che potrebbero dare indicazioni fuorvianti.

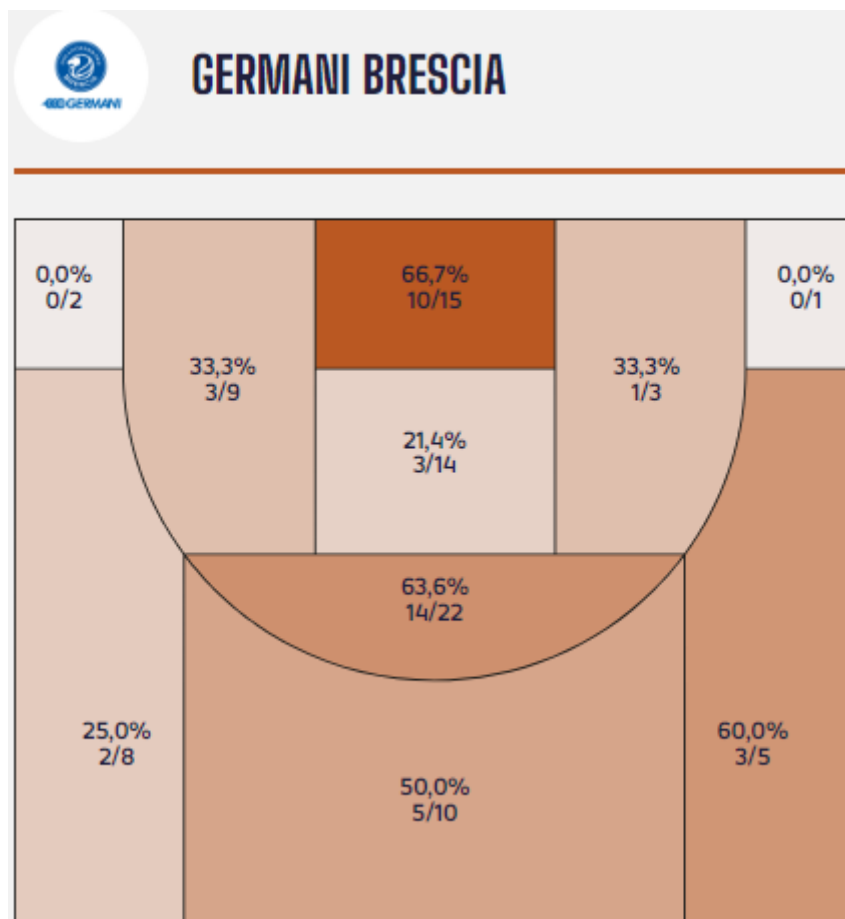
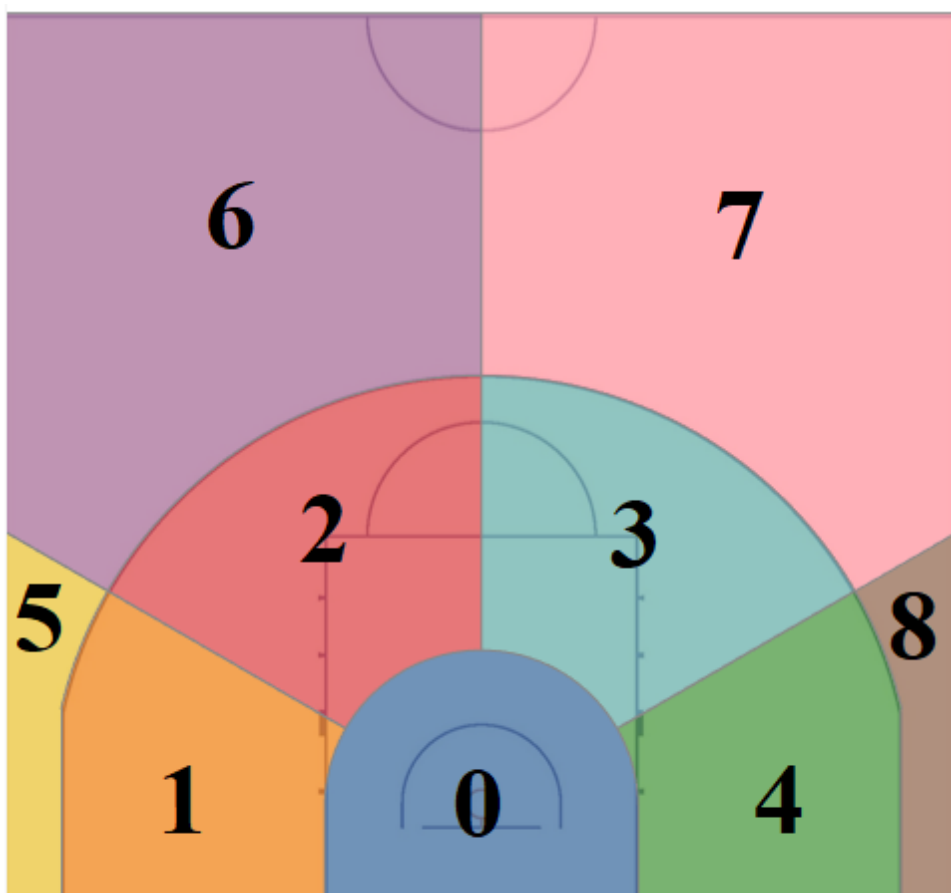


Figura 7. Shotchart relativa ad una partita della Germani Brescia dal sito della [Lega Serie A](#).

Si è deciso dunque di creare dei poligoni più uniformi per creare una shotchart che si discostasse da quella già disponibile sul sito. L'idea era di realizzare dei poligoni che considerassero sostanzialmente i tre tipi in cui si dividono le conclusioni: tiro da tre punti, tiro dalla media distanza e tiro sotto canestro; dividendo anche le conclusioni centrali al canestro da quelle laterali. Il campo doveva essere quindi diviso in nove poligoni, numerati da 0 ad 8:

- 0: la zona sotto il canestro;
- Da 1 a 4: le zone comprese tra l'area sotto il canestro e l'arco da tre punti; le zone interne sono delimitate da tre linee di 30, 90 e 150 gradi che partono dal centro del canestro;
- Da 5 a 8: le zone al di fuori della linea da tre punti; anch'esse sono divise tra loro dalle tre linee partenti dal canestro.



*Figura 8. Rappresentazione della mappa realizzata.*

Ogni area è stata creata calcolando la posizione dei vertici dei poligoni e l'ampiezza e la lunghezza degli archi formati dalle linee del campo; dopodiché, tramite un codice Python, sono stati generati uno per volta tutti i settori congiungendo i punti del piano che costituivano il perimetro di ognuno di essi. Infine, le coordinate create sono state esportate in un file CSV. Questo file conteneva dunque una lista di coordinate a cui era associato il numero del poligono corrispondente.

## 2.6 L'integrazione tra i poligoni e i dati raccolti

A questo punto, bisognava trovare un modo per associare i dati sui tiri con i settori realizzati. Perciò, sempre lavorando con Python, è stato realizzato un codice che importasse due file CSV: quello riguardante le coordinate dei perimetri dei nove poligoni e una delle tabelle realizzate in precedenza riguardo le conclusioni tentate, trattate nel [paragrafo 2.4](#).

Dato in input il file CSV delle conclusioni, l'algoritmo si occupava di associare ad ogni osservazione nel file il settore in cui il tiro è stato effettuato.

Assegnato ad ogni tiro il suo settore, l'algoritmo aggiornava le tre tabelle create in precedenza con la nuova variabile "sector\_id". Dopo aver realizzato l'algoritmo, bastava inserire, al termine di ogni giornata, le tre tabelle aggiornate in formato CSV, una alla volta, per poter associare ad ogni tiro la sua area specifica.

Dopodiché, con l'uso di R è stato realizzato un dataframe contenente alcune informazioni ricavate dalle tabelle precedenti. Per ognuna di esse sono stati calcolati, in ogni settore:

- Il numero di tentativi di tiro;
- Il numero di canestri segnati;
- La percentuale di tiro, si ottiene dal rapporto tra canestri segnati e tentati;
- La tendenza di tiro in quel settore, si ottiene dal rapporto tra tiri tentati nel settore e tiri totali tentati;
- Una media di quanti punti vengono realizzati per ogni tentativo in uno specifico settore, si ottiene moltiplicando il valore del canestro (2 o 3 punti) per la percentuale di tiro nel settore. Questo valore dà un'idea dell'efficienza del tiro in una specifica zona del campo.

Oltre a questi indicatori, venivano conteggiati anche il numero di tiri totali e, per quanto riguarda la Germani Brescia, la differenza tra i valori ottenuti dalla percentuale e dalla tendenza di tiro con i rispettivi valori calcolati su tutte le squadre del campionato. Queste due voci davano un'idea delle zone del campo in cui la squadra era più efficiente oppure quelle in cui il volume di tentativi era più elevato.

## 2.7 La rappresentazione grafica

In seguito, non restava altro che cercare di presentare nel modo più semplice ed efficace i dati raccolti, per rendere le analisi effettuate il più comprensibili possibile per lo staff tecnico. La rappresentazione grafica delle analisi è stata svolta tramite Tableau, un software di data visualization molto popolare nell'analisi di dati, che consente di visualizzare dati attraverso tabelle, mappe e dashboard interattive, facilitando la comprensione delle statistiche realizzate. Le figure 2 e 4 sono esempi di dashboard che sono state realizzate con questo software.

Tramite questo software è stato possibile, inserendo l'immagine utilizzata per realizzare i poligoni, ricreare il campo da gioco diviso per settori. Dopodiché, associando il file dei poligoni con l'ultimo file realizzato in R era possibile realizzare una serie di dashboard



inserendo i dati da voler visualizzare, anche in modalità interattiva. Durante il tirocinio, seguendo le necessità dello staff tecnico della società, sono state realizzate otto dashboard, che sono state raggruppate in un report di cinque pagine, che venivano aggiornati nei giorni successivi ad ogni partita della squadra. Il modo in cui queste dashboard sono state costruite verrà poi descritto nello specifico al [paragrafo 3.3](#).

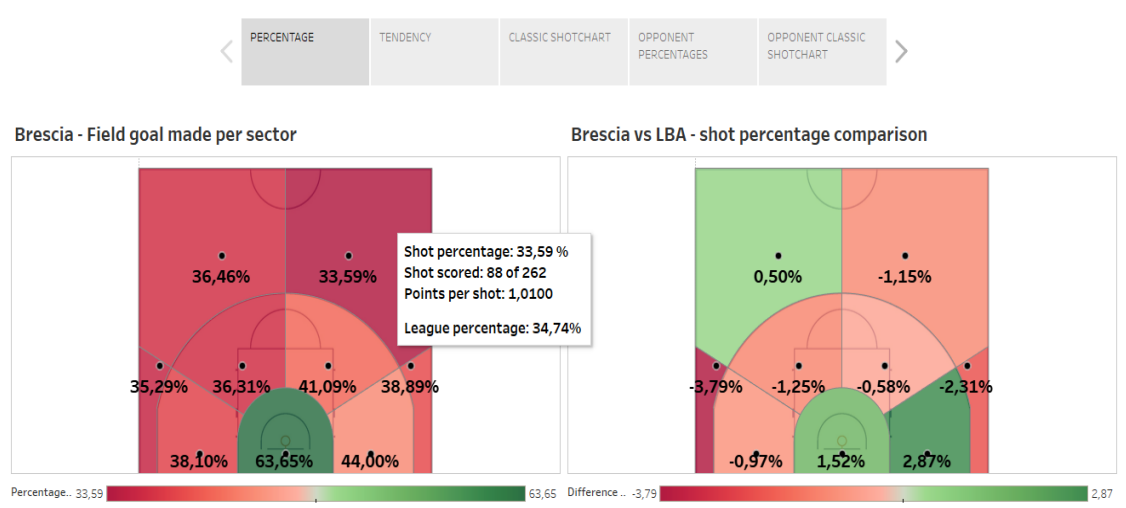


Figura 9. Dashboard relativa all'efficienza al tiro della Germani Brescia. Pubblicata il 6 maggio 2024.

La figura 9 mostra la prima pagina visualizzata nel report contenente due dashboard sviluppate e correlate tra loro. Nella prima, sono visualizzate le percentuali di tiro della Germani Brescia nei vari settori, ognuno dei quali è colorato secondo una scala che passa dal rosso al verde a seconda delle percentuali registrate. Grazie all'interattività di queste dashboard, posizionando il puntatore su uno specifico settore del campo si possono visualizzare altri dati: il volume di tiri segnati rispetto al totale, la media dei punti segnati con un tiro e la percentuale della lega nel settore. Nella figura accanto invece vengono confrontate le percentuali della squadra con quelle del resto del campionato, i settori in verde sono quelli in cui la squadra fa meglio degli avversari, in rosso quelli in cui fa peggio. La percentuale indicata in figura rappresenta la differenza tra la percentuale della Germani con quella dell'intera LBA. Posizionando il cursore sul settore interessato vengono specificate le due percentuali.

Nella figura 10 viene mostrata la seconda slide del report, che presenta un'analisi simile a quella precedente ma riguardante le tendenze di tiro della squadra. Per tendenza si intende il volume di conclusioni che la squadra tenta in un determinato settore rispetto al totale dei tiri. Come nella pagina precedente, vi sono due dashboard dove la prima rappresenta il volume di tiri rapportato al totale, con i settori colorati dal rosso al verde a

seconda delle percentuali; mentre nella seconda vengono paragonati i dati della Germani Brescia con quelli dell'intera lega.

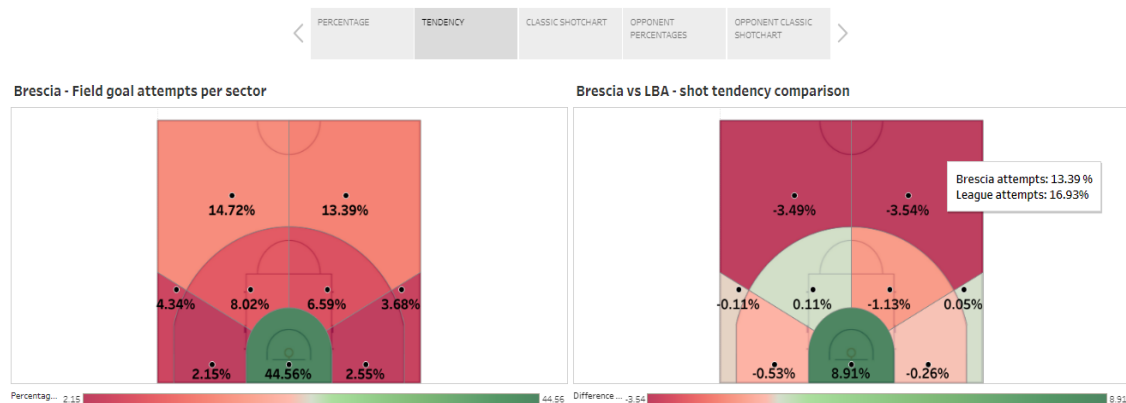


Figura 10. Dashboard relativa alle tendenze di tiro della Germani Brescia.

Queste due slide danno un'idea di quelli che sono i punti di forza della squadra dal punto di vista realizzativo nel corso della stagione; per esempio, l'efficienza sotto il canestro è stata la chiave dei successi della Germani Brescia, che ha chiuso la fase di campionato al secondo posto. Come si può notare, infatti, la squadra ha registrato percentuali notevoli nel settore (63,65%) nonostante un volume di tiri tentati decisamente alto; il 44,65% dei tiri, infatti, è stato tentato nella zona, quasi il 9% in più del resto della lega.

Le figure 11 e 12 sono, invece, delle semplici rappresentazioni del campo da basket in cui sono visualizzati tutte le conclusioni tentate dalla squadra; posizionando il cursore su una conclusione si può conoscere il giocatore che ha effettuato il tiro. Poiché visualizzando tutte le conclusioni della stagione il grafico diventa poco leggibile, è stata inserita la possibilità di filtrare i dati visualizzati in due modi:

- Per giocatore;
- Per partita.

In questo modo era possibile personalizzare la visualizzazione della dashboard a proprio piacimento. Nello stesso modo, è stato inserito un grafico interattivo analogo, ma relativo alla squadra avversaria della Germani nella partita successiva; un esempio di questa dashboard è visualizzato nella figura 13. Di seguito, quindi, sono mostrate alcune possibili visualizzazioni delle shotcharts.

### Brescia - Classic shotchart - Playoff

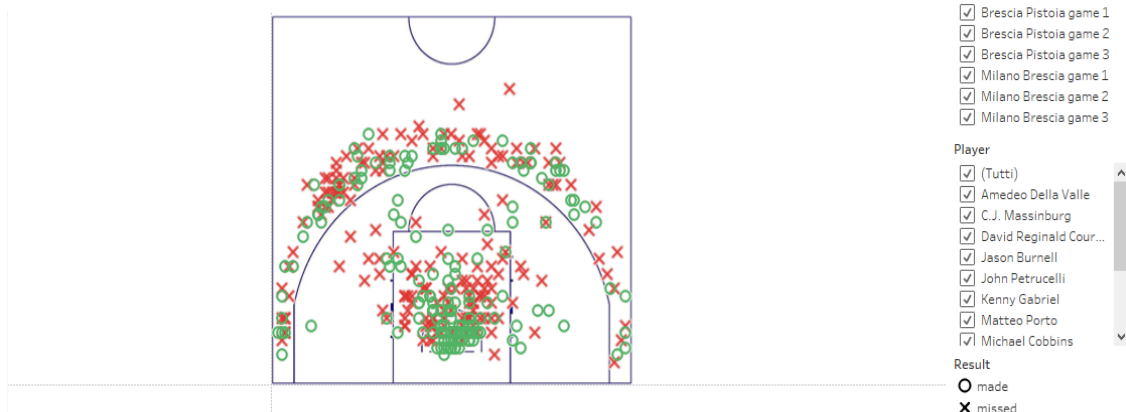


Figura 11. Tiri della Germani Brescia nei playoff.

### Brescia - Classic shotchart - Playoff

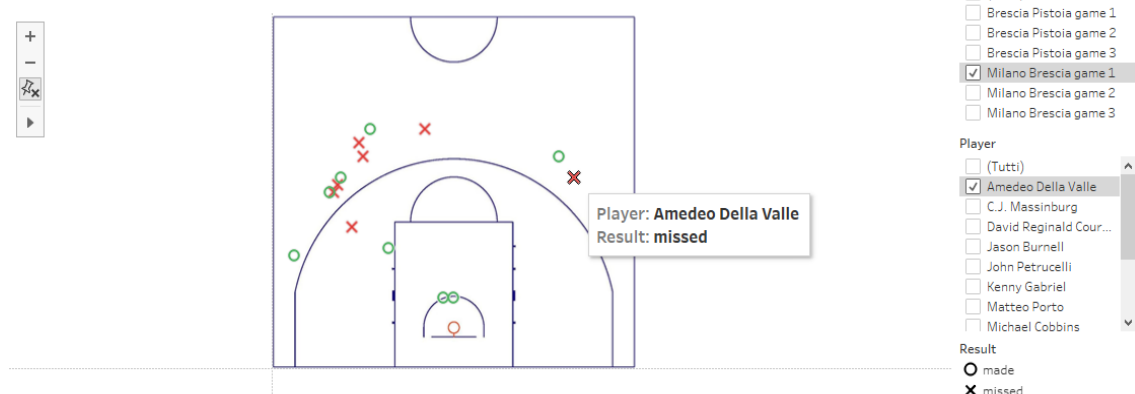


Figura 12. Tiri di Amedeo Della Valle durante la gara 1 dei playoff tra Milano e Brescia.

### Milano - Classic shotchart - Playoff

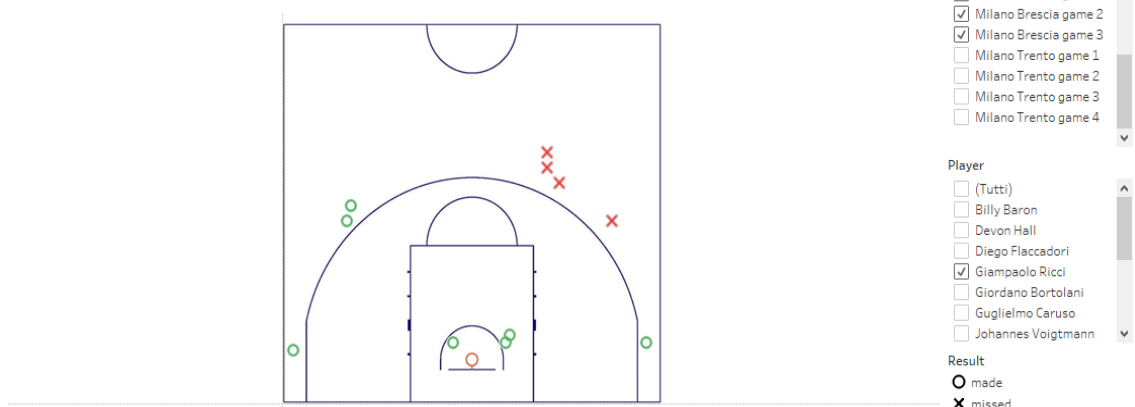


Figura 13. Tiri di Giampaolo Ricci durante la serie dei playoff tra Milano e Brescia (3 gare).

L'ultima slide è dedicata a due dashboard relative all'imminente avversaria della Germani. La prima tratta le percentuali di tiri segnati mentre la seconda le tendenze nei

vari settori. Nella figura 14 sono mostrate le statistiche relative ai playoff dell'Olimpia Milano, la squadra che ha vinto il campionato LBA nella scorsa stagione, battendo la Germani Brescia 3-0 in semifinale.

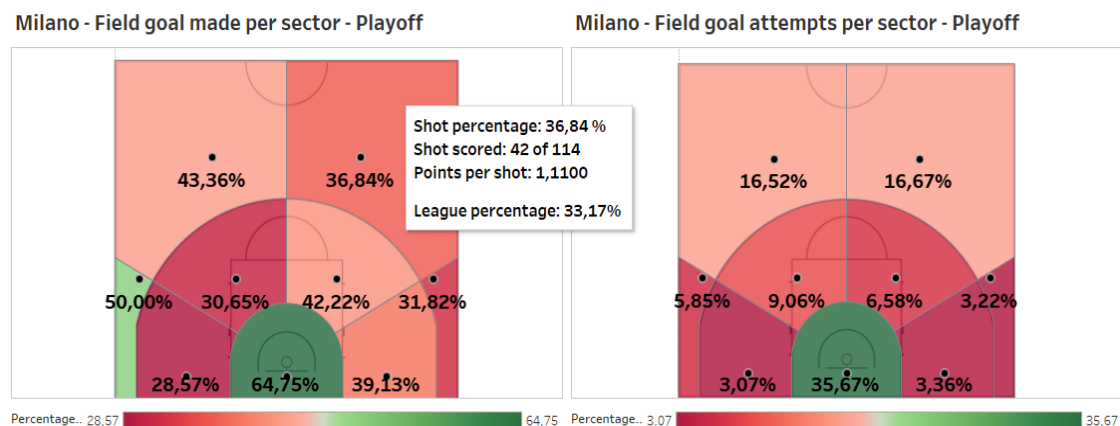


Figura 14. Statistiche Olimpia Milano playoff.

Confrontando questi dati con le medie delle altre squadre nei playoff, si nota come le percentuali nel tiro da tre punti e sotto canestro hanno fatto la differenza nel riuscire a vincere la competizione.

Durante il tirocinio, e quindi per tutta la seconda parte della stagione, queste analisi venivano pubblicate periodicamente su Tableau Public per fornire alla società un'infografica aggiornata, consentendo così allo staff tecnico di avere a disposizione un maggior numero di analisi per preparare le partite in modo più consapevole e strategico.

[<https://public.tableau.com/app/profile/francesco.riboni/vizzes>]

Nel capitolo successivo verranno approfonditi, dal punto di vista tecnico, i software e i codici realizzati per produrre i report statistici finali.

## 3 Tecniche utilizzate

In questo capitolo saranno trattate le tecniche statistiche ed informatiche che sono state impiegate nel lavoro svolto. In particolare, ci si soffermerà sulle tre piattaforme utilizzate nella produzione dei dati: Python, R e Tableau; definendo il modo in cui sono state sfruttate, analizzando gli algoritmi e le tecniche di analisi che hanno permesso la realizzazione delle dashboard.

### 3.1 Python

Python è un linguaggio di programmazione ad alto livello, realizzato dal programmatore olandese Guido Van Rossum nel 1991. Si tratta di un software completamente gratuito che, attualmente, è sviluppato e gestito dall'organizzazione no profit Python Software Foundation che favorisce una costante evoluzione del linguaggio [Van Rossum, 2007].

Grazie alla sua semplicità e flessibilità, Python è uno dei linguaggi di programmazione più utilizzati sia dagli studenti che si affacciano a questo mondo, sia da aziende e operatori più esperti, tanto da essere adottato perfino da alcune tra le multinazionali più importanti al mondo, come Google e IBM.

Questo linguaggio offre altri vantaggi che lo rendono molto versatile: può essere utilizzato su gran parte dei sistemi operativi più comunemente utilizzati (macOS, Windows, Linux) e supporta diversi paradigmi di programmazione come quella funzionale, quella dinamica e quella imperativa.

Inoltre, per Python sono disponibili una vasta gamma di librerie disponibili che facilitano la realizzazione di applicazioni più complesse.

Per questi motivi, questa piattaforma è utilizzata in ambiti come la data science e il machine learning, lo sviluppo web, la creazione di videogiochi, per la sicurezza informatica e in molti altri settori [<https://www.python.it/>].

Nel capitolo precedente è stato anticipato l'uso di questo linguaggio in tre fasi distinte del lavoro di tirocinio:

1. La scrittura dello scraper per estrarre i dati grezzi;
2. La realizzazione dei poligoni;
3. La creazione di un algoritmo che associa ad ogni tiro il suo settore corrispondente.

### 3.1.1 Scraper

Come anticipato nel capitolo precedente, inizialmente l'estrapolazione dei dati era facilitata dall'utilizzo di uno [scraper realizzato da Francesco Olivo](#) per la stagione 2021/22. Nel corso dell'anno, sono state apportate importanti modifiche al sito web della LBA e, di conseguenza, si è dovuti ricorrere a delle modifiche anche nel codice utilizzato.

È possibile visionare il codice modificato ed utilizzato per ricavare i dati iniziali a questo [link](#).

Lo scraper recupera le informazioni dal sito della lega basket. Il recupero delle informazioni viene effettuato tramite API. Le API (Application Programming Interface) di un sito web sono un insieme di regole e definizioni che permettono la creazione e l'integrazione di diverse applicazioni e software.

Per ogni partita giocata nel campionato, identificata dalla variabile `game_id`, vengono fatte due chiamate:

- Scores: serve per recuperare il box score della partita. La chiamata restituisce un file JSON che contiene le statistiche generali di tutti i giocatori e la lista dei giocatori titolari ad inizio partita. Questa informazione è utile quando bisogna gestire le sostituzioni e si vuole sapere quali giocatori sono in campo ad ogni azione del play by play.
- Play by play: serve per recuperare la lista delle azioni registrate di ogni partita. Il file JSON ottenuto è il punto di partenza per ottenere il play by play su cui sono state effettuate le analisi.

Lo scraper si occupa di unire le informazioni ottenute con queste due chiamate e compie delle procedure di miglioramento del dato. Tra queste vi sono anche delle operazioni di mapping nella descrizione dell'evento in modo da adattare le informazioni in italiano ad uno standard in lingua inglese che permette facilmente l'uso della libreria [BasketballAnalyzeR](#), sviluppata da Marco Sandri con le professoressse Zuccolotto e Manisera dell'Università di Brescia, per l'analisi dei dati della pallacanestro in R.

L'output di questo codice è un file CSV contenente la lista di tutte le azioni effettuate in ogni partita del campionato. Al termine della stagione regolare, il file conteneva circa 87000 osservazioni e 45 variabili. Oltre alle 9 variabili descritte nel [paragrafo 2.4](#), vi erano altri parametri che avrebbero potuto essere interessanti per altre analisi, come per esempio

la data dell'incontro, la lista dei giocatori in campo nel momento in cui avviene l'azione e i punteggi delle due squadre in quell'istante.

Lo stesso processo è stato applicato alle partite dei playoff ottenendo in questo caso un file di 10700 osservazioni e 45 variabili.

### 3.1.2 Creazione poligoni

Per creare i poligoni alla base delle shotchart è stato creato un file CSV contenente i punti dei nove poligoni. Ogni poligono è identificato da una lista di coordinate (X,Y), l'unione di queste coordinate genera il poligono.

Per generare correttamente ogni lista di punti è stato utilizzato il tool “CBI Studio by Interworks” [<https://cbistudio.interworks.com>], in cui è possibile definire dei poligoni come sequenze di punti. Dato che il tool non permetteva di generare dei poligoni curvilinei, si è deciso di produrre un elevato numero di coordinate lungo gli archi del campo in modo tale da riprodurre nel modo migliore possibile le aree del campo.

Questi punti sono stati generati tramite uno [script python](#) e importati nel tool per verificarne la correttezza.

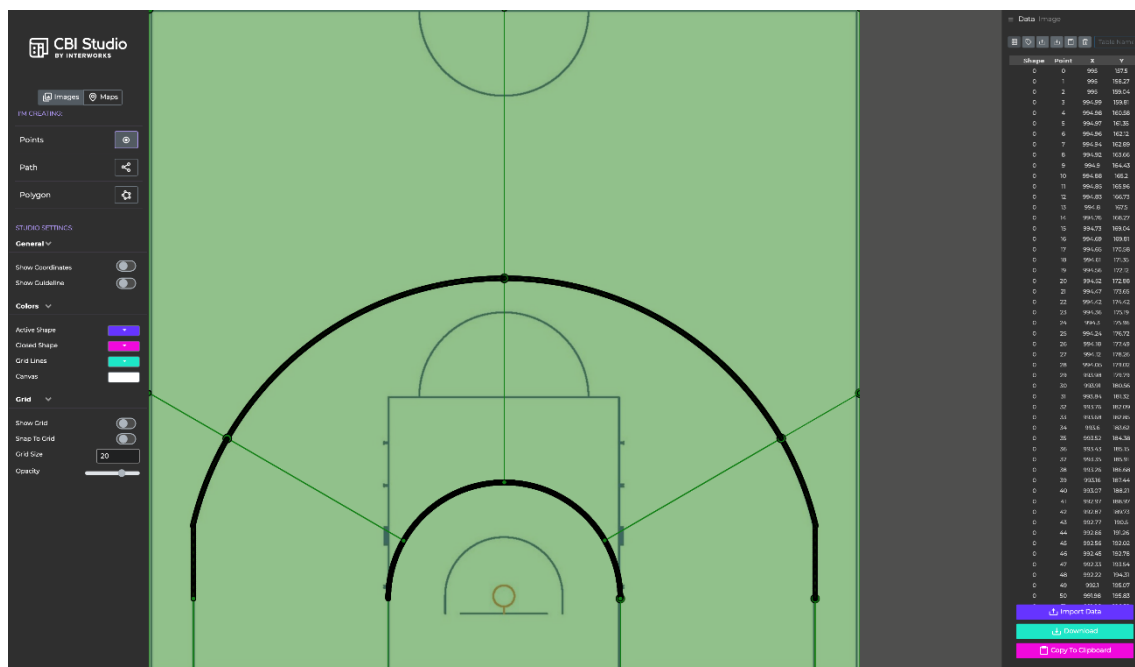


Figura 15. Visualizzazione dei punti dei poligoni con CBI Studio.

### 3.1.3 Associazione tra dati e settori

Questo script python ha il compito di associare alle coordinate (X,Y) delle conclusioni di tiro il settore corrispondente. Questo step viene effettuato sui tre dataset ottenuti in precedenza con R e descritti nel [paragrafo 2.4](#).

Questa operazione è stata effettuata tramite la libreria [shapely](#), la quale consente di creare degli oggetti di tipo Point e Polygon, e verificare se un punto si trova all'interno di un poligono attraverso le funzioni contains e touches della classe Polygon.

Lo script si limita sostanzialmente a ricreare un CSV identico a quello di partenza con l'aggiunta della colonna sector\_id che indica l'id del poligono in cui è stato effettuato il tiro. In seguito, sono mostrate le prime osservazioni di un file ottenuto utilizzando questo script:

	game_id	description	assist	player	result	periodTime	totalTime	xx	yy	sector_id
0	Milano Trento game 1	2 punti segnato	NaN	Shavon Shields	made	34	34	1020	644	3
1	Milano Trento game 1	2 punti sbagliato	NaN	Stefano Tonut	missed	102	102	825	196	0
2	Milano Trento game 1	2 punti sbagliato	NaN	Nicolò Melli	missed	102	102	825	196	0
3	Milano Trento game 1	3 punti segnato	Nicolò Melli	Shavon Shields	made	128	128	1020	896	7
4	Milano Trento game 1	3 punti segnato	NaN	Shabazz Napier	made	218	218	300	840	6

*Figura 16. Esempio di osservazioni a cui viene associato un settore.*

È possibile visualizzare il codice realizzato a questo [link](#).

## 3.2 R ed RStudio

R è un linguaggio di programmazione per l'elaborazione statistica e grafica, realizzato dagli statistici Robert Gentleman e Ross Ihaka. È utilizzato soprattutto per l'impiego di tecniche statistiche avanzate in molti settori come la finanza, l'econometria, la medicina e il marketing [Ihaka, 1996].

Come Python, R è una piattaforma gratuita e open source, in cui una comunità globale contribuisce al suo sviluppo e al suo continuo aggiornamento; inoltre, anch'esso è utilizzabile su diverse piattaforme.

R dispone di una vasta gamma di tecniche statistiche (regressioni lineari e non lineari, analisi delle componenti principali, test statistici e molto altro) e grafiche con cui produrre diagrammi e plot di qualità. Fornisce anche diverse interfacce grafiche, come RStudio, che consentono di integrare R con altri pacchetti [<https://www.r-project.org/about.html>].



L'utilizzo di R in questo lavoro si è limitato alla produzione di dataframe realizzati attraverso i dati grezzi ottenuti nelle prime fasi e al calcolo delle statistiche visualizzate successivamente sulle dashboard.

### 3.2.1 Script realizzati in R

In R i codici sono stati scritti con l'ausilio di RStudio e hanno subito spesso aggiornamenti e piccole modifiche a causa delle richieste fatte dallo staff tecnico durante la stagione, per questo motivo gli script che verranno trattati nel dettaglio saranno quelli utilizzati nel finale di stagione durante i playoff.

Il primo script aveva lo scopo di estrarre dal play by play ricavato dallo scraper i dati necessari alle analisi successive. Per prima cosa è stato importato il file CSV intitolandolo "PbP", mentre in una lista sono stati inserite tutte le descrizioni possibili delle azioni che riguardavano i tiri dei giocatori. Dopodiché, si è andati a creare tre dataset contenenti le informazioni riguardanti le conclusioni. Nella figura seguente sono mostrati i codici utilizzati per creare il file CSV "LBA\_shot", in cui sono contenute tutte le conclusioni tentate dai giocatori del campionato durante i playoff.

```
tiri <- c("2 punti segnato", "2 punti sbagliato", "3 punti segnato", "3 punti sbagliato")
# Creazione dataset LBA_shot
subdata <- subset(PbP, PbP$description %in% tiri)
# Trasformazione coordinate
subdata$xx <- ifelse(subdata$original_x > 50, (100-subdata$original_y)*15, subdata$original_y * 15)
subdata$yy <- ifelse(subdata$original_x > 50, (100-subdata$original_x)*28, subdata$original_x*28)
# Selezione delle variabili necessarie
LBA_shot <- subdata[, c("game_id", "description", "assist",
                        "player", "result", "periodTime", "totalTime", "xx", "yy")]
# Creazione file csv
percorso_csv <- "C:/Users/ribon/Desktop/Tirocinio BDsports/Notebooks/csv/LBA_shot.csv"
write.csv(LBA_shot, file = percorso_csv, row.names = FALSE)
```

Figura 17. Creazione del file "LBA\_shot".

Tramite il comando subset vengono estratti dalla tabella PbP solamente le righe riguardanti azioni di tiro; poi vengono trasformate le coordinate delle conclusioni, poiché inizialmente erano predisposte per un campo intero con coordinate da 0 a 100 in entrambe le dimensioni, mentre per poter utilizzare le coordinate sul campo da gioco creato in precedenza, queste dovevano essere calibrate su una sola metà del campo con misurazioni da 0 a 1500 sull'asse X e da 0 a 1400 sull'asse Y. Infine, nel dataset "LBA\_shot" sono state inserite le variabili che sarebbero servite in seguito ed è stato esportato il file in formato CSV. In questo modo, il file iniziale contenente tutte le azioni registrate nei playoff è stato ridotto da 10700 a 3400 righe e da 45 a 9 variabili. Lo stesso procedimento

è stato ripetuto nella creazione degli altri due dataset più specifici: “BRESCIA\_shot” e “AVVERSARIO\_shot”, in cui però sono state selezionate solo le righe in cui a compiere l’azione era una determinata squadra e, inoltre, è stata rinominata la variabile “game\_id”, inserendo il nome della partita anziché il codice associato a essa.

```
# creazione dataset BRESCIA_shot
subdata <- subset(PbP, PbP$description %in% tiri & team == "Germani Brescia")
# Trasformazione coordinate
subdata$xx <- ifelse(subdata$original_x > 50, (100-subdata$original_y)*15, subdata$original_y * 15)
subdata$yy <- ifelse(subdata$original_x > 50, (100-subdata$original_x)*28, subdata$original_x*28)
# Selezione delle variabili necessarie
BRESCIA_shot <- subdata[, c("game_id", "description", "assist",
                           "player", "result", "periodTime", "totalTime", "xx", "yy")]

# Rinomina delle partite giocate
BRESCIA_shot$game_id[BRESCIA_shot_map$game_id == "24430"] <- "Brescia Pistoia game 1"
BRESCIA_shot$game_id[BRESCIA_shot_map$game_id == "24434"] <- "Brescia Pistoia game 2"
BRESCIA_shot$game_id[BRESCIA_shot_map$game_id == "24437"] <- "Brescia Pistoia game 3"
BRESCIA_shot$game_id[BRESCIA_shot_map$game_id == "24444"] <- "Milano Brescia game 1"
BRESCIA_shot$game_id[BRESCIA_shot_map$game_id == "24445"] <- "Milano Brescia game 2"
BRESCIA_shot$game_id[BRESCIA_shot_map$game_id == "24446"] <- "Milano Brescia game 3"

# creazione file csv
percorso_csv <- "C:/Users/ribon/Desktop/Tirocinio BDsports/Notebooks/csv/BRESCIA_shot.csv"
write.csv(BRESCIA_shot, file = percorso_csv, row.names = FALSE)

# Creazione dataset AVVERSARIO_shot
subdata <- subset(PbP, PbP$description %in% tiri & PbP$team == "EA7 Emporio Armani Milano")
# Trasformazione coordinate
subdata$xx <- ifelse(subdata$original_x > 50, (100-subdata$original_y)*15, subdata$original_y * 15)
subdata$yy <- ifelse(subdata$original_x > 50, (100-subdata$original_x)*28, subdata$original_x*28)
# Selezione delle variabili necessarie
AVVERSARIO_shot <- subdata[, c("game_id", "description", "assist",
                              "player", "result", "periodTime", "totalTime", "xx", "yy")]

# Rinomina delle partite giocate
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24429"] <- "Milano Trento game 1"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24433"] <- "Milano Trento game 2"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24438"] <- "Milano Trento game 3"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24440"] <- "Milano Trento game 4"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24444"] <- "Milano Brescia game 1"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24445"] <- "Milano Brescia game 2"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24446"] <- "Milano Brescia game 3"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24451"] <- "Bologna Milano game 1"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24452"] <- "Bologna Milano game 2"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24453"] <- "Bologna Milano game 3"
AVVERSARIO_shot$game_id[AVVERSARIO_shot_map$game_id == "24454"] <- "Bologna Milano game 4"

# creazione file csv
percorso_csv <- "C:/Users/ribon/Desktop/Tirocinio BDsports/Notebooks/csv/AVVERSARIO_shot.csv"
write.csv(AVVERSARIO_shot, file = percorso_csv, row.names = FALSE)
```

Figura 18. Creazione dei file “BRESCIA\_shot” e “AVVERSARIO\_shot”.

Le osservazioni presenti nei tre dataset sono state associate ai relativi settori tramite lo script Python descritto in precedenza. Successivamente, con l’ausilio di R, è stata creata una tabella contenente una serie di dati e statistiche ricavate dai dati elaborati, che è stata poi integrata nelle dashboard di Tableau. Il codice importa in R i file CSV dei tre dataset delle conclusioni con il settore associato con i seguenti nomi: “Shot\_sector\_LBA”, “Shot\_sector\_BS” e “Shot\_sector\_AVV”; viene poi caricato il pacchetto “dplyr” per poter eseguire lo script riportato.

```
# Caricamento pacchetto dplyr
library(dplyr)

# Calcolo somma di made e missed per ogni settore
lba_shot <- shot_sector_LBA %>%
  group_by(sector_id) %>%
  summarise(total = n(),
            made = sum(result == "made"))
# calcolo di alcune statistiche
lba_shot$percentage <- round(lba_shot$made / lba_shot$total, 4)*100
lba_shot$tendency <- round(lba_shot$total / sum(lba_shot$total), 4) * 100
lba_shot$points_per_shot <- round(ifelse(lba_shot$sector_id >= 5,
                                         lba_shot$percentage*3/100, lba_shot$percentage*2/100), 2)
```

Figura 19. Codici R per ottenere alcune statistiche del campionato.

Il pacchetto caricato permette di raggruppare le osservazioni in base alla variabile `sector_id`, riuscendo così ad ottenere la somma di tiri segnati (`made`) e tiri tentati (`total`) dalla somma di `made` e `missed`. Dopo aver ottenuto questi dati, sono state ricavate le percentuali di tiri segnati, le tendenze di tiro nei diversi settori e la media di punti realizzati per tiro in ogni area. Questo procedimento è stato ripetuto per tutti i tre file. Le tabelle ottenute avevano la seguente forma:

sector_id	total	made	percentage	tendency	points_per_shot
0	1314	793	60.35	38.14	1.21
1	89	29	32.58	2.58	0.65
2	256	92	35.94	7.43	0.72
3	195	70	35.90	5.66	0.72
4	76	35	46.05	2.21	0.92
5	166	69	41.57	4.82	1.25
6	634	233	36.75	18.40	1.10
7	597	198	33.17	17.33	1.00
8	118	47	39.83	3.43	1.19

Figura 20. Dati riferiti alle partite dei playoff della stagione 2023/24 ottenuti in R.

Questi dati sono stati raggruppati in un unico dataframe, inserendo anche il numero totale di tiri e le differenze percentuali tra i dati della Germani e quelli della lega.

Il dataframe ottenuto è stato denominato “shotchart\_stats”, composto da 9 righe e 21 variabili, che saranno poi interpretate su Tableau nelle dashboard interattive. La tabella è stata perciò esportata in un file CSV.

```

# Tabella unica
shotchart_stats <- data.frame(
  sector_id = lba_shot$sector_id,
  BS_made = bs_shot$made,
  BS_attempts = bs_shot$total,
  BS_percentage = bs_shot$percentage,
  BS_tendency = bs_shot$tendency,
  BS_points_per_shot = bs_shot$points_per_shot,
  LBA_made = lba_shot$made,
  LBA_attempts = lba_shot$total,
  LBA_percentage = lba_shot$percentage,
  LBA_tendency = lba_shot$tendency,
  LBA_points_per_shot = lba_shot$points_per_shot,
  AVV_made = avv_shot$made,
  AVV_attempts = avv_shot$total,
  AVV_percentage = avv_shot$percentage,
  AVV_tendency = avv_shot$tendency,
  AVV_points_per_shot = avv_shot$points_per_shot)

# Totale conclusioni
shotchart_stats$BS_totalshots <- sum(shotchart_stats$BS_attempts)
shotchart_stats$LBA_totalshots <- sum(shotchart_stats$LBA_attempts)
shotchart_stats$AVV_totalshots <- sum(shotchart_stats$AVV_attempts)
# Differenze tra Brescia e LBA
shotchart_stats$BS_Diff_percentage <- round(shotchart_stats$BS_percentage - shotchart_stats$LBA_percentage, 2)
shotchart_stats$BS_Diff_tendency <- round(shotchart_stats$BS_tendency - shotchart_stats$LBA_tendency, 2)

```

*Figura 21. Calcolo delle statistiche utili per le shotchart.*

## 3.3 Tableau

Tableau Software è una società statunitense nata nel 2003, che si occupa di software di visualizzazione dei dati focalizzati sulla business intelligence.

Si tratta di una piattaforma di data visualization che trasforma il modo di utilizzare i dati per risolvere i problemi; questo strumento è molto popolare, perché è in grado di produrre grafici, tabelle e dashboard interattive, senza la necessità di alcuna competenza nella programmazione [Hanrahan, 2007].

Attraverso l'analisi visiva e intuitiva per tutti, Tableau ha trasformato la business intelligence, ovvero l'insieme dei processi e delle tecnologie che aiutano le imprese a trasformare i dati grezzi in informazioni utili per le decisioni aziendali, migliorando l'efficienza e la comunicazione [<https://www.tableau.com/it-it/why-tableau/what-is-tableau>].

### 3.3.1 La costruzione delle dashboard

Per poter visualizzare i dati ottenuti nel campo di gioco, sono stati importati i file utili per le analisi ed è stato effettuato un join tra le tabelle “polygons.csv” e “shotchart\_stats” mediante le rispettive variabili “Shape Id” e “Sector Id”. Questa opzione permette di collegare due o più tabelle correlate tra loro da una variabile in comune.

polygons.csv è composto da 2 tabelle. ⓘ

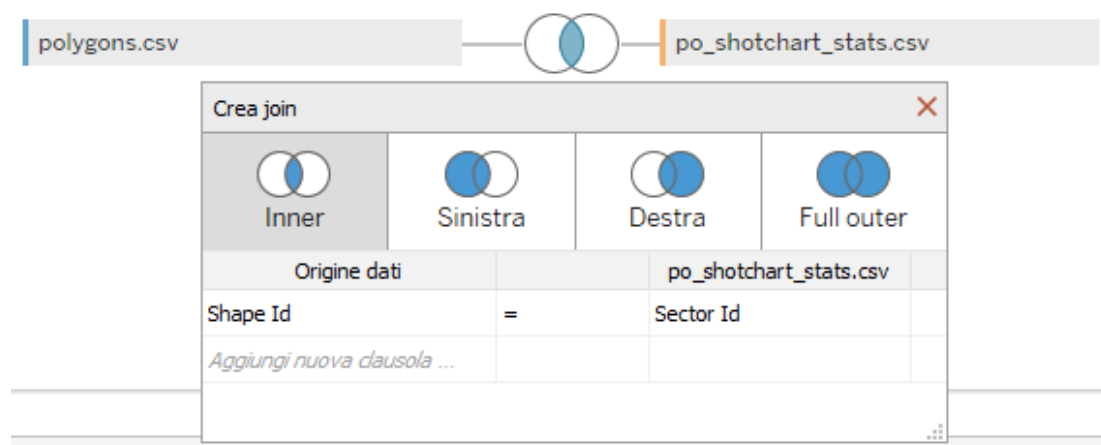


Figura 22. Join tra due tabelle su Tableau.

Successivamente, sono state modificate una serie di variabili considerate come stringhe da Tableau, ma che invece erano variabili numeriche e andavano considerate come tali. In seguito, sono stati creati i primi fogli di lavoro relativi alle dashboard che fornivano i dati in base ai settori creati: per prima cosa doveva essere importata la mappa del campo utilizzata nella creazione dei poligoni nel foglio e sono state inserite sulle barre delle colonne e delle righe le coordinate X e Y in modo da rappresentare il campo da gioco nel foglio in maniera corretta; inserendo poi le variabili Shape Id e Point Id tra gli indicatori sono stati individuati i 9 settori.

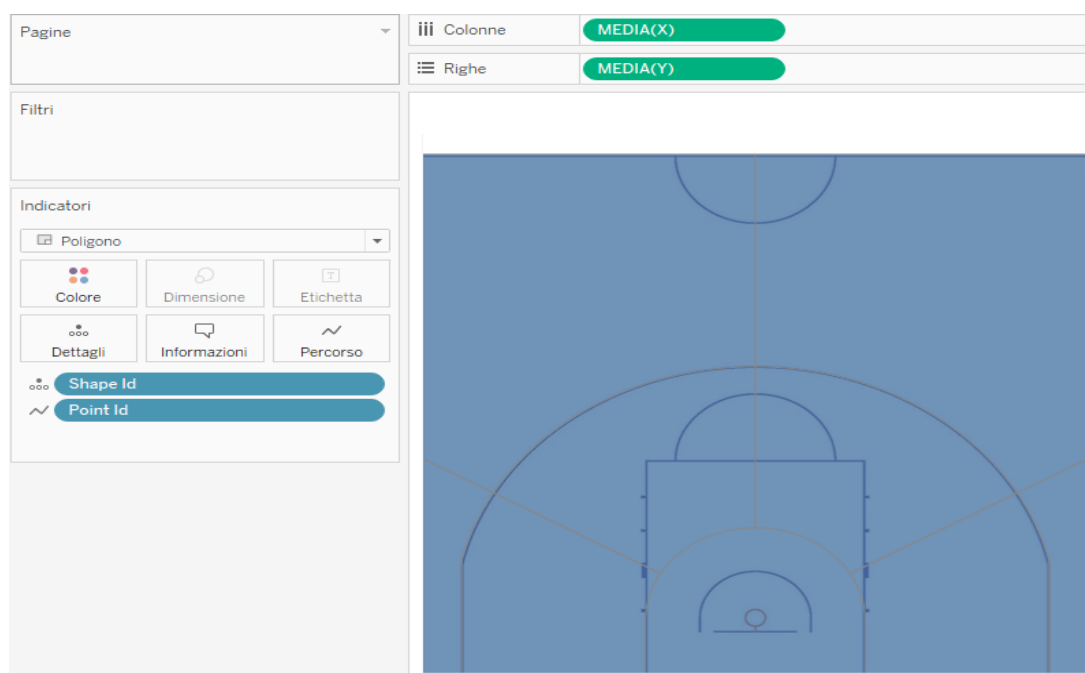


Figura 23. Visualizzazione del campo con i poligoni.

Per creare le dashboard da pubblicare era sufficiente inserire tra gli indicatori nella sezione colore la statistica da cui voler far dipendere le tonalità dei rettangoli. Per mettere in risalto le statistiche migliori e peggiori delle squadre, si è scelto di colorare i settori secondo una scala che andasse dal rosso acceso per le percentuali più basse al verde scuro per quelle più alte, secondo le percentuali registrate, sfumando verso un colore più chiaro per quelle intermedie. Sono stati poi aggiunte tra gli indicatori tutte le variabili di interesse, che sarebbero state mostrate nella dashboard.

Nella sezione informazioni era possibile inserire la descrizione di ciò che si sarebbe voluto visualizzare nei grafici interattivi una volta puntato il cursore su un settore di interesse.

Nella figura 24 sono mostrati gli indicatori e le informazioni utilizzati per descrivere ogni settore della prima dashboard del report.

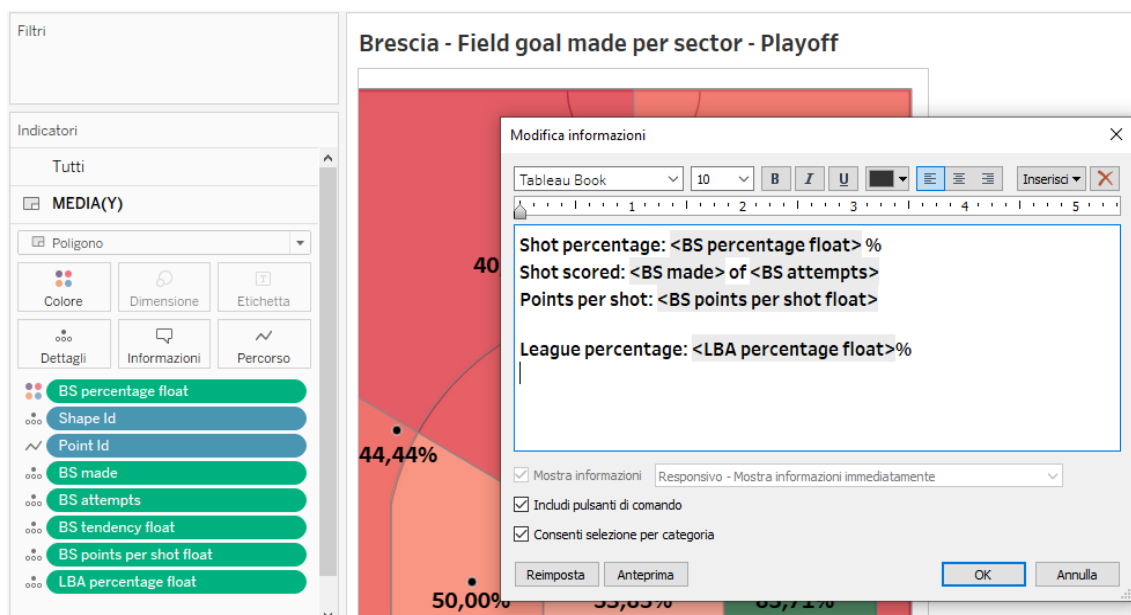


Figura 24. Scelta delle informazioni visualizzabili in ogni settore.

In questo modo sono state create tutte le dashboard che rappresentavano le conclusioni raggruppate per settori.

Per quanto riguarda le shotcharts che mostravano tutte le conclusioni delle squadre, il processo è stato diverso. È stato sufficiente associare la mappa del campo di gioco al file delle conclusioni della squadra, inserendo tra le variabili le coordinate dei tiri.

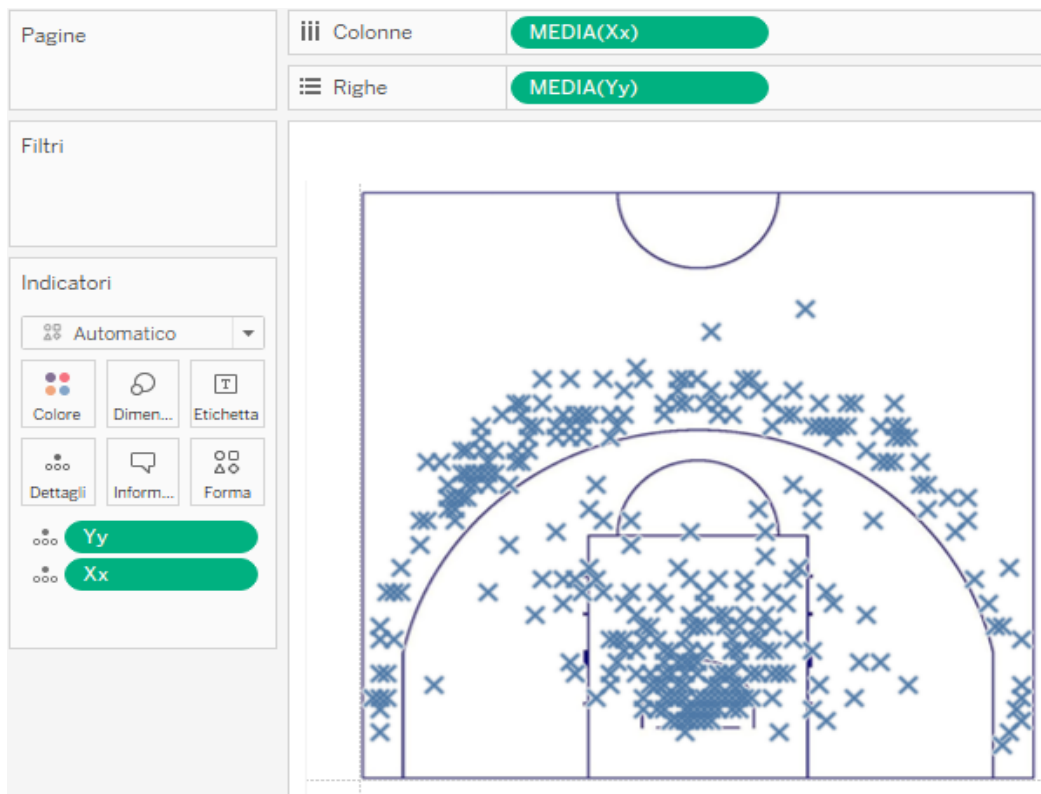


Figura 25. Visualizzazione delle conclusioni nella mappa.

Dopodiché, sono stati modificati la forma ed il colore delle conclusioni in base alla variabile “Result”, indicando con un cerchio verde i canestri segnati (made) e con una croce rossa quelli sbagliati (missed). Per rendere queste dashboard più interessanti, è stata aggiunta l’informazione relativa al nome del giocatore cliccando su ogni conclusione e, come già anticipato nel secondo capitolo, anche la possibilità di selezionare le conclusioni a proprio piacimento attraverso dei filtri, scegliendo i giocatori e le partite interessate.

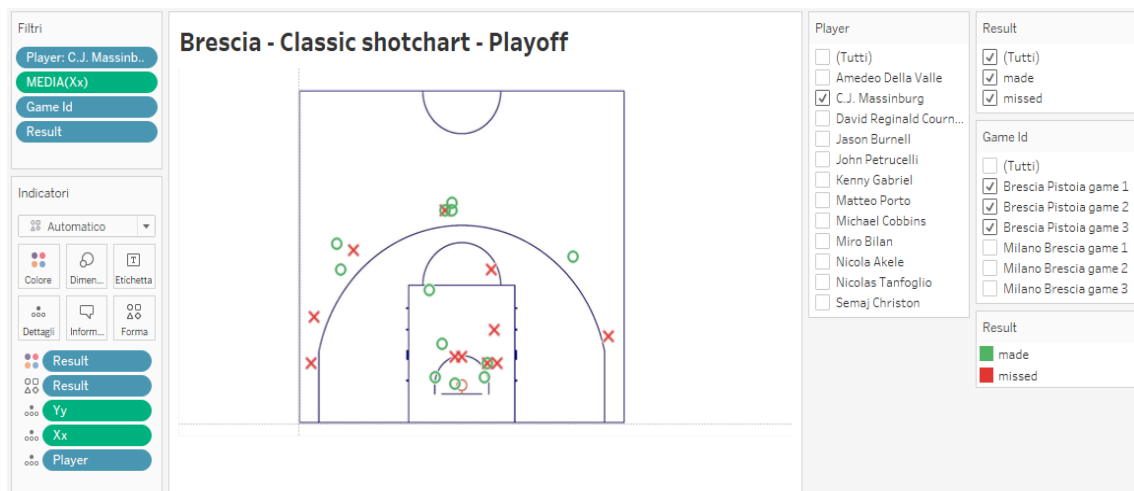


Figura 26. Esempio di selezione dei filtri disponibili.

Dopo aver creato tutte le dashboard necessarie per il report, alcune sono state raggruppate nello stesso foglio per facilitare la lettura e l'interpretazione delle dashboard correlate. Infine, utilizzando l'opzione "storia", tutte le visualizzazioni sono state raccolte in cinque fogli, che hanno costituito il report finale, successivamente pubblicato su Tableau Public e condiviso con lo staff tecnico della Germani Brescia.



## Conclusioni

Con questo elaborato, si è voluto evidenziare come lo sport, ed in particolare la pallacanestro moderna, sono sempre più influenzati e condizionati dalle informazioni ricavabili attraverso l'analisi dei dati; provocando così un forte cambiamento nelle strategie utilizzate dalle società sportive in tutti i campi in cui queste si trovano ad affrontare delle scelte.

L'analisi di grandi quantità di dati nelle prestazioni sportive è in grado di mettere a disposizione una grande mole di informazioni non ricavabili dalle semplici statistiche riportare nei tabellini delle partite, dando così una visione sempre più ampia e precisa delle performance individuali e collettive dei membri di una squadra.

Nel basket, soprattutto, gli sport analysts utilizzano questi dati per cercare di sintetizzare i sistemi di gioco delle squadre, individuando i punti di forza e di debolezza dei giocatori. In questo contesto, le shotchart forniscono delle indicazioni importanti sulle prestazioni offensive degli atleti; ad esempio, riguardo al modo in cui una squadra o un giocatore distribuisce le proprie conclusioni e quanto queste siano efficienti. In quest'ottica, la ricerca di un sistema sempre più efficiente sta rivoluzionando questo gioco, portando le squadre a privilegiare le conclusioni con un rendimento maggiore, i tiri da tre punti e quelli ravvicinati al canestro.

Grazie a queste analisi, gli allenatori possono ottimizzare ogni aspetto del gioco, potenziando le performance dei propri giocatori e affinando le scelte che essi fanno sul campo, migliorando così le prestazioni di tutta la squadra.

Lo stage presso la Germani Brescia ha fornito un esempio concreto di come l'analisi dei big data e di statistiche avanzate sia diventata uno strumento imprescindibile per tutte le squadre che mirano a coprire un ruolo di primo piano nel basket professionistico. Questa esperienza ha mostrato chiaramente come queste dinamiche si concretizzano nel contesto reale, dando una visione dettagliata di quanto la statistica possa migliorare le performance sportive a tutti i livelli.

Tutte queste tecnologie non sostituiscono ciò che il talento, l'intuito e la determinazione possono offrire nel successo di una squadra, ma ne rafforzano l'efficacia, aiutando a prendere decisioni basate su dati concreti e modelli predittivi avanzati. In futuro, questa interazione tra la componente umana e quella analitica sarà sempre più intensa e centrale,

ed è lecito aspettarsi un'ulteriore evoluzione degli strumenti di analisi, spingendo sempre più in là i confini dell'innovazione sportiva.

## Bibliografia

- [1] Catalfano, J. J. (2015). *Moneyball to moreyball: How analytics have shaped the NBA today*.
- [2] Guseo, R. (2006). *Statistica* (pp. 1-763). Cedam.
- [3] Hanrahan, P., Stolte, C., & Mackinlay, J. (2007). visual analysis for everyone. *Tableau White paper*, 4, 23.
- [4] Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3), 299-314.
- [5] Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.
- [6] Rezzani, A. (2017). *Big Data Analytics: Il manuale del data scientist*. Maggioli.
- [7] Sarlis, V., & Tjortjis, C. (2020). Sports analytics—Evaluation of basketball players and team performance. *Information Systems*, 93, 101562.
- [8] Van Rossum, G. (2007, June). Python programming language. In *USENIX annual technical conference* (Vol. 41, No. 1, pp. 1-36).
- [9] Vermeulen, E., & Venkata, S. (2018, October). Big data in sport analytics: applications and risks. In *Industrial Engineering and Operations Management (Presidencia). Proceedings of the International Conference on Industrial Engineering and Operations Management. Conferencia llevada a cabo en IEOM Society International. Pretoria/Johannesburg, South Africa. Recuperado de <https://bit.ly/3ojotk9>*.
- [10] Zuccolotto, P., & Manisera, M. (2020). *Basketball data science: With applications in R*. CRC Press.

## Sitografia

- [1] FIBA. (2024). *FIBA official basketball rules*. International Basketball Federation.  
<https://www.fiba.basketball/documents/official-basketball-rules/current.pdf>
- [2] <https://bodai.unibs.it/bdsports/>
- [3] <https://bodai.unibs.it/bdsports/basketballanalyzer/>
- [4] <https://cbistudio.interworks.com>
- [5] <https://github.com/alericboni/BasketballScraper2024>
- [6] <https://github.com/francescoolivo/BasketballScrapeR>
- [7] [https://github.com/FrancescoRiboni/basketball-analytics-thesis/blob/main/python/1\\_create\\_polygons.ipynb](https://github.com/FrancescoRiboni/basketball-analytics-thesis/blob/main/python/1_create_polygons.ipynb)
- [8] [https://github.com/FrancescoRiboni/basketball-analytics-thesis/main/python/2\\_retrieve\\_polygons.ipynb](https://github.com/FrancescoRiboni/basketball-analytics-thesis/main/python/2_retrieve_polygons.ipynb)
- [9] <https://pypi.org/project/shapely/>
- [10] <https://public.tableau.com/app/profile/fabio.galante/vizzes>
- [11] <https://public.tableau.com/app/profile/francesco.riboni/vizzes>
- [12] <https://teamkeys.it/>
- [13] <https://www.legabasket.it/game/24434/germani-brescia-estra-pistoia/play-by-play>
- [14] <https://www.legabasket.it/la-lega/storia>
- [15] <https://www.python.it/>
- [16] <https://www.r-project.org/about.html>
- [17] <https://www.tableau.com/it-it/why-tableau/what-is-tableau>
- [18] Tiao, S. (2024, 11 marzo). *What Is Big Data?* Oracle.  
<https://www.oracle.com/it/big-data/what-is-big-data/>
- [19] Zanni, A. (2022, 11 novembre). *Moneyball, la rivoluzione dei numeri*. style.corriere.  
<https://style.corriere.it/attualita/moneyball-la-rivoluzione-dei-numeri/>

# Ringraziamenti

Con questi ringraziamenti desidero esprimere la mia gratitudine a tutte le persone che hanno contribuito, direttamente o indirettamente, al raggiungimento di questa importante tappa della mia vita.

Innanzitutto, ringrazio l'Università di Milano Bicocca per avermi concesso la possibilità di intraprendere un percorso di studio che mi ha permesso di esplorare e approfondire un ambito che mi affascina profondamente.

Un profondo grazie va anche al mio relatore, il Prof. Marletta che mi ha affiancato nella realizzazione di questa tesi con una grandissima disponibilità, dandomi numerosi e preziosi consigli che sono stati di grande aiuto.

Ringrazio, poi, le professoresse Manisera e Zuccolotto dell'Università di Brescia, Francesco Taccetti e tutti gli altri membri dello staff della Germani Brescia che hanno reso possibile questa incredibile opportunità di stage, consentendomi di vivere e conoscere da vicino una realtà così tanto stimolante e interessante.

Vorrei poi dire un immenso grazie a quelle persone che, praticamente da sempre, sono vicine in tutti i momenti importanti della mia vita e che mi hanno sempre dato un affetto incondizionato.

A partire dai nonni Maria, Fausto, Marco e la nonna Franca, che, pur non essendo più con noi, sono certo sarebbe orgogliosa di me. Insieme a loro ringrazio tutti gli zii, che sono sempre disponibili e pronti a offrire il loro supporto ogni volta che ce n'è bisogno e rendono ogni momento speciale e significativo.

A tutti i cugini: Nic, Ste, Fede, Gian, Enrica, Ila, Edoardo e Mariasole, vi ringrazio per essere una presenza costante nella mia vita con cui condividere risate, gioie e momenti felici.

A tutti i miei amici: Ale, Matti, Hulk, Albi, Gabry, Gio, Zuck, Penna, Sofi e Pit. Ognuno di voi ha contribuito ad arrivare a questo traguardo. Vi ringrazio per essere sempre stati di supporto, per i numerosi consigli, i momenti di sfogo tra una sessione e l'altra, e per le risate in compagnia.

Alla Emy, che nell'ultimo anno mi hai dato un supporto incondizionato. Con te ho potuto condividere ogni momento speciale, e mi hai insegnato a valorizzare anche le cose più

piccole e semplici, come una giornata di studio o un pomeriggio in compagnia, aiutandomi anche a credere di più in me stesso. Ti ringrazio per il sorriso che porti nella mia vita e per il modo in cui rendi ogni giornata migliore. Grazie per le avventure passate insieme e per tutte quelle che vivremo ancora.

Un grazie enorme va, poi, ai miei genitori, che fin dall'adolescenza hanno sempre accolto e sostenuto ogni mia scelta, appoggiando ogni mio percorso e spronandomi nei momenti in cui le energie o la motivazione venivano meno. La vostra presenza, il vostro aiuto e il vostro supporto sono stati fondamentali per raggiungere questo traguardo.

Un ringraziamento speciale va a mio fratello Ale. Da sempre, e in particolare in questi ultimi tre anni, sei stato sempre al mio fianco, pronto a offrirmi aiuto, consigli e supporto ogni volta che ne avessi bisogno. Mi hai insegnato moltissime cose e mi hai accompagnato in questo percorso come se fosse il tuo. Sei una persona di grande ispirazione per me e la tua presenza ha reso questa sfida più facile.

Concludo ringraziando tutte le persone che sono qui oggi o che saranno vicino a me in queste giornate importanti, forse non ve lo dico mai ma vi voglio bene.