

Analisi descrittiva NBA



Domenico Izzo, Ciro Maccarone, Adelio Antonini

Anno accademico 2023-2024

1 Introduzione

La presente relazione è parte di un'analisi in congiunta con le altre fatte con PowerBI e Tableau, in questo contesto, presentiamo una panoramica delle prestazioni delle squadre e dei giocatori durante le partite regolari delle squadre della NBA dalle stagioni del 1996 fino a quella del 2023.

L'analisi sarà svolta con l'app Qlik, che è molto portata nel generare grafici con il quale sarà possibile interagire con filtri e granularità drill down, separando la natura delle analisi per fogli, delle dashboard con il quale presentare delle analisi sotto un determinato aspetto.

2 Dataset

Il dataset proviene dal sito di Kaggle^[1], dal quale abbiamo prelevato il file contenente i dati. L'autore del file ha prelevato queste informazioni attraverso una API con la quale estrapola le informazioni direttamente dal sito dell'NBA^[2].

Il dataset è un database relazionale contenente i seguenti attributi:

player_name: nome del giocatore.

team_abbreviation: abbreviazione della squadra di appartenenza.

age: età del giocatore.

player_height: altezza in centimetri del giocatore.

player_weight: peso in chilogrammi del giocatore.

[1][NBA Players \(kaggle.com\)](https://www.kaggle.com/datasets/nba/nba-players)

[2]<https://stats.nba.com/>

college: nome dell'università di provenienza del giocatore.

country: Nome del paese di nascita del giocatore, non corrisponde necessariamente alla sua nazionalità.

draft_year: anno di arruolamento.

draft_round: si riferisce al processo annuale di selezione dei giocatori nuovi da parte delle squadre: le squadre scelgono i giocatori in base a un ordine predeterminato, con le squadre più deboli della stagione precedente che scelgono prima. Il draft è composto da una serie di turni, ciascuno dei quali consiste in un numero predeterminato di selezioni. Una selezione nel primo round, ad esempio, è considerata più prestigiosa e spesso associata a giocatori di maggior talento rispetto alle selezioni nei round successivi.

draft_number: si riferisce all'ordine nel quale all'interno del proprio round di arruolamento un giocatore sia stato scelto.

gp: numero di partite giocate durante l'intera stagione.

pts: numero medio di punti segnati.

reb: numero medio di rebound presi dal giocatore, cioè un rimbalzo, cioè l'atto di catturare una palla dopo un tentativo fallito di fare punto.

ast: numero medio di assist forniti

net_rating: è il "differenziale di punti della squadra per 100 possesi quando il giocatore è in campo". Indica quanto la squadra ha un vantaggio o uno svantaggio nei punti durante il tempo in cui il giocatore specifico è in campo, considerando una base di 100 possedimenti per normalizzare i dati. Un valore positivo suggerirebbe che la squadra ha un vantaggio durante quel periodo, mentre un valore negativo indicherebbe uno svantaggio.

oreb_pct: indica la percentuale di offensive rebounds presi dal giocatore, Si verifica quando un giocatore recupera la palla dopo un proprio tiro sbagliato. L'offensivo rebound è prezioso perché offre alla squadra un'altra opportunità di segnare, creando una nuova possesso offensivo. I rimbalzi sono una statistica chiave nel basket e sono spesso utilizzati per valutare l'efficacia di un giocatore nel gioco sotto canestro.

dreb_pct: indica la percentuale di defensive rebound, che consiste nel catturare la palla dopo un tentativo fallito della squadra avversaria di segnare, impedendo quindi ulteriori azioni.

usg_pct: indica la "percentuale d'uso" del giocatore mentre è in campo, cioè quante volte il giocatore mentre era in campo è stato coinvolto nelle azioni della squadra.

ts_pct: indica l'efficienza del giocatore nei tiri a canestro

ast_pct: indica la percentuale di assist commessi del giocatore

season: indica la stagione

Il dataset presenta quindi ad ogni record le prestazioni che il giocatore mediamente per partita ha mantenuto per l'intera stagione indicata: quindi il numero ad esempio indicato nel campo "pts" indica il numero di punti che il giocatore ha segnato in media per partita durante la stagione indicata.

Si noti come non è presente anche in maniera esplicita la squadra, durante la pulizia dei dati abbiamo infatti scoperto che alcune delle abbreviazioni delle squadre, che sono l'unico riferimento che abbiamo, si riferiscono a squadre che hanno cambiato nel tempo nome o hanno cessato di esistere. Menzioniamo dunque i casi particolari trovati:

La sigla "SEA" faceva riferimento ai "Seattle SuperSonics". Tuttavia, la franchigia dei SuperSonics si è trasferita a Oklahoma City nel 2008 e ora è conosciuta come "Oklahoma City Thunder". La sigla "SEA" non è più in uso nella NBA per identificare una squadra attiva.

La sigla "NOK" faceva riferimento ai "New Orleans/Oklahoma City Hornets". Questa sigla fu utilizzata quando la squadra fu temporaneamente trasferita a Oklahoma City a seguito degli effetti dell'Uragano Katrina. Successivamente, la squadra è tornata a New Orleans e ha cambiato il nome in "New Orleans Hornets", e poi ulteriormente rinominata "New Orleans Pelicans". La sigla "NOK" non è più in uso nella NBA.

La sigla "NOH" faceva riferimento ai "New Orleans Hornets". Tuttavia, nel 2013 la squadra è stata rinominata "New Orleans Pelicans". Quindi, attualmente, la sigla "NOP" è quella ufficiale per la squadra di New Orleans nella NBA. La sigla "NOH" non è più in uso.

La sigla "VAN" faceva riferimento ai "Vancouver Grizzlies". Tuttavia, la squadra si è trasferita a Memphis nel 2001 e ora è conosciuta come "Memphis Grizzlies". La sigla "VAN" non è più in uso nella NBA per identificare una squadra attiva.

La sigla CHH e CHA fanno entrambe riferimento ai Charlotte Hornets, abbiamo risolto l'ambiguità sostituendo tutte le ricorrenze della prima abbreviazione con la seconda, in quanto è quella universalmente riconosciuta.

Per facilitare il riconoscimento delle squadre con le sigle, abbiamo deciso di generare una nuova tabella manualmente con tutte le corrispondenze delle abbreviazioni dei team con il rispettivo nome intero. Il campo renderà più semplice la selezione della squadra di interesse durante le analisi, lasciando le abbreviazioni per la legenda. Data la corrispondenza uno ad uno della squadra, con la sua abbreviazione, abbiamo deciso di affiancare un nuovo campo per le squadre: la conferenza. La conferenza si divide in est ed ovest, ed è una suddivisione delle squadre con criterio geografico: le squadre della stessa conferenza si confronteranno tra loro per l'intera stagione. Al termine di questa, le migliori squadre delle due conferenze si sfideranno per nominare il campione della stagione.

Per rendere le analisi ancora più esaustive, si è deciso di aggiungere anche una nuova tabella, nella quale abbiamo riportato il numero di vittorie e sconfitte di ogni squadra per le stagioni considerate. I dati sono stati recuperati dai motori di ricerca e disposti in modo tabellare, in modo da facilitarne la copia nell'app.

Il risultato finale è quindi un dataset che presenta una panoramica completa sulle prestazioni dei giocatori e delle squadre per tutte le stagioni. Le caratteristiche fisiche dei giocatori nella presente vengono ignorate, in quanto trattate con più dettaglio nelle altre analisi menzionate nel primo paragrafo.

3 Definizione delle misure e delle dimensioni

Abbiamo deciso di creare queste dimensioni:

stagione: ci consentirà di osservare come si evolvono le prestazioni nel tempo

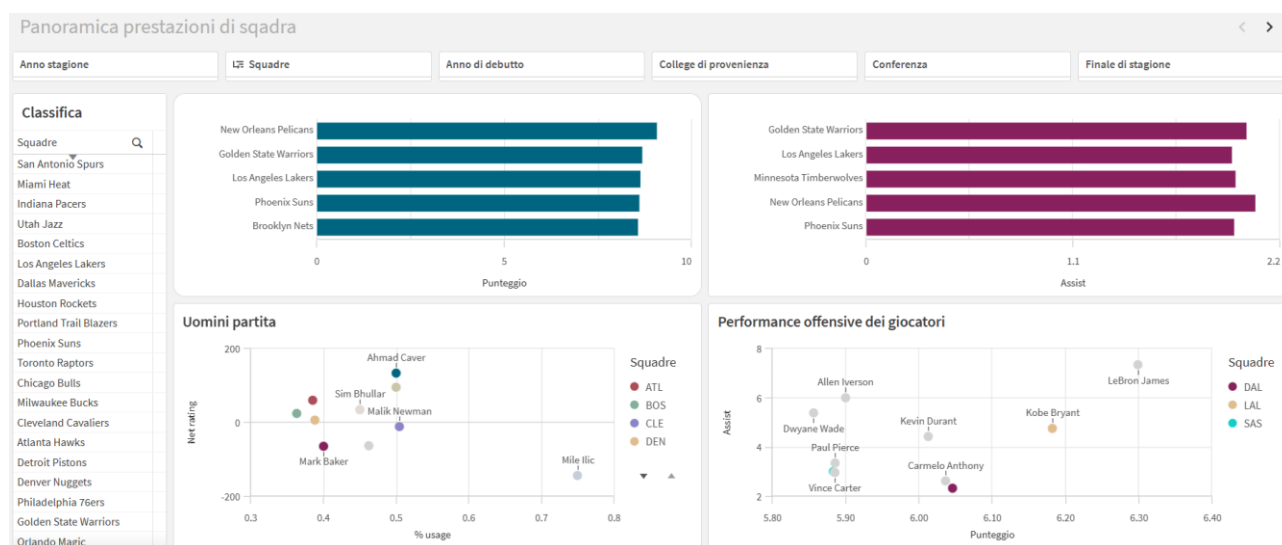
Squadre: la dimensione squadra è di due livelli, la prima ci consente di valutare le prestazioni dell'intera squadra, per poi scendere a livello di giocatore, per valutare le loro prestazioni personali. Osserveremo più avanti come alcune analisi vedranno le analisi per questi due livelli eseguite in maniera separata, in quanto è ricorrente che i giocatori cambino varie squadre nel corso degli anni.

Conferenza: ci permette di selezionare le prestazioni con le squadre con le quale i soggetti al quale siamo interessati effettivamente si confronteranno.

Le misure come gli assist, i punti, i rebound, la percentuale di presenza o di net rating saranno aggregati per media, dando quindi come risultato le prestazioni medie della squadra per partita

4 Foglio 1: Panoramica prestazioni di squadra

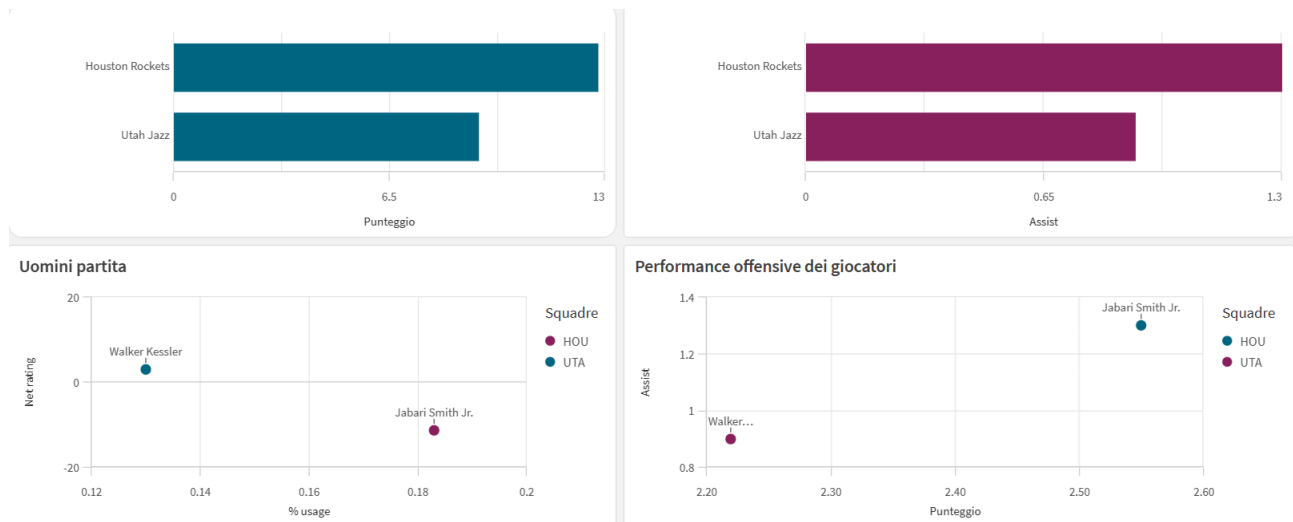
Mostriamo innanzitutto come si presenta il foglio, andando nel seguito a descrivere i filtri e i grafici:



Il foglio fornisce una panoramica sulle prestazioni dei giocatori per stagione e per squadra, consentendo anche di selezionare i giocatori di interesse particolari. Di questi possiamo osservare con l'istogramma orizzontale i punteggi e gli assist fatti mediamente per partita, mettendo in risalto le squadre con le medie più alte o i giocatori più capaci, se selezioniamo dal filtro una squadra, mettendo in risalto i loro migliori giocatori. I grafici a dispersione sotto mostrano in maniera generica attraverso i colori delle bolle le diverse squadre di appartenenza indicate dalla legenda, il grafico a sinistra mostrerà il vantaggio dei giocatori nelle azioni rispetto al net rating, mettendo in risalto quindi i giocatori che sono risultati importanti per le loro squadre, mentre il grafico a dispersione a destra mostrerà i giocatori attaccanti migliori, contraddistinti dal numero di assist e di punteggio più alto. Per terminare, è possibile selezionare dal filtro finale di stagione, la stagione di interesse per mostrare la classifica finale, ottenuta mettendo in ordine decrescente di vittorie.

a. Foglio 1: Analisi dei debuttanti

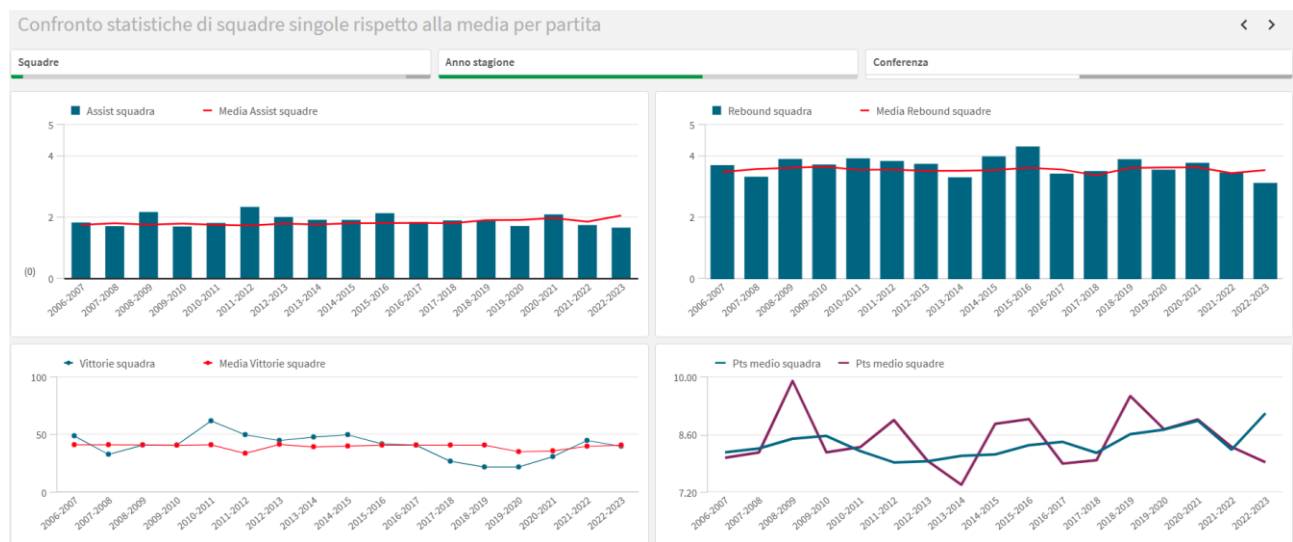
Il foglio visto tiene in considerazione anche gli esordienti: il filtro consente anche di selezionare i giocatori per l'anno di debutto nella associazione di basket nazionale e per il college di provenienza. Mostriamo una vista del foglio selezionando a mero scopo illustrativo, gli esordienti del 2022 provenienti dal college Auburn:



Il foglio rivela dunque due giocatori che rispondono ai criteri: Walker Kessler e Jabari Smith Jr., il foglio ci rivela che quest'ultimo è sicuramente un giocatore che ricopre un ruolo offensivo, dato il numero di assist e punteggio alto, così come una percentuale di intervento nelle partite in media più alto.

5 Foglio 2: Confronto delle prestazioni di una squadra rispetto alla media

Nel foglio precedente abbiamo dato risalto alle squadre e ai giocatori più abili, dando una panoramica che ci consente di capire subito quali sono state le squadre migliori e dando spazio anche ai giocatori a questo riguardo. In questo foglio ci concentriamo invece nel valutare le prestazioni di una squadra alla volta rispetto alla media generale.



Per far osservare come il foglio agisce, abbiamo deciso di selezionare i Chicago Bulls per far comprendere il suo funzionamento: i bin e le linee in blu rappresentano la statistica della squadra selezionata, mentre quella in rosso (e viola per il grafico in basso a destra) la media generale. Qualora non sia fatta alcuna selezione, entrambe le misure rappresenteranno la media generale. Il foglio è quindi stato pensato per fare confronti di questo genere, le misure usate sono state:

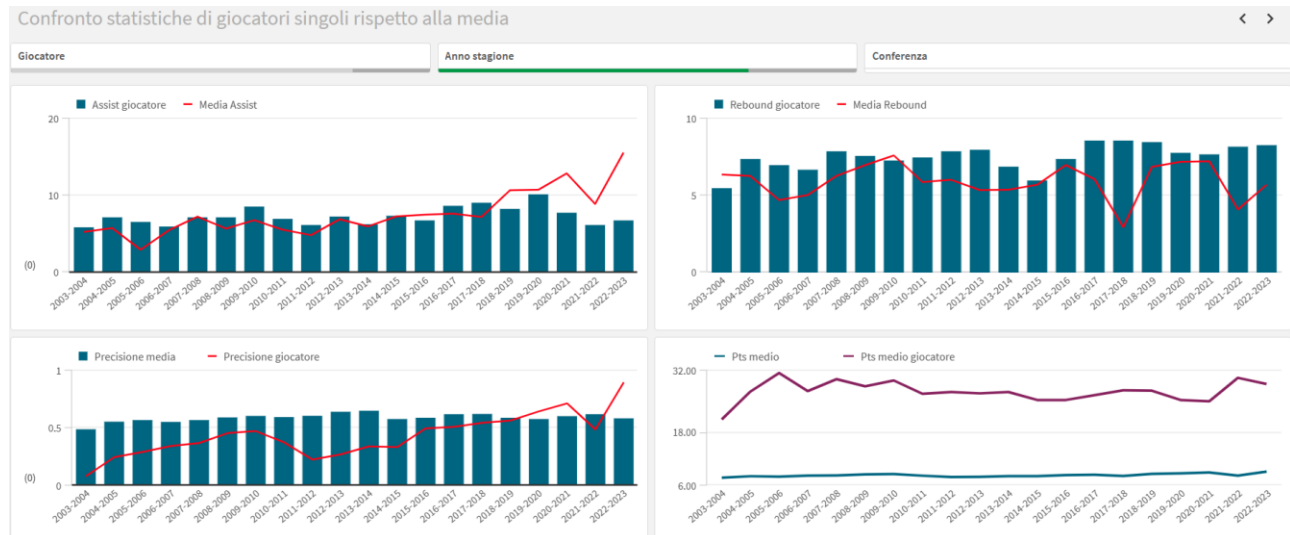
`avg(misura)` per ottenere la media generica per tutte le squadre

`avg({<[Team]>} misura)` che ci consente di calcolare la media per una particolare dimensione

I grafici presentano un confronto per assist, per rebound effettuati, le vittorie e i punti segnati, rispettivamente partendo dal grafico in alto a sinistra fino ad arrivare in basso a destra.

6 Foglio 3: Confronto delle prestazioni dei giocatori rispetto alla media

Presentiamo come nel capitolo precedente, il foglio nella sua interezza:

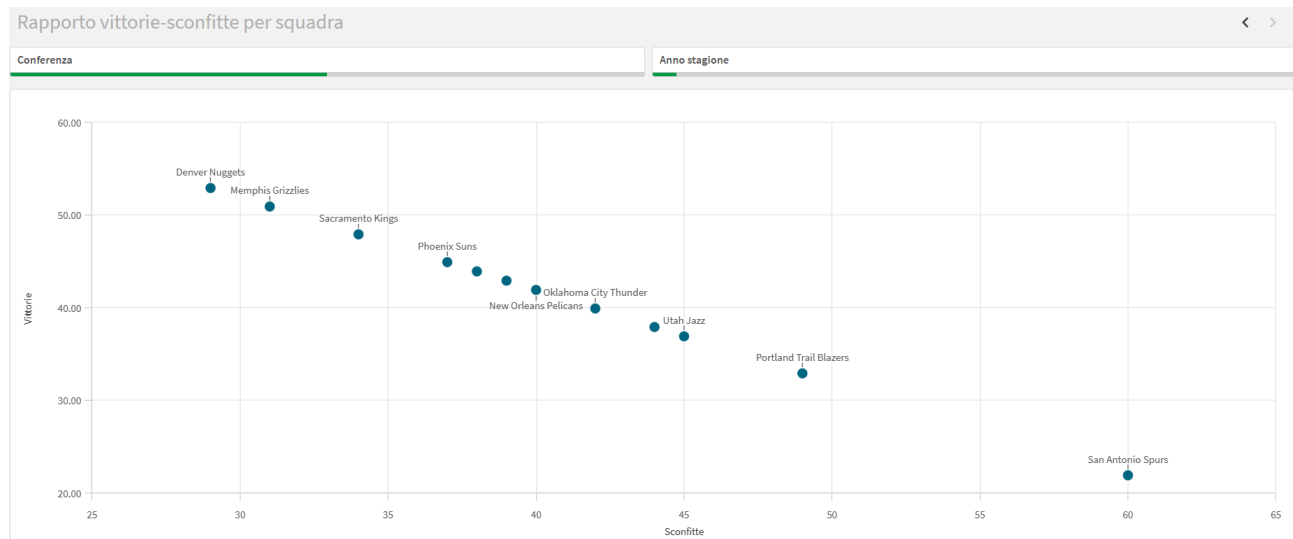


Il foglio presenta gli stessi grafici visti nel foglio precedente, le misure distinte sono calcolate nella stessa maniera come visto precedentemente, con la differenza di mettere il campo del nome del giocatore al posto del campo Team. Il foglio mostra le prestazioni di LeBron James, un giocatore fuoriclasse di fama internazionale, come anche i grafici attestano.

Dato che le vittorie e le sconfitte sono riportate solo a livello di squadra e non riporta infatti la granularità a livello di giocatore, abbiamo pensato di valutare la sua precisione di tiro rispetto alla media, dato che è un indice che non abbiamo esaminato in precedenza, e che ci consente di apprezzare come man mano il giocatore sia diventato via via più accurato, oltre a poter osservare il confronto con gli altri giocatori.

7 Foglio 4: Confronto dei risultati di fine stagione per squadra

Abbiamo potuto osservare le prestazioni delle squadre in termini di capacità di gioco e di risultati raggiunti a fine stagione, ma finora non abbiamo mai osservato il rapporto tra le vittorie e le sconfitte. In questo foglio ci occuperemo proprio di questo, osserveremo con un grafico a dispersione la disposizione delle squadre sulla base delle vittorie e sconfitte subite, che, come vedremo, metterà in risalto il divario esistente tra le squadre.



Nella vista corrente osserviamo la conferenza ovest nella stagione 2022-2023: osserviamo come le squadre si dispongono dalle più forti disposte a sinistra in alto verso l'angolo basso a destra lungo una traiettoria rettilinea ben visibile; si faccia anche caso della distanza che intercorre tra i vari punti, che mette in risalto la differenza delle prestazioni tra le diverse squadre.

8 Conclusioni

L'app creata permette di raggiungere un buon grado di conoscenza della situazione del campionato di pallacanestro più famoso del mondo, è stata pensata per fornire un quadro generico ed intuitivo anche per coloro che non seguono e non conoscono lo sport, consentendo di avere una buona conoscenza anche storica su quali sono le squadre che in genere si trovano in cima alla classifica così come si possono individuare i giocatori di talento sia veterani che per tenere d'occhio i nuovi arrivi. Abbiamo osservato la differenza esistente fra i giocatori nella media ed i giganti di questo mondo, enfatizzando le differenze nelle prestazioni con la possibilità di apprezzare come negli anni siano diventati man mano quello che sono attualmente.