

SPORT ANALYTICS: LA STATISTICA DIVERTENTE

PAOLA ZUCCOLOTTO E MARICA MANISERA

Nei primi mesi del 2016 nasce presso l'Università di Brescia il *Big&Open Data Innovation Laboratory (BODaI-Lab, bodai.unibs.it)*, un laboratorio di ricerca scientifica finalizzato alla creazione di gruppi di lavoro che predispongano - all'interno di specifici progetti - metodi, tecniche e strumenti innovativi per il reperimento, la gestione e l'analisi di dati, con approccio multidisciplinare. Nell'ambito del *BODaI-Lab* le docenti di Statistica Paola Zuccolotto e Marica Manisera hanno dato vita al progetto internazionale *Big Data Analytics in Sports (BDSports, bodai.unibs.it/BDSports)*, con lo scopo di creare una collaborazione tra esperti interessati alle analisi quantitative nello Sport. *BDSports* è impegnato su quattro canali: la ricerca scientifica, l'applicazione sul campo, la didattica e la divulgazione. Le competenze dei ricercatori partecipanti al progetto coprono una vasta gamma di strumenti quantitativi nel campo dei modelli statistici, dell'analisi multivariata, del *data mining*, degli algoritmi di intelligenza artificiale e apprendimento automatico, con particolare riferimento all'analisi dei *big data*.

Questo articolo si inserisce nell'ambito delle attività svolte dal *BDSports* con finalità didattiche. Una missione che ci sta particolarmente a cuore è insegnare agli studenti che le materie di tipo informatico-quantitativo possono anche essere divertenti. Tutti sappiamo che dover decifrare formule matematiche e risolvere problemi complicati non è esattamente ai vertici delle classifiche di gradimento delle materie scolastiche. Nel caso della Statistica, questo problema è aggravato da una frequente marginalità di questa materia rispetto ai programmi usuali di Matematica, che la rende complessivamente poco conosciuta e spesso fraintesa (Zuccolotto, 2016). Tuttavia riteniamo che, come per tutti gli insegnamenti, esistano vari modi per coinvolgere e appassionare gli studenti (almeno quelli più volenterosi), grazie a metodi didattici alternativi e allo sviluppo di progetti applicativi finalizzati, in cui sia chiaramente percepibile l'utilità dello studio della materia. Il binomio Statistica-Sport è perfetto per questo fine, come ben evidenziato anche da Guthrie Donald nella recensione, apparsa sul prestigioso *Journal of the American Statistical Association*, del libro di Antonio Mussino, "Statistica e Sport: non solo numeri", uno dei pochi editi in italiano su questo tema. L'importanza della questione è anche testimoniata dall'interesse dedicatole in occasione della recente MIT SLOAN Sports Analytics Conference, probabilmente l'appuntamento internazionale più importante del settore, che si svolge annualmente dal 2006 a Boston, MA (Drazan et al., 2017). Dal punto di vista del metodo didattico, è possibile organizzare sperimentazioni multidisciplinari (coinvolgendo gli insegnanti di matematica, informatica, educazione fisica), mentre dal punto di vista dei progetti applicativi da sviluppare assieme agli studenti, si può ideare un'ampia gamma di percorsi con differenti

obiettivi specifici (analisi delle prestazioni di squadre ed atleti, monitoraggio nel tempo, studio degli atteggiamenti verso lo sport, analisi del profilo psicologico degli atleti, ...).

Nei paragrafi che seguono sottoporremo all'attenzione dei docenti tre diversi progetti che possono essere realizzati nell'ambito di un percorso formativo scolastico su queste tematiche, pensati per la scuola secondaria superiore. Ovviamente è possibile, partendo da questi spunti, adattare i percorsi di lavoro e ideare nuovi progetti collegati con qualsiasi sport e adatti a qualunque livello di istruzione. I tre progetti differiscono tra loro per la tipologia di lavoro e per gli strumenti (metodologici e informatici) necessari, quindi in definitiva per le competenze che sono in grado di sviluppare (Tabella 1). Per ogni progetto sono descritte nel dettaglio le modalità esecutive e i principali passi da svolgere nell'analisi dei dati. A supporto degli insegnanti che intendessero seguire uno dei percorsi è reso disponibile il relativo materiale didattico, da richiedere via *email* o scaricabile all'indirizzo bodai.unibs.it/BDSports/Students.htm, sezione *Projects for School and University* (le modalità per ottenere la *password* che dà accesso all'area protetta sono riportate sul sito). Inoltre i ricercatori del *BDSports* sono disponibili a incontri di formazione per docenti e studenti, e anche alla creazione di nuovi progetti "su misura", per sport specifici o diversi livelli di istruzione scolastica e universitaria.

Tabella 1: Sintesi dei progetti: competenze sviluppate e disponibilità di materiale didattico dal *BDSports*

	Competenze sviluppate	Materiale didattico <i>BDSports</i>
Progetto 1	<ul style="list-style-type: none"> • Ricerca e consultazione web • Interpretazione dati e grafici statistici 	-
Progetto 2	<ul style="list-style-type: none"> • Reperimento dati sul web • Analisi statistiche descrittive di base (frequenze relative, percentuali, indici di variabilità e di concentrazione): teoria ed applicazione con interpretazione dei risultati • Costruzione matrici di dati, utilizzo di formule e grafici di Microsoft Excel 	File Progetto2.xlsx per la riproduzione delle analisi statistiche descritte
Progetto 3	<ul style="list-style-type: none"> • Costruzione e gestione di <i>form online</i> • Utilizzo funzioni Tabella e Grafico Pivot di Microsoft Excel 	File Progetto3.xlsx per la riproduzione delle analisi statistiche descritte. Su richiesta, possibilità di condivisione del <i>form online</i> già predisposto

PROGETTO 1: ESAME DEI DATI E DELLE STATISTICHE PRESENTI ONLINE

Il primo progetto è prevalentemente finalizzato a far scoprire agli studenti il vasto mondo della Statistica nello Sport. Sarà cura del docente esplorare preventivamente i siti *web* ufficiali relativi a vari Sport (selezionati secondo gli interessi degli studenti) e analizzare in classe il contenuto delle pagine dedicate alle statistiche *online*, focalizzando l'attenzione sulle analisi presentate e sui dati disponibili, da scaricare per svolgere ulteriori elaborazioni.

Le Figure da 1 a 4 riportano alcune pagine di esempio.

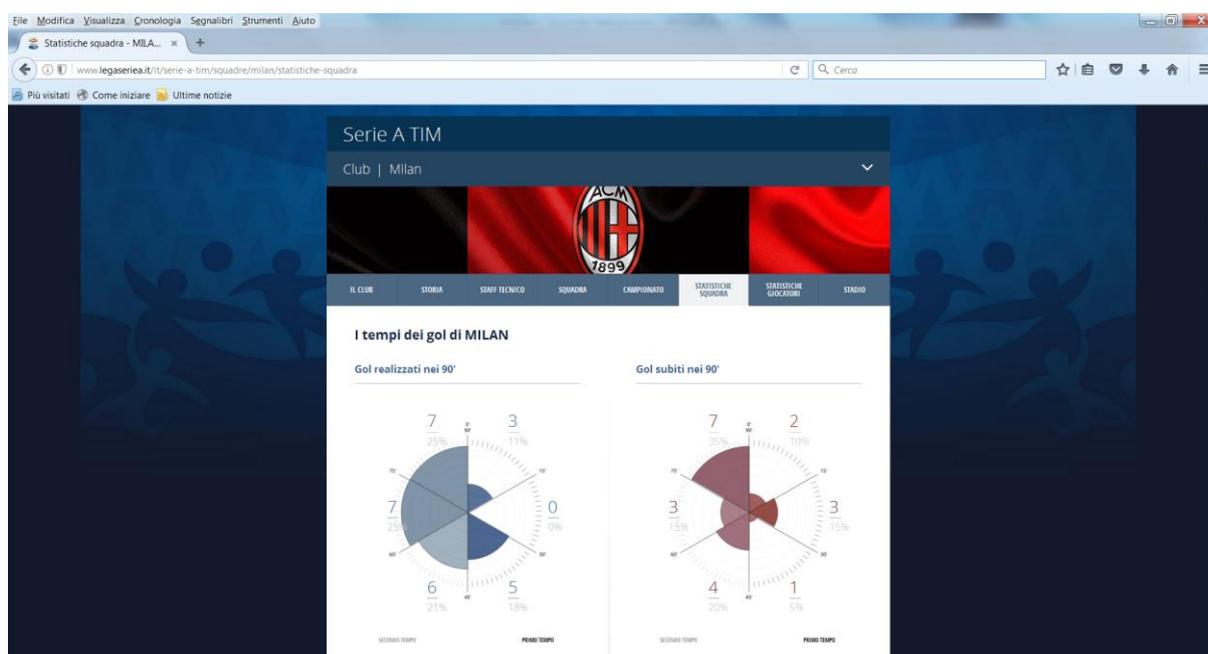


Figura 1: www.legaseriea.it/it/serie-a-tim/squadre/milan/statistiche-squadra (12 gennaio 2017).

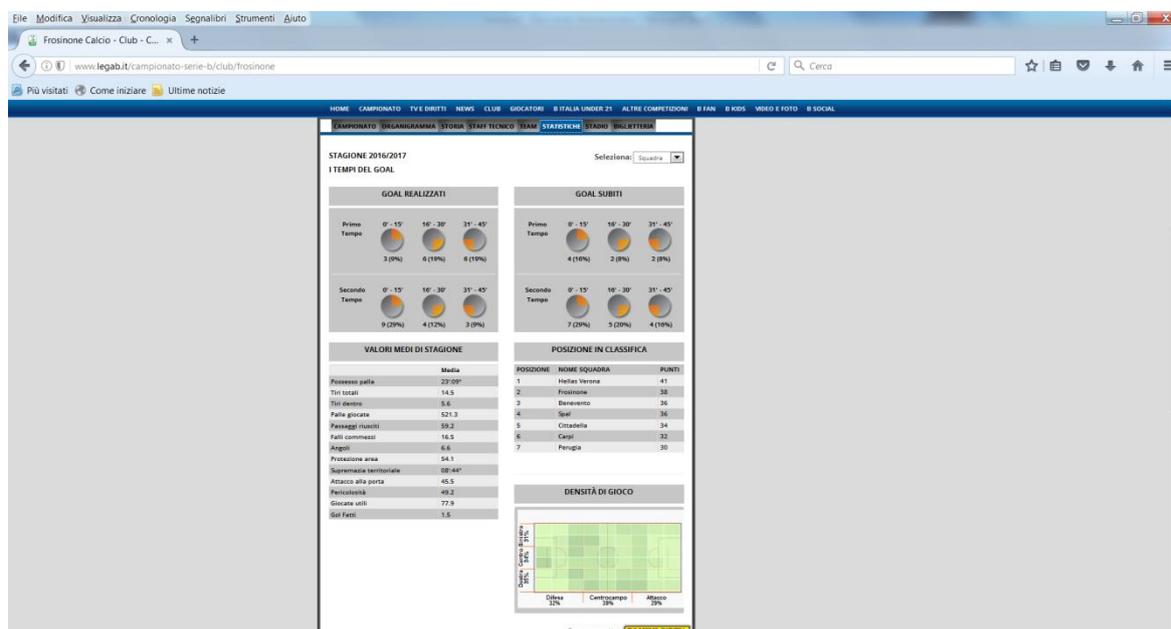


Figura 2: <http://www.legab.it/campionato-serie-b/club/frosinone> (12 gennaio 2017).

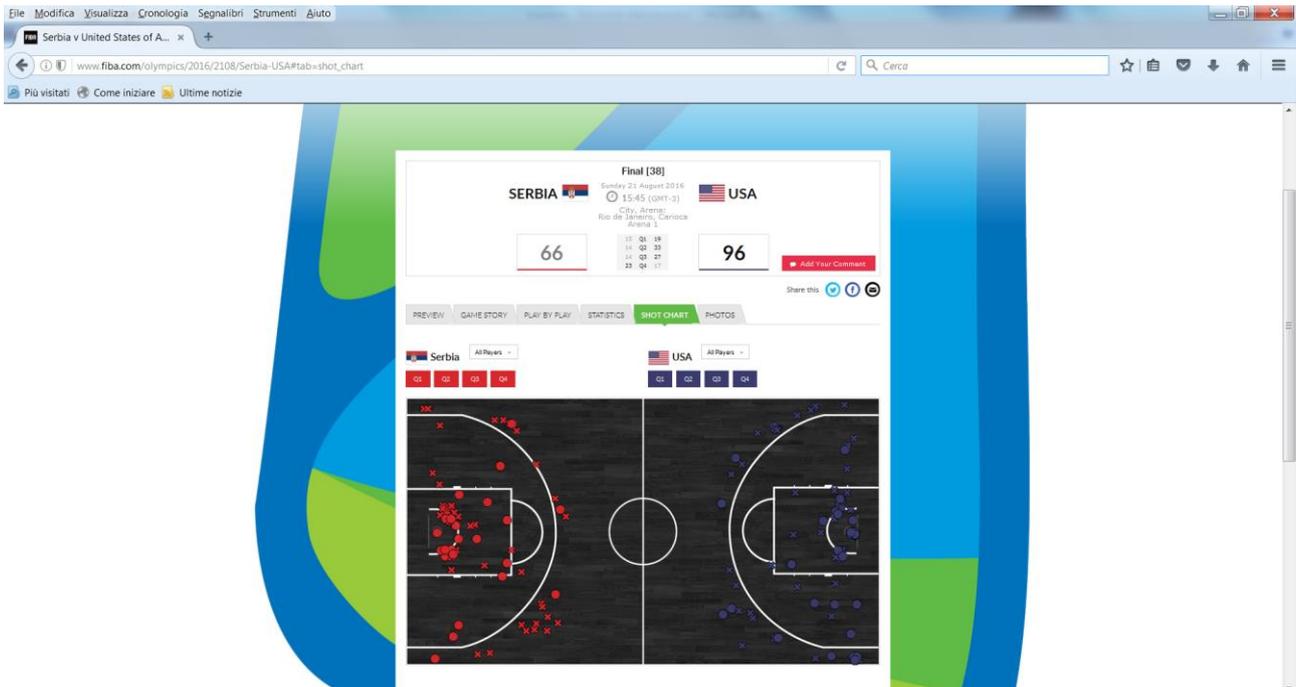


Figura 3: www.fiba.com/olympics/2016/2108/Serbia-USA#tab=shot_chart (12 gennaio 2017).

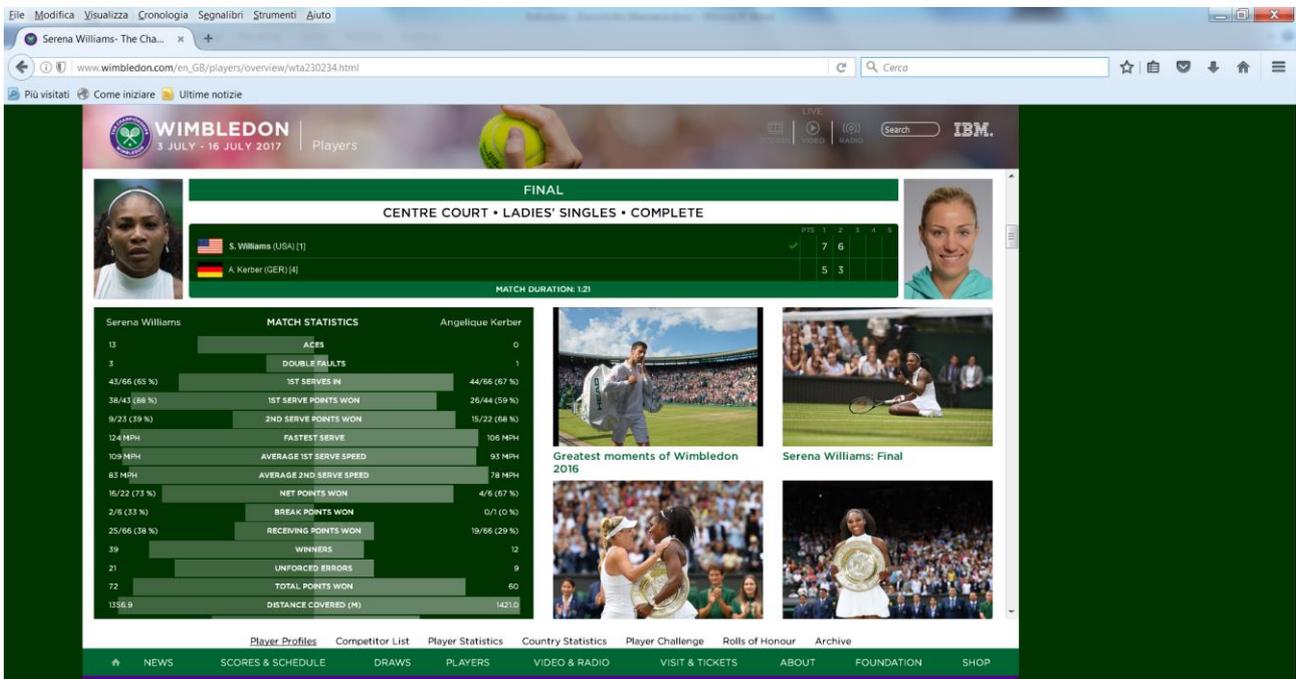


Figura 4: www.wimbledon.com/en_GB/players/overview/wta230234.html (12 gennaio 2017).

PROGETTO 2: ANALISI DELLE PRESTAZIONI DI SQUADRE E GIOCATORI

Il secondo progetto consiste nell'utilizzare strumenti di statistica descrittiva per analizzare le prestazioni di una squadra di pallacanestro e dei suoi giocatori sulla base dei dati, che si possono ricavare online, relativi alle partite giocate in un campionato. Scegliamo ad esempio la squadra Virtus Bologna, militante nel campionato di serie A2 2016-2017 ed analizziamo i dati del girone di andata (partite dal 2 ottobre 2016 al 6 gennaio 2017). Tutte le analisi presentate di seguito sono riproducibili e modificabili (per altre squadre o la stessa squadra, con l'aggiornamento delle nuove partite giocate) utilizzando il file *Progetto2.xlsx* disponibile *online* in una sezione protetta del sito *BDSports* (la *password* di accesso è fornita gratuitamente a chiunque ne faccia richiesta secondo le modalità indicate) e seguendo attentamente le istruzioni in esso riportate. Le analisi proposte sono divise in 6 fasi, in ordine crescente di complessità dal punto di vista delle competenze statistiche e informatiche richieste. Il progetto può, su valutazione del docente, essere svolto in forma parziale, fermandosi alla fase che si ritiene opportuna. Dal punto di vista delle competenze acquisite, attraverso questo progetto è possibile insegnare agli studenti alcune nozioni non banali di Statistica (frequenze relative, percentuali, indici di variabilità e concentrazione) e la loro implementazione pratica attraverso formule e grafici di Microsoft Excel.

FASE I: REPERIMENTO DATI, DESCRIZIONE VARIABILI, ORGANIZZAZIONE IN TABELLA SU MICROSOFT EXCEL

I dati utili per queste analisi sono liberamente disponibili sul sito della Lega Nazionale Pallacanestro, in particolare – per la squadra selezionata – all'indirizzo www.legapallacanestro.com/serie-a2/segafredo-virtus-bologna.

Il *dataset* viene organizzato in matrice dei dati su un foglio di Microsoft Excel, come illustrato in Figura 5. Per ottenere statistiche bene interpretabili è consigliabile selezionare solamente i giocatori che hanno un minutaggio consistente. In questo caso sono stati inclusi nell'analisi coloro che hanno giocato almeno 100 minuti nel periodo considerato.

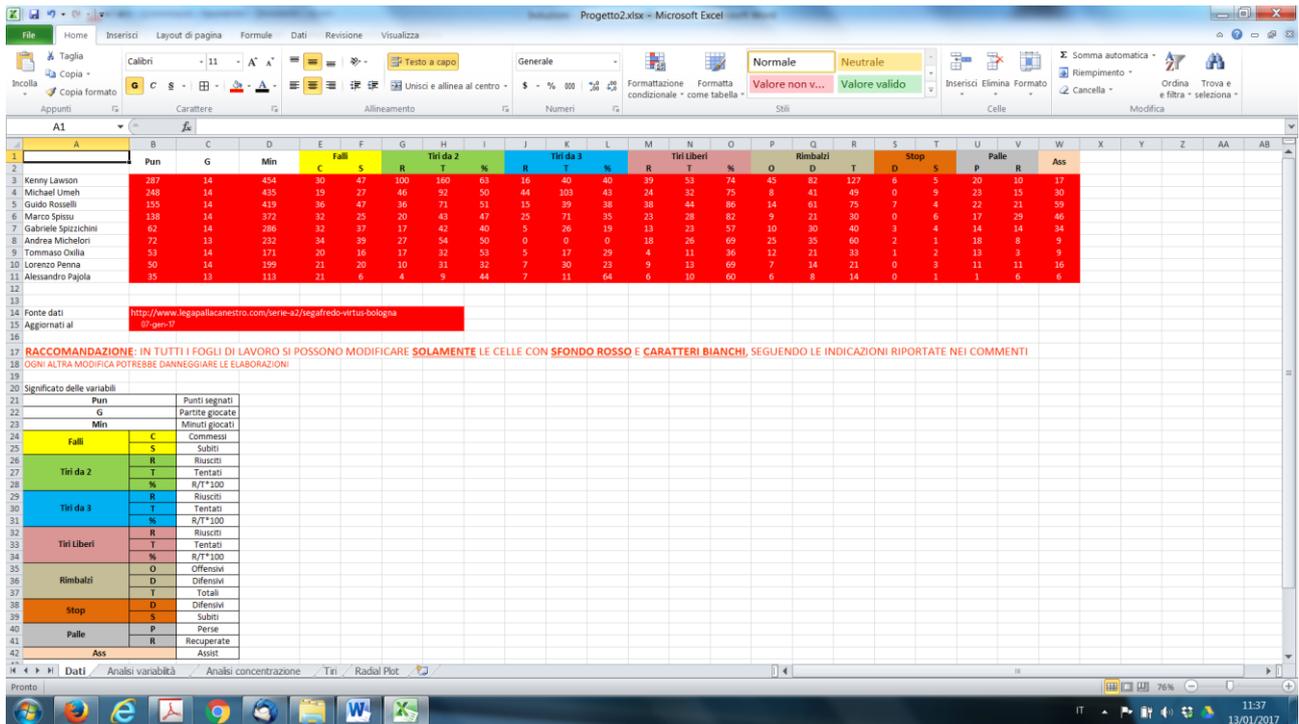


Figura 5: Matrice dei dati su Microsoft Excel e legenda di descrizione delle variabili.

FASE II: ANALISI DELLE PRESTAZIONI DI TIRO

Questa analisi viene fatta attraverso rappresentazioni grafiche che mettano in luce il numero di tiri riusciti e falliti. In Figura 6 è riportato a titolo di esempio il solo grafico relativo ai tiri da 2 punti. Emerge in maniera chiara la superiorità di Kenny Lawson, che ha fatto il numero di tiri più elevato (160) e con la percentuale di riuscita più elevata, pari al 62.5%. Inoltre, la rappresentazione grafica ha anche il pregio di combinare la percentuale di riuscita con il numero di canestri fatti e realizzati, numero che è importante perché i valori percentuali sono tanto più affidabili quanto più è elevato il numero di tiri sui quali essi sono calcolati.

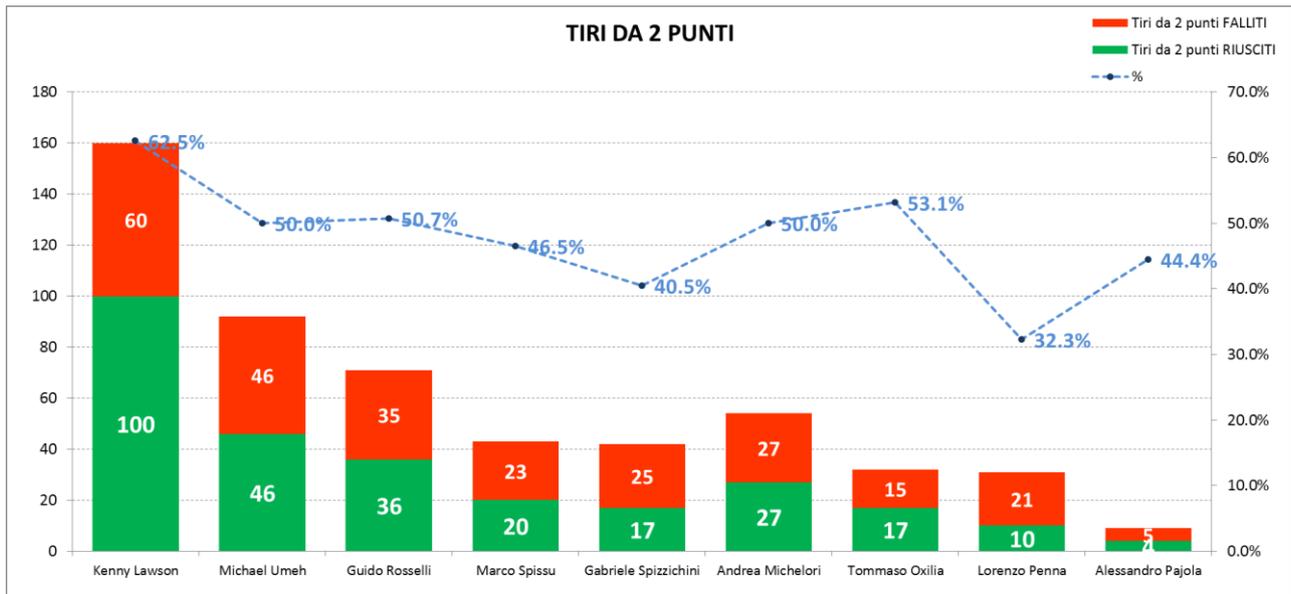


Figura 6: Analisi delle prestazioni dei giocatori nei tiri da 2 punti: tiri riusciti (barre verdi), tiri falliti (barre rosse), tiri totali (altezza totale delle barre verdi + rosse), percentuale di riuscita dei tiri (riusciti/totali*100, riga blu tratteggiata).

FASE III: ANALISI DELLA PRODUTTIVITÀ IN TERMINI DI PUNTI

Per ogni giocatore vengono calcolati i punti segnati al minuto (Punti/Minuti giocati). Tale statistica viene poi convenzionalmente moltiplicata per 40 minuti (la durata di un'intera partita), per migliorarne l'interpretazione. Il grafico di Figura 7 riporta i risultati dei vari giocatori, unitamente all'indicazione dei minuti complessivamente giocati nella porzione di campionato analizzata. I punti medi (per 40 minuti di gioco) segnati da giocatori con diverso minutaggio sono confrontabili; tuttavia, il numero di punti medi di ogni giocatore è tanto più affidabile quanto più elevato è il numero di minuti giocati. Kenny Lawson e Michael Umeh risultano essere i giocatori che segnano di più; il loro numero elevato di minuti giocati aggiunge valore a tale conclusione.

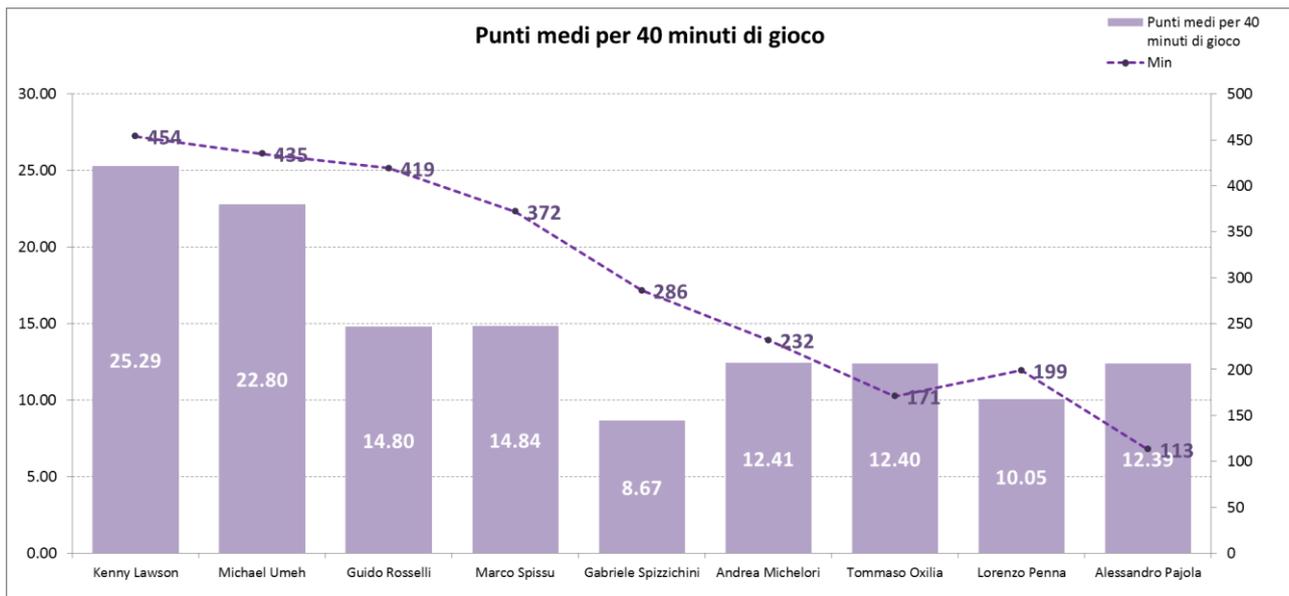


Figura 7: Analisi della produttività in termini di punti: punti medi per 40 minuti giocati (barre viola), minuti complessivamente giocati (linea tratteggiata).

FASE IV: PROFILO COMPLESSIVO DEL GIOCATORE

Lo scopo di questa fase è produrre dei grafici in grado di delineare il profilo di ogni singolo giocatore con riferimento a diverse variabili di gioco. In particolare, interessano i punti segnati, i tiri da 2 e 3 punti e i tiri liberi (tentati e segnati), i rimbalzi totali (offensivi e difensivi), gli stop difensivi e gli assist. Tutte le variabili vengono rapportate ai minuti di gioco (e successivamente moltiplicate per 40, come specificato sopra), in modo da rendere confrontabili i profili di giocatori che hanno minutaggi differenti. Un grafico molto utile per questo tipo di analisi è il grafico radiale, che colloca i valori delle variabili su assi che si diramano da un centro, con angoli di uguale ampiezza (Figura 8). Questo tipo di grafico considera congiuntamente più variabili (in Figura 8 le variabili sono 7, quindi i poligoni che si ottengono sono ettagoni). Giocatori associati a forme poligonali simili hanno profili simili (dal punto di vista delle variabili impiegate). Il poligono tracciato con linea continua si differenzia da quello tracciato con linea tratteggiata solo con riferimento ai tiri, segnati - linea continua - o tentati - linea tratteggiata. Tanto più essi si sovrappongono, tanto migliore è la percentuale di riuscita del giocatore.

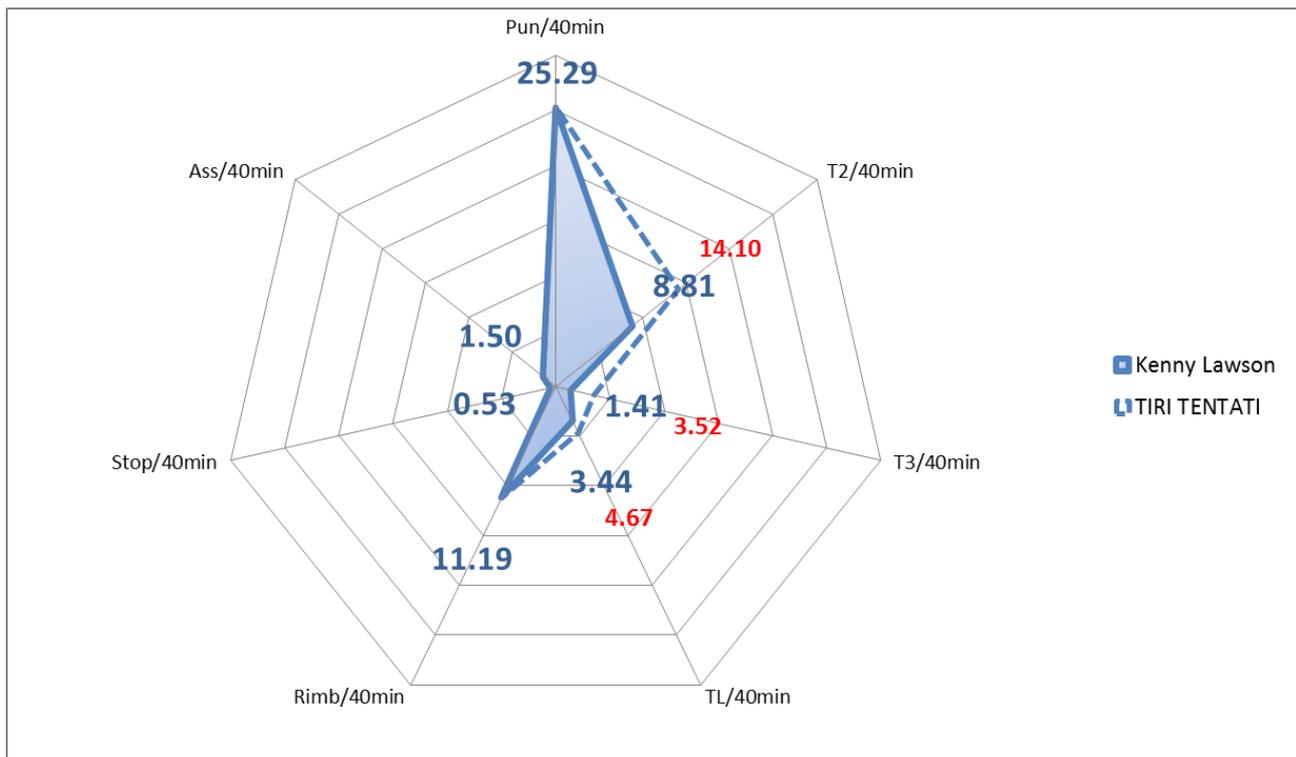


Figura 8: Grafico radiale del profilo di Kenny Lawson secondo le variabili (rapportate a 40 minuti di gioco) Punti, Tiri da 2 punti riusciti, Tiri da 3 punti riusciti, Tiri liberi riusciti, Rimbalzi totali, Stop difensivi, Assist; Tiri da 2 e 3 punti e tiri liberi tentati rappresentati dalla linea tratteggiata e dalle etichette rosse.

FASE V: ANALISI DELLA VARIABILITÀ DI TIRO

Con riferimento separatamente ai tiri da 2 e 3 punti e ai tiri liberi, si considera il numero medio di tiri riusciti per ogni giocatore rapportato ai 40 minuti di gioco. Su queste statistiche si possono calcolare indici di variabilità (scarto quadratico medio e coefficiente di variazione; quest'ultimo, dato dal rapporto tra scarto quadratico medio e media aritmetica, è un indice relativo che consente di effettuare confronti di variabilità; si veda ad esempio il testo di Montinaro e Nicolini, 2010), che informano sulla tendenza della squadra ad essere composta da giocatori con prestazioni molto diverse tra loro da questo punto di vista. I risultati per la squadra in questione sono riportati in Tabella 2. In particolare, osservando i coefficienti di variazione, notiamo che la variabilità è piuttosto alta per i tiri da 3 punti (0.6636), di livello medio-alto per i tiri da 2 punti (0.5715), medio-bassa per i tiri liberi (0.3420). Questo tipo di analisi assume molto più significato se vengono effettuati confronti tra squadre diverse o se si tiene monitorata l'evoluzione temporale del fenomeno per una data squadra lungo l'intero campionato.

Tabella 2: Media, mediana, scarto quadratico medio (SQM), coefficiente di variazione del numero medio di tiri per 40 minuti

	T2R/40min	T3R/40min	TLR/40min
Kenny Lawson	8.81	1.41	3.44
Michael Umeh	4.23	4.05	2.21
Guido Rosselli	3.44	1.43	3.63
Marco Spissu	2.15	2.69	2.47
Gabriele Spizzichini	2.38	0.70	1.82
Andrea Michelori	4.66	0.00	3.10
Tommaso Oxilia	3.98	1.17	0.94
Lorenzo Penna	2.01	1.41	1.81
Alessandro Pajola	1.42	2.48	2.12
MEDIA	3.6737	1.7033	2.3927
MEDIANA	3.4368	1.4097	2.2069
SQM	2.0996	1.1304	0.8183
COEFF. VARIAZIONE	0.5715	0.6636	0.3420

Anche per questo tipo di analisi è possibile costruire una idonea rappresentazione grafica (Figura 9), dalla quale emerge con chiarezza come la media sia una sintesi di valori molto diversi tra loro. Ad esempio, per i canestri da 2 punti, si va da 1,42 a 8,81 canestri per 40 minuti. Inoltre appare immediatamente visibile che il valore 8,81 costituisce una sorta di *outlier*, che segnala una prestazione particolarmente buona di un singolo giocatore.

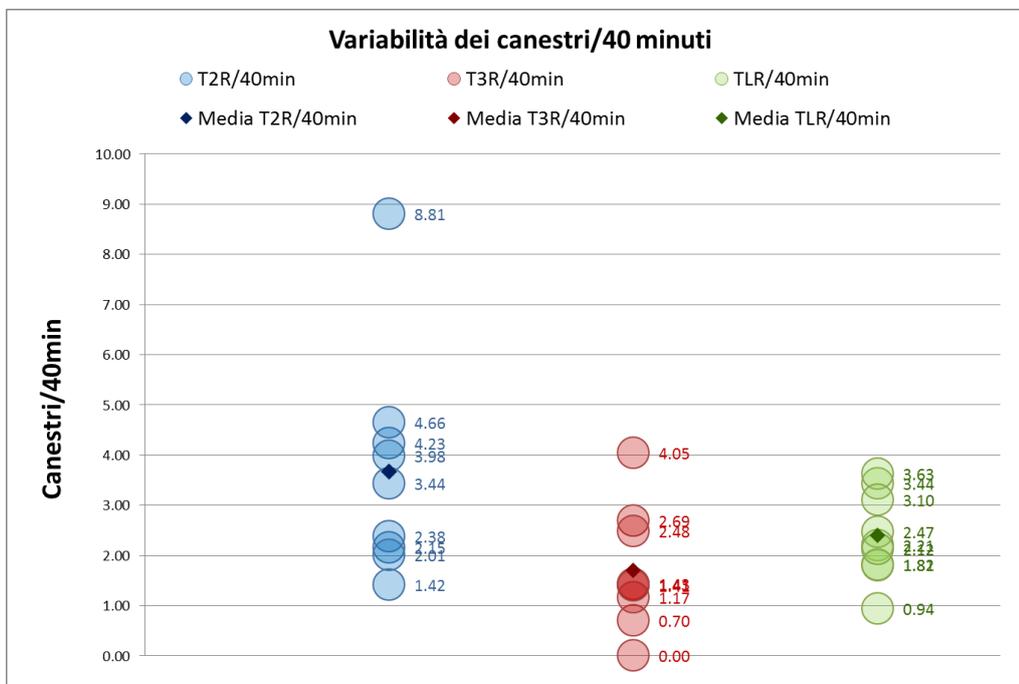


Figura 9: Tiri riusciti da 2 punti (blu), da 3 punti (rosso), liberi (verde), in media sui 40 minuti di gioco; valore medio della squadra (rombo colorato) e valori dei singoli giocatori (bolle).

FASE VI: ANALISI DELLA CONCENTRAZIONE DEI PUNTI

Vi sono squadre in cui il numero di punti è (relativamente) equidistribuito tra i vari membri del *roaster*, e squadre che invece poggiano in modo molto più consistente su un numero esiguo di giocatori. In statistica questa circostanza si analizza facendo riferimento a indici e grafici per lo studio della concentrazione. Relativamente ai punti segnati complessivamente dai vari giocatori, l'analisi della concentrazione prevede l'esistenza (da un punto di vista teorico) di due situazioni limite: quella in cui tutti i punti sono segnati solo giocatore (massima concentrazione) e quella in cui i punti sono esattamente equidistribuiti tra tutti i giocatori (minima concentrazione o equidistribuzione). Usualmente vengono utilizzati un grafico – il diagramma di Lorenz – e un indice – il rapporto di concentrazione di Gini. Per le modalità di costruzione, calcolo e interpretazione si rimanda a un testo di statistica di base (Montinaro e Nicolini, 2010). Qui basterà ricordare che il rapporto di concentrazione di Gini, usualmente utilizzato per misurare la concentrazione dei redditi (per esempio, in un determinato Paese), è un indice che assume valori compresi tra 0 e 1 (100%); tanto più elevato è il valore dell'indice, tanto maggiore è la concentrazione del fenomeno analizzato. L'indice è normalizzato, quindi può essere utilizzato direttamente per fare confronti. Relativamente all'analisi qui presentata, il diagramma di Lorenz è riportato in Figura 10 e il rapporto di concentrazione di Gini risulta pari al 42.77% del suo massimo teorico. Si può concludere che nel girone di andata del campionato la squadra Virtus Bologna ha avuto una concentrazione medio-bassa del numero di punti segnati, il che significa che il carico è ben distribuito tra i diversi giocatori. Anche in questo caso l'analisi accresce il suo interesse se si effettuano confronti tra squadre diverse o se si tiene monitorata l'evoluzione temporale del fenomeno per una data squadra lungo l'intero campionato.

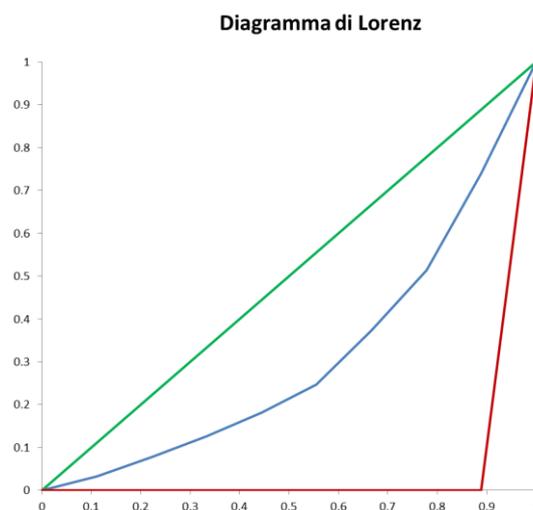


Figura 10: Diagramma di Lorenz dei punti segnati; retta di equidistribuzione (verde), spezzata di massima concentrazione (rosso), spezzata di Lorenz (blu). Il grafico si interpreta valutando quanto la spezzata blu si avvicina alle due situazioni estreme.

PROGETTO 3: MONITORAGGIO PRESTAZIONI DEGLI STUDENTI

In questo progetto gli studenti analizzano le proprie prestazioni sportive, le confrontano con quelle dei compagni e tengono monitorato nel tempo il loro andamento.

Per questo progetto è essenziale l'interazione con l'insegnante di Educazione Fisica, in quanto la rilevazione dei dati relativi alle prestazioni deve avvenire durante le sue lezioni. Il progetto prevede che siano selezionate alcune discipline sportive (in questo esempio, a scopo puramente esemplificativo, si è optato per le seguenti di atletica leggera: 80 m piani, 200 m piani e salto in lungo) e che vengano misurate le prestazioni degli studenti a intervalli regolari durante il corso dell'anno. Ad ogni sessione di misurazione vanno registrate, tramite un apposito *form online*, alcune variabili relative allo studente (alcune di esse ovviamente immutabili come classe, sezione, genere,...) e i risultati ottenuti.

Oltre che per la finalità e le modalità esecutive, questo progetto differisce dal precedente per gli strumenti utilizzati e le competenze sviluppate. Infatti prevede che gli studenti imparino a costruire il *form online* nominato in precedenza per la raccolta e l'archiviazione automatica dei dati, da elaborare successivamente con le funzioni Tabella e Grafico Pivot di Microsoft Excel. Come per il Progetto 2, le analisi in Excel sono riproducibili e modificabili utilizzando il file *Progetto3.xlsx* disponibile *online* in una sezione protetta del sito *BDSports* (la *password* di accesso è fornita gratuitamente a chiunque ne faccia richiesta secondo le modalità indicate) e seguendo attentamente le istruzioni in esso riportate.

FASE I: COSTRUZIONE *FORM ONLINE* E RACCOLTA DATI

Un *form online* è l'interfaccia di un'applicazione che consente di inserire informazioni digitate dall'utente direttamente su una scheda di raccolta dati disponibile sul *web* e archivarle in modo da renderle successivamente disponibili. Si tratta di strumenti di facile costruzione e ancor più semplice utilizzo, accessibili via *computer*, *smartphone* e *tablet*, che consentono svariate applicazioni di tipo statistico.

A questo fine proponiamo l'utilizzo del noto strumento Google Forms (www.google.com/forms/about/), gratuito e intuitivo, che richiede soltanto l'attivazione preliminare di un account Google. In Figura 11 è riportato lo schema della scheda di raccolta dati, suddivisa in due sezioni (Informazioni sullo studente e Risultati dei test) e disponibile per consultazione e prova al *link* goo.gl/forms/60Te0a82ZAVv9qDF2 (Figura 12). La prima fase di esecuzione del progetto consiste nel riprodurre assieme agli studenti il *form online*. In alternativa, è possibile chiedere via *email* a *BDSports* di attivare una copia

condivisa (accessibile solo a chi ne fa richiesta) del *form* già predisposto, con successivo accesso riservato ai fogli dati.

Come già illustrato, il progetto richiede che durante l'anno vengano stabilite un certo numero di sessioni di rilevazione delle prestazioni degli studenti nelle discipline selezionate. Ad ogni sessione di rilevazione sarà cura degli studenti stessi registrare i dati, utilizzando, tramite *smartphone* o *tablet*, il *form online* predisposto. Al momento della predisposizione del *form online*, viene automaticamente creato un foglio di tipo Google Sheet (Figura 13) residente sul *cloud* e accessibile tramite Google Drive, che viene aggiornato in tempo reale ad ogni nuova immissione dati. In ogni momento, e in particolare alla fine di ogni sessione di misurazione, i dati possono essere consultati *online* e scaricati in vari formati, tra cui il foglio di lavoro Microsoft Excel.

1. Cognome dello studente *	7. Numero sessione di misurazione *
2. Classe *	8. 80 m piani (centesimi di secondo) *
3. Sezione *	9. 200 m piani (centesimi di secondo) *
4. Genere <i>Contrassegna solo un ovale.</i> <input type="radio"/> Maschio <input type="radio"/> Femmina	10. Salto in lungo (centimetri) *
5. Svolgi attività sportiva agonistica * <i>Contrassegna solo un ovale.</i> <input type="radio"/> Sì, atletica leggera <input type="radio"/> Sì, altro sport <input type="radio"/> No	
6. Quanto ti sei allenato rispetto al test precedente? * <i>Contrassegna solo un ovale.</i> <input type="radio"/> Per nulla <input type="radio"/> Poco <input type="radio"/> Abbastanza <input type="radio"/> Molto	

Figura 11: Scheda di sintesi del *form online* (SINISTRA: sezione 1 - Informazioni sullo studente; DESTRA: sezione 2 - Risultati dei test).

UNIVERSITÀ DI BRESCIA - HEALTH
BDSports
 Big DATA Analytics in SPORTS

Progetto didattico "Sport e Statistica"
ATLETICA LEGGERA

INFORMAZIONI SULLO STUDENTE
*Campo obbligatorio

Cognome dello studente *
 La tua risposta

Classe *
 La tua risposta

Sezione *

Figura 12: *Form online* del Progetto didattico "Sport e Statistica" Atletica Leggera.

Progetto3
 File Modifica Visualizza Inserisci Formato Dati Strumenti Modulo Componenti aggiuntivi Guida Tutte le modifiche sono state salvate in Drive

Informazioni cronologiche											
A	B	C	D	E	F	G	H	I	J	K	
1	Informazioni cronologiche	Cognome dello studente	Classe	Sezione	Genere	Svolgi attività sportiva agonistica	Quanto ti sei allenato rispetto al test precedente?	Numero sessione di misurazione	80 m piani (centesimi di secondo)	200 m piani (centesimi di secondo)	Salto in lungo (centimetri)
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											

Figura 13: Foglio di raccolta dati automaticamente predisposto in Google Sheet per il *form online*.

FASE II: ELABORAZIONE DATI

I dati raccolti durante le sessioni di misurazione possono essere elaborati in vari modi. Proponiamo qui l'utilizzo delle funzioni Tabella e Grafico Pivot di Microsoft Excel, strumenti efficienti e flessibili per creare svariate tipologie di tabelle di sintesi. Le analisi verranno svolte con le seguenti finalità specifiche:

1. valutare le prestazioni medie degli studenti nelle varie discipline, complessivamente e separatamente per le varie sessioni di misurazione;
2. monitorare l'andamento nel tempo delle prestazioni per singolo studente;
3. mettere in luce le differenze nelle prestazioni medie per attività sportiva svolta all'esterno della scuola, classe, genere.

Tutte le analisi presentate di seguito sono riproducibili sui propri dati utilizzando il file `Progetto3.xlsx` anch'esso disponibile *online* nella sezione protetta del sito *BDSports*. Il punto di partenza è costituito dalla predisposizione di una griglia per Tabella Pivot contenente tutte le colonne del dataset, che verrà poi utilizzata per le varie analisi (Figura 14).

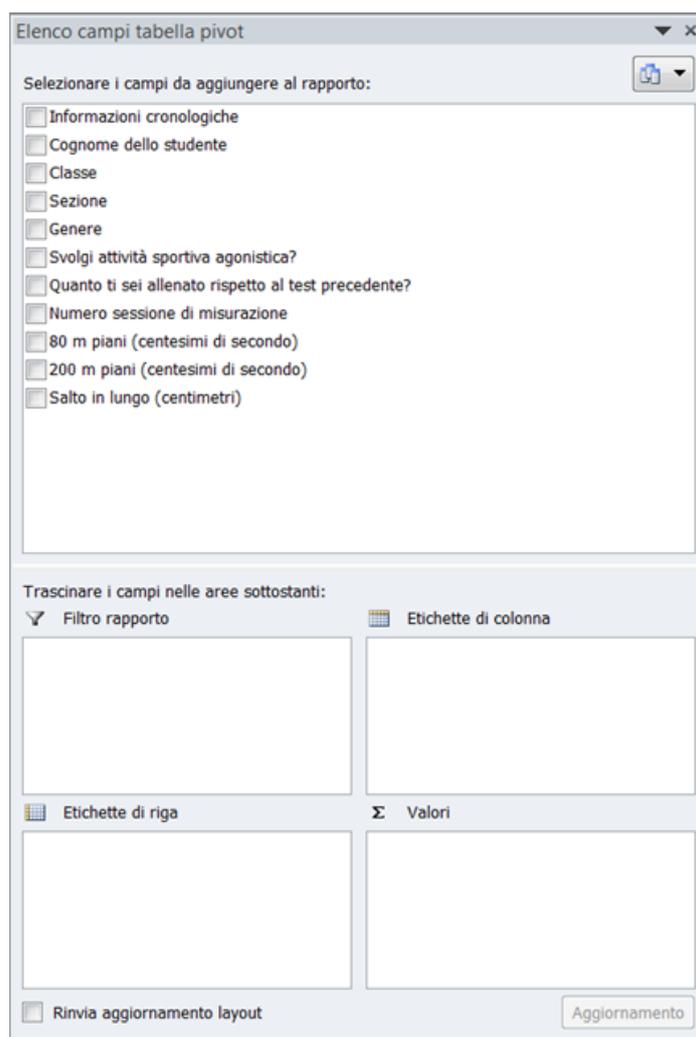


Figura 14: Elenco campi per Tabella Pivot (menù Inserisci/Tabella pivot).

Di seguito analizzeremo le tre finalità specifiche separatamente, utilizzando una matrice di dati simulati artificialmente e focalizzando l'attenzione a scopo esemplificativo sulle sole prestazioni relative agli 80 m piani.

1. Valutare le prestazioni medie degli studenti nelle varie discipline, complessivamente e separatamente per le varie sessioni di misurazione

Selezionando i campi come mostrato nel blocco sinistro di Figura 15 (attenzione al campo valori: bisogna selezionare “Impostazioni campo valore...” e successivamente “Riepiloga campo valore per... Media”) si ottiene la tabella in alto a destra, che riassume le prestazioni medie lungo le varie sessioni di misurazione (6 sessioni in questo esempio). Con i dati della tabella può poi essere costruito il grafico sottostante, avendo cura di impostare una scala sull'asse delle ordinate che contenga le prestazioni individuali minime e massime.

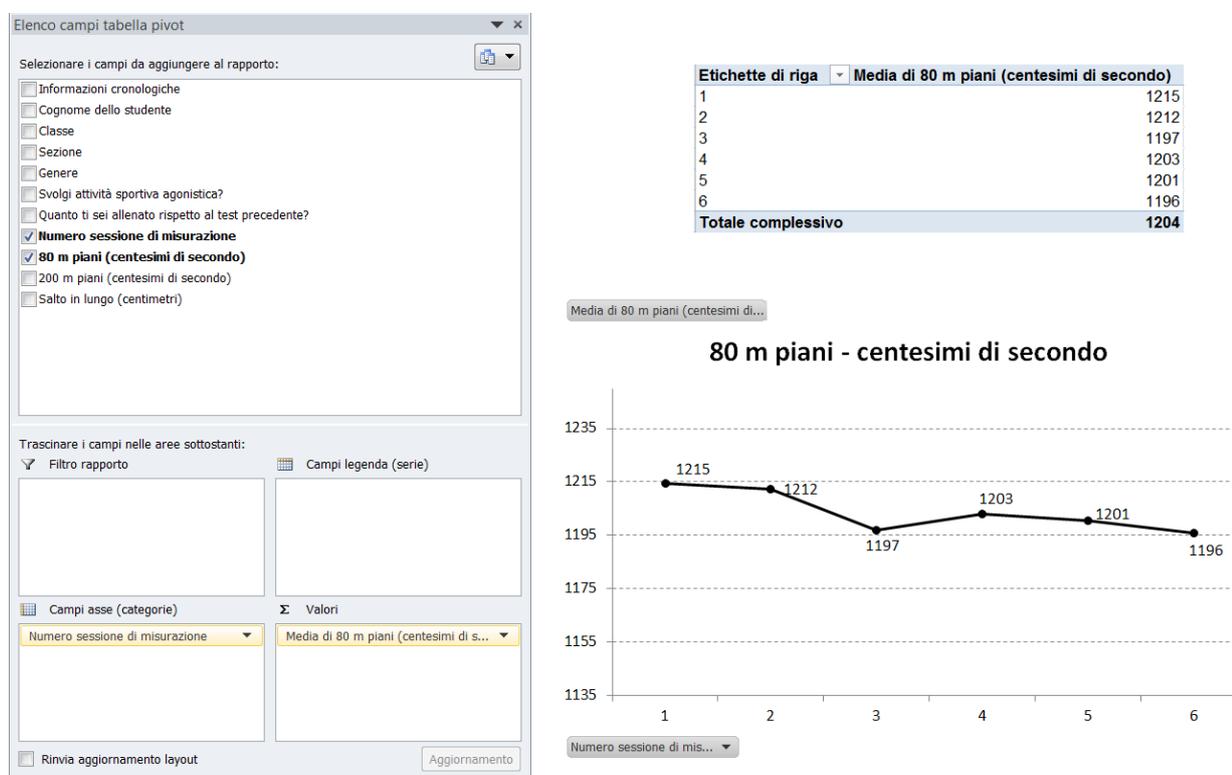


Figura 15: Analisi delle prestazioni medie negli 80 m piani lungo le varie sessioni di misurazione.

Il grafico ottenuto mostra che, in questo esempio, le prestazioni hanno subito un leggerissimo miglioramento lungo le varie sessioni, passando da una media di 12”15 a 11”96. La media complessiva si attesta su 12”04.

Sono evidentemente necessari approfondimenti per capire come tale andamento si differenzia per singolo studente (punto 2.) e se esistono differenze nella prestazione media a seconda della classe frequentata, del genere e del fatto di svolgere o meno attività sportiva agonistica (punto 3.).

2. Monitorare l'andamento nel tempo delle prestazioni per singolo studente

Per il punto 2. è sufficiente inserire la variabile “Cognome dello studente” nell’area denominata “Filtro rapporto”. Appare nella tabella e nel grafico un menù a tendina che consente di selezionare i vari studenti, rappresentando così automaticamente i risultati delle loro prestazioni e il relativo grafico. Si osserva che l’andamento medio complessivo rappresentato in Figura 15 è una sintesi di andamenti singoli profondamente diversi tra loro (Figure 16 e 17).

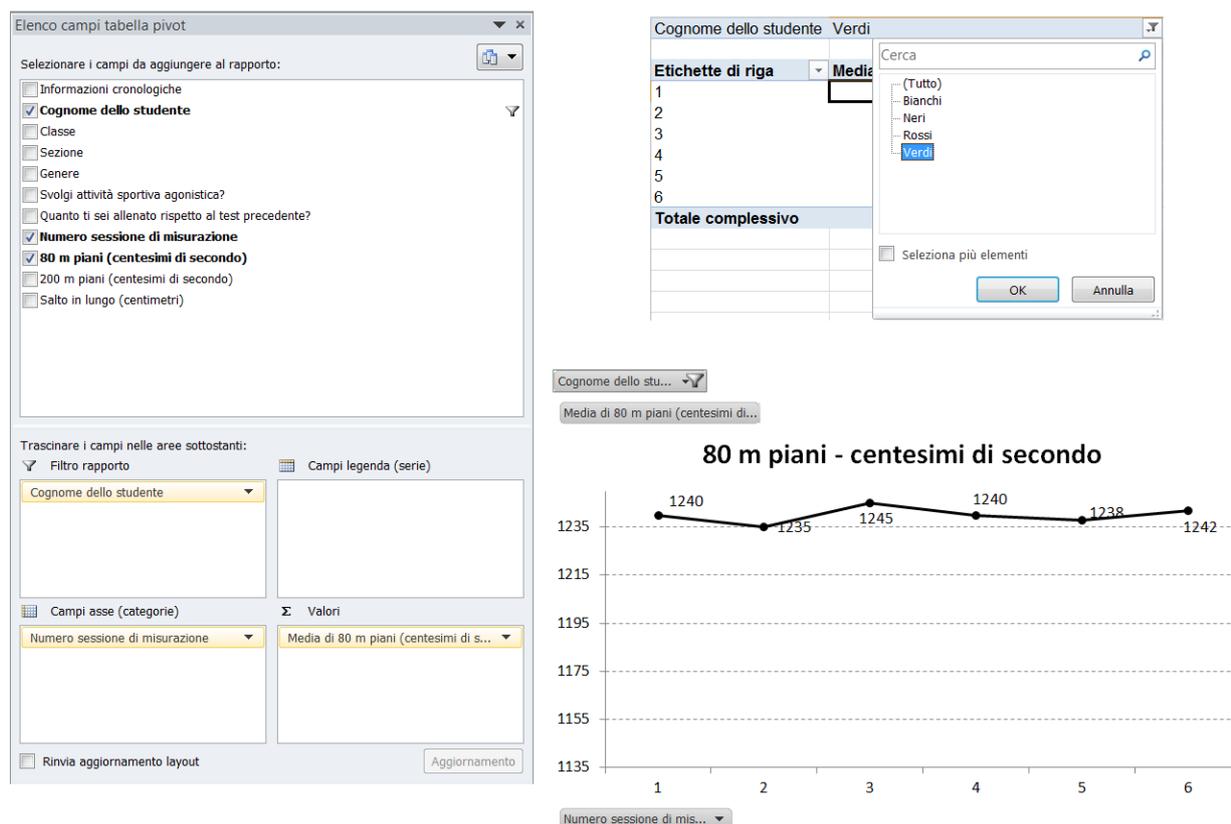


Figura 16: Analisi delle prestazioni medie negli 80 m piani lungo le varie sessioni di misurazione – selezione dello studente Verdi.

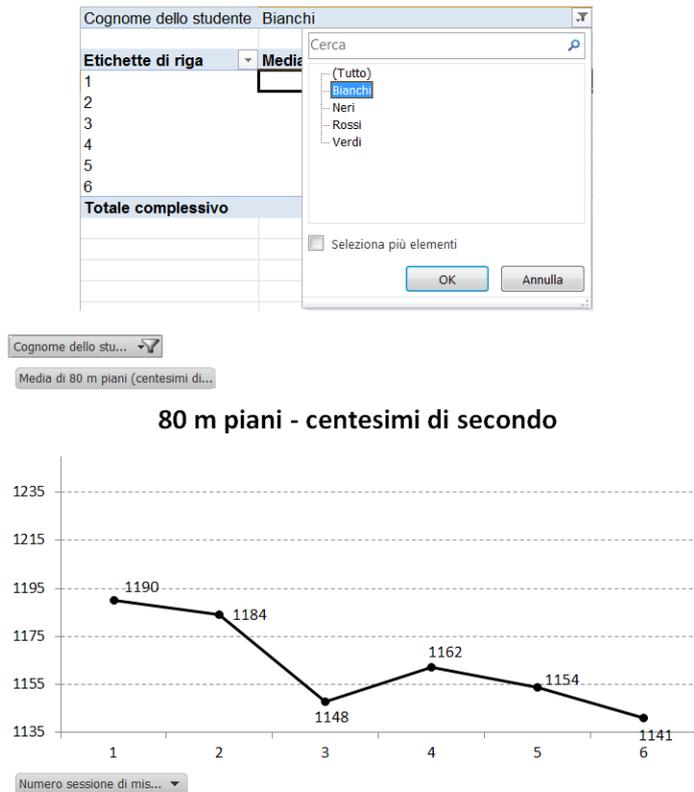
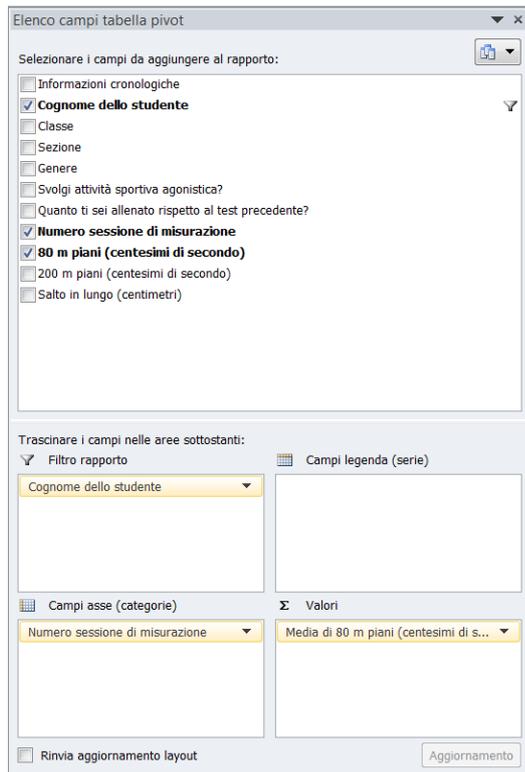


Figura 17: Analisi delle prestazioni medie negli 80 m piani lungo le varie sessioni di misurazione – selezione dello studente Bianchi.

A questo punto è anche possibile, dopo aver illustrato agli studenti da un punto di vista teorico le modalità di interpolazione di una retta a una nube di punti con il metodo dei minimi quadrati (Montinaro e Nicolini, 2010), tracciare la retta interpolante i vari profili per offrire una sintesi dell'andamento.

Tale operazione si effettua selezionando “Aggiungi linea di tendenza” nell'apposito menù e scegliendo poi l'interpolante lineare con visualizzazione dell'equazione sul grafico, come illustrato in Figura 18. Osservando il valore del coefficiente angolare della retta, emerge che lo studente Bianchi ha avuto un miglioramento medio di 9.17 centesimi di secondo per sessione di misurazione. Cambiando la scelta dello studente nel menù a tendina la retta interpolante e la relativa equazione si aggiornano automaticamente.

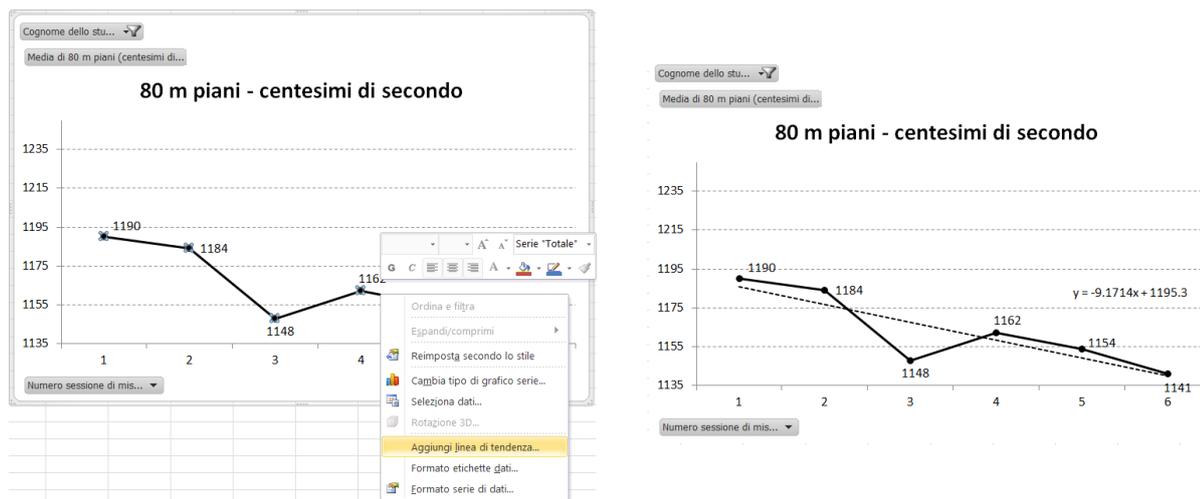


Figura 18: Interpolazione lineare per sintetizzare l'andamento delle prestazioni dei singoli studenti.

3. Mettere in luce le differenze nelle prestazioni medie per attività sportiva svolta all'esterno della scuola, classe, genere

Su questo punto sono possibili un gran numero di analisi diverse, finalizzate a capire in che misura le prestazioni risentono di variabili di contesto come la classe frequentata, il genere, il fatto che lo studente svolga o meno attività sportiva agonistica extrascolastica,

Se si ha ancora interesse per il monitoraggio dell'andamento nel tempo, si può utilizzare la tabella già impostata ai punti 1. e 2. e semplicemente aggiungere la variabile di contesto nell'area "Filtro rapporto", ottenendo così un menù a tendina analogo a quello utilizzato per rappresentare gli andamenti dei singoli studenti, che calcolerà in questo caso le prestazioni medie per gruppi di studenti (ad esempio, maschi e femmine, se la variabile inserita come filtro è il Genere). In alternativa, si può spostare la variabile nell'area "Campi legenda", per ottenere un'analisi comparata sulla stessa tabella/grafico, come rappresentato in Figura 19, relativamente alla variabile che indica l'attività sportiva extrascolastica svolta dallo studente.

Elenco campi tabella pivot

Selezionare i campi da aggiungere al rapporto:

- Informazioni cronologiche
- Cognome dello studente
- Classe
- Sezione
- Genere
- Svolgi attività sportiva agonistica?**
- Quanto ti sei allenato rispetto al test precedente?
- Numero sessione di misurazione**
- 80 m piani (centesimi di secondo)**
- 200 m piani (centesimi di secondo)
- Salto in lungo (centimetri)

Trascinare i campi nelle aree sottostanti:

Filtro rapporto

Etichette di colonna: Svolgi attività sportiva agonistica?

Etichette di riga: Numero sessione di misurazione

Valori: Media di 80 m piani (centesimi di s...)

Rinvia aggiornamento layout

Aggiornamento

Media di 80 m piani (centesimi di secondo)		Etichette di colonna			
Etichette di riga		No	Si, altro sport	Si, atletica leggera	Totale complessivo
1		1240	1214	1190	1215
2		1235	1215	1184	1212
3		1245	1198	1148	1197
4		1240	1205	1162	1203
5		1238	1205	1154	1201
6		1242	1200	1141	1196
Totale complessivo		1240	1206	1163	1204

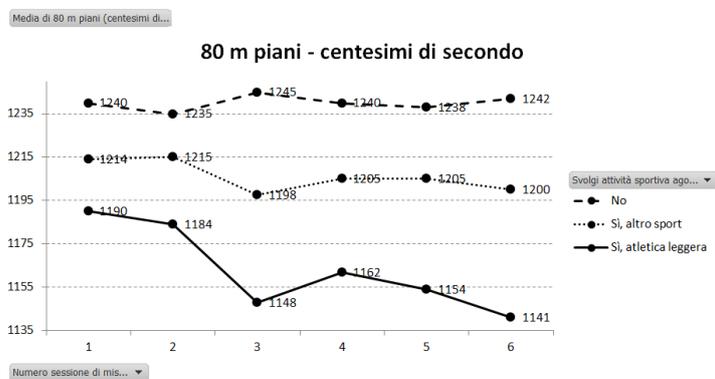


Figura 19: Analisi delle prestazioni medie negli 80 m piani lungo le varie sessioni di misurazione – influenza dell’attività sportiva agonistica extrascolastica.

Se invece si desidera tralasciare la parte dinamica dell’analisi e focalizzarsi sulle medie complessive, relative all’intero periodo in cui si sono svolte le varie sessioni di misurazione, è consigliabile cambiare tipologia di tabella, eliminando la variabile “Numero sessione di misurazione” e sostituendola con la variabile di contesto di interesse. Conviene cambiare anche tipologia di rappresentazione e utilizzare in questo caso un grafico a barre, come illustrato in Figura 20 relativamente alla classe frequentata. Si nota che, nell’esempio qui analizzato, le prestazioni tendono a essere migliori per gli studenti più grandi, con un salto deciso nel passaggio tra il biennio e il triennio. Cambiando variabile nell’area “Etichette di riga” tabella e grafico si aggiornano automaticamente, pertanto sarà possibile ripetere l’analisi utilizzando altre caratteristiche degli studenti.

Le funzionalità dello strumento Tabella Pivot e Grafico Pivot sono svariate ed estremamente flessibili. Pertanto da questo punto in avanti si possono condurre molte altre analisi più approfondite.

È possibile esaminare l’effetto congiunto di più caratteristiche (ad esempio, il fatto di svolgere attività sportiva extrascolastica e il livello di impegno profuso nell’allenamento), semplicemente trascinando più variabili nelle varie aree della Tabella Pivot. Quando si utilizzano più variabili bisogna tenere presente che gli studenti vengono suddivisi in un numero sempre crescente di categorie (date da tutti i possibili incroci delle variabili selezionate), pertanto è possibile che qualche categoria sia composta da un numero esiguo di studenti. In questo caso è necessario assicurarsi preventivamente che il numero di studenti in ogni categoria sia tale da non privare di significato statistico le medie che vengono calcolate.

Inoltre è possibile condurre analisi di correlazione (Montinaro e Nicolini, 2010) tra le prestazioni nelle varie discipline, prendendo così l'occasione per trattare un altro importantissimo tema della Statistica Descrittiva.

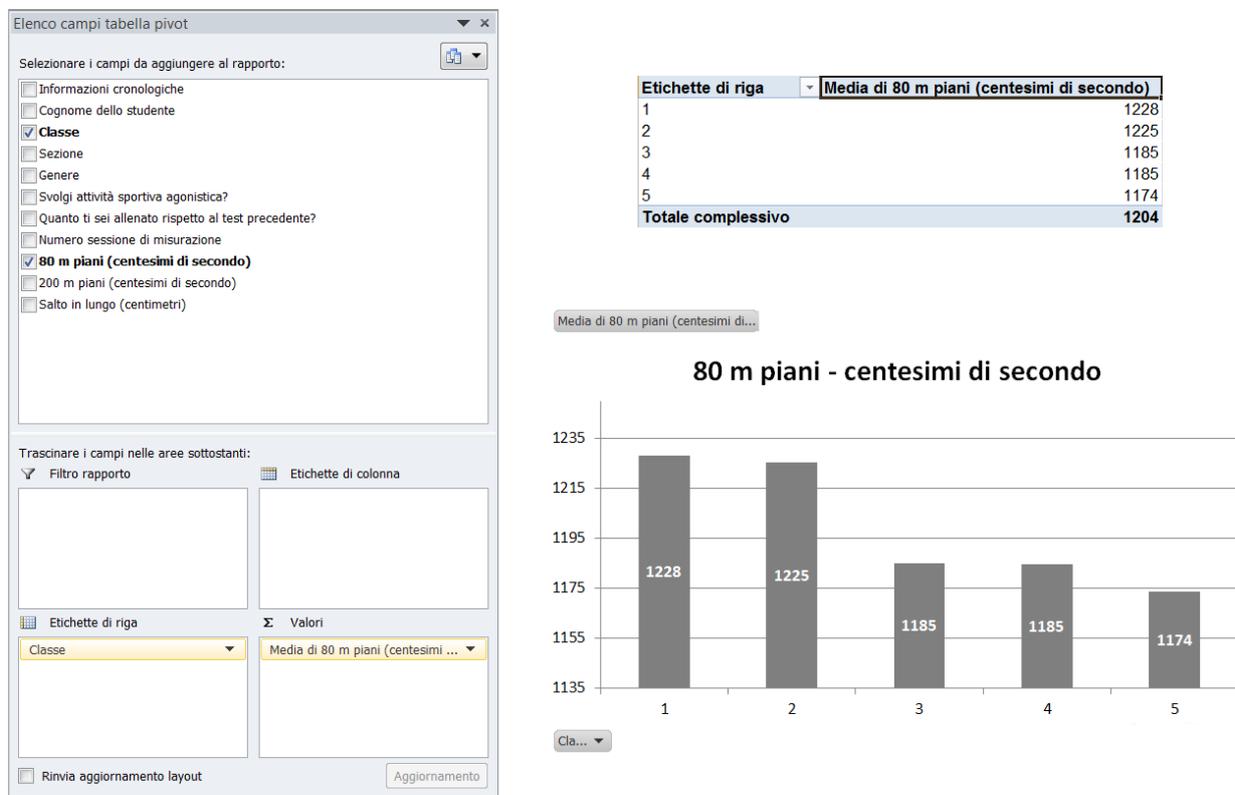


Figura 19: Prestazioni medie negli 80 m piani suddivise per classe frequentata.

CONSIDERAZIONI CONCLUSIVE

In questo articolo sono stati presentati tre progetti che supportano gli insegnanti della scuola secondaria superiore nel trasmettere ai propri studenti, attraverso l'applicazione di tecniche statistiche per l'analisi di dati legati allo Sport, che la Statistica e, in generale, le materie informatico-quantitative posso essere anche divertenti.

Il Progetto 1 mira ad insegnare come sia possibile esaminare ed interpretare i dati e le statistiche sullo sport presenti online, per far scoprire loro l'utilità della statistica nello sport attraverso esempi che li appassionino.

Il Progetto 2 si focalizza sulla pallacanestro ed è finalizzato a insegnare l'uso di strumenti di statistica descrittiva in Excel per l'analisi delle performance di squadre e giocatori, partendo dallo scaricamento dei dati da siti web.

Il Progetto 3 prevede la rilevazione dei dati relativi a prestazioni sportive degli studenti stessi e la successiva analisi, con confronti tra studenti e nel tempo. Il

punto di partenza è la costruzione di moduli on line per la raccolta dei dati, che saranno poi analizzati con diverse finalità mediante tabelle e grafici pivot in Excel.

Gli insegnanti che volessero sviluppare, con i propri studenti, un percorso formativo appassionante e, in taluni casi, multidisciplinare, coinvolgendo insegnanti di matematica, informatica, educazione fisica, possono reperire il materiale didattico di ciascun progetto on line o su richiesta via email (tutte le informazioni utili si trovano al link bodai.unibs.it/BDSports/Students.htm).

BIBLIOGRAFIA

Drazan John F., Loya Amy K., Horne Benjamin D., Eglash Ron, *From Sports to Science: Using Basketball Analytics to Broaden the Appeal of Math and Science Among Youth*, MIT SLOAN Sports Analytics Conference, 2017 (<http://www.sloansportsconference.com/wp-content/uploads/2017/02/1595.pdf>).

Guthrie Donald, *review of the book "Statistica e Sport: non solo numeri" (In Italian) by Antonio Mussino*, Journal of the American Statistical Association, Vol. 94, No. 448, 1999, pp. 1383-1384.

Montinaro Mario, Nicolini Giovanna, *Elementi di Statistica descrittiva*, UTET università, 2010.

Mussino Antonio, *Statistica e Sport: non solo numeri*, Roma: Società Stampa Sportiva, 1997.

Zuccolotto Paola, *Ah, sei uno Statistico! Quindi, in pratica...?*, Il Gatto di Schrödinger, 11, 2016, <http://www.gdsdolomiti.org>.