



**Padova, 20 aprile 2018**

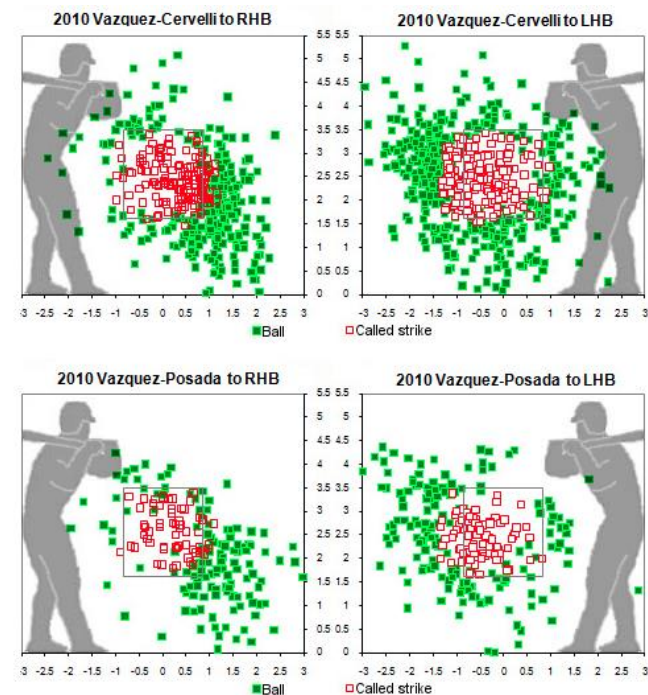
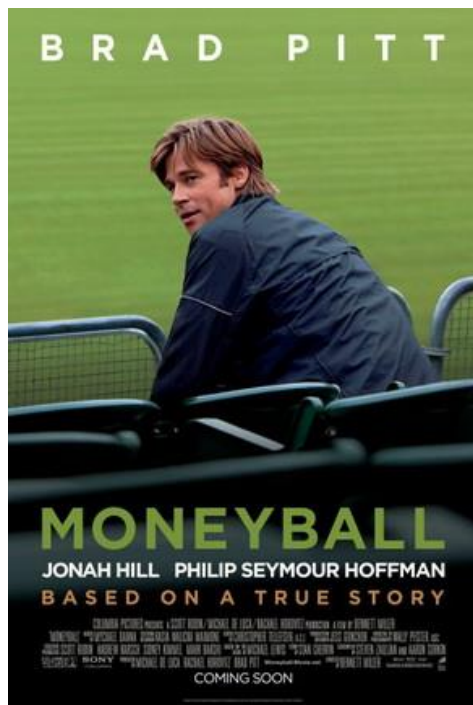
# Nell'era dei Big Data gli Statistici contano.... anche nello Sport

Paola Zuccolotto e Marica Manisera - Big&Open Data Innovation Laboratory (BODaI-Lab)  
Università di Brescia

# Sport e Statistica

*Moneyball*: La storia di Billy Beane e gli esordi della *Sabermetrics*....

<https://www.youtube.com/watch?v=9v7nEpXSc8E>





# Sport e Statistica

Oggi la Statistica viene utilizzata in tutti gli sport.



# Sport e Statistica - calcio

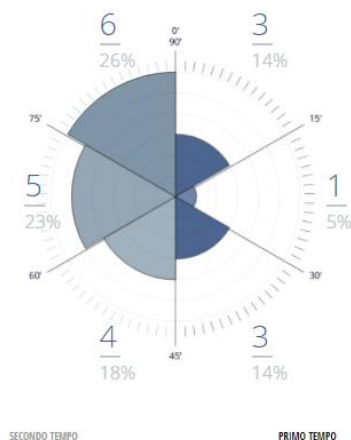
Ai CT delle squadre di calcio è vietato avere PC, tablet o comunicare con gli analisti durante le partite!! <https://www.youtube.com/watch?v=gxEa6B6eW5o>

<http://cartesiani.it/euro-championship-scorers/top/>

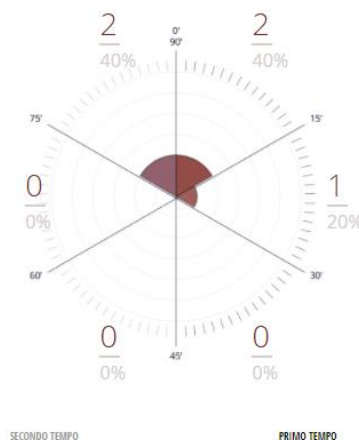


## I tempi dei gol di NAPOLI

### Gol realizzati nei 90'



### Gol subiti nei 90'



CAMPIONATO ORGANIGRAMMA STORIA STAFF TECNICO TEAM **STATISTICHE** STADIO BIGLIETTERIA

STAGIONE 2016/2017

Selezione: Squadra

I TEMPI DEL GOAL

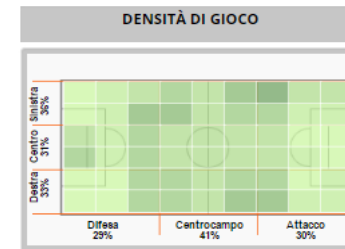


VALORI MEDI DI STAGIONE

|                         | Media  |
|-------------------------|--------|
| Possesso palla          | 23'56" |
| Tiri totali             | 11.9   |
| Tiri dentro             | 4.5    |
| Palle giocate           | 539.1  |
| Passaggi riusciti       | 60.5   |
| Falli commessi          | 17.0   |
| Angoli                  | 5.1    |
| Protezione area         | 54.7   |
| Supremazia territoriale | 10'06" |
| Attacco alla porta      | 47.2   |
| Pericolosità            | 42.9   |
| Giocate utili           | 81.3   |
| Gol Fatti               | 1.1    |

POSIZIONE IN CLASSIFICA

| POSIZIONE | NOME SQUADRA | PUNTI |
|-----------|--------------|-------|
| 10        | Ascoli       | 33    |
| 11        | Carpi        | 33    |
| 12        | Novara       | 31    |
| 13        | Salernitana  | 31    |
| 14        | Avellino     | 29    |
| 15        | Cesena       | 28    |
| 16        | Brescia      | 28    |



Data powered by **PANINI DIGITAL**



# Sport e Statistica - tennis

[http://www.wimbledon.com/en\\_GB/players/overview/wta310761.html](http://www.wimbledon.com/en_GB/players/overview/wta310761.html)





**FIRST ROUND**

**COURT 14 • LADIES' SINGLES • COMPLETE**

|                      | PTS | 1 | 2 | 3 | 4 | 5 |
|----------------------|-----|---|---|---|---|---|
| S. Errani (ITA) [20] | ✓   | 6 | 6 |   |   |   |
| P. Tig (ROU)         |     | 4 | 4 |   |   |   |

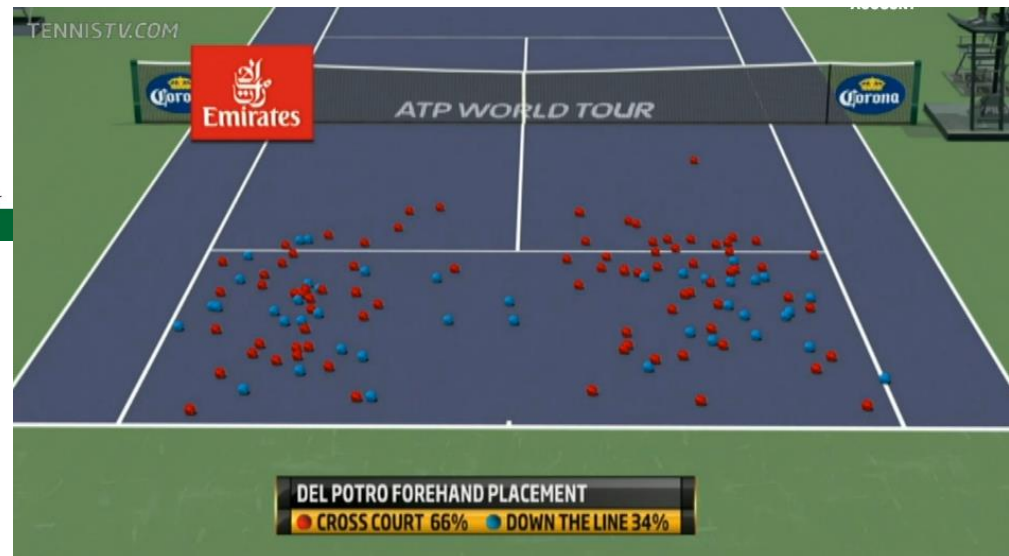
**MATCH DURATION: 1:31**



| Sara Errani  | MATCH STATISTICS     | Patricia Maria Tig |
|--------------|----------------------|--------------------|
| 0            | ACES                 | 3                  |
| 1            | DOUBLE FAULTS        | 7                  |
| 56/67 (84 %) | 1ST SERVES IN        | 38/66 (58 %)       |
| 34/56 (61 %) | 1ST SERVE POINTS WON | 27/38 (71 %)       |
| 7/11 (64 %)  | 2ND SERVE POINTS WON | 9/28 (32 %)        |
| 8/14 (57 %)  | NET POINTS WON       | 11/15 (73 %)       |
| 4/9 (44 %)   | BREAK POINTS WON     | 2/10 (20 %)        |
| 30/66 (45 %) | RECEIVING POINTS WON | 26/67 (39 %)       |
| 12           | WINNERS              | 31                 |
| 14           | UNFORCED ERRORS      | 37                 |
| 71           | TOTAL POINTS WON     | 62                 |



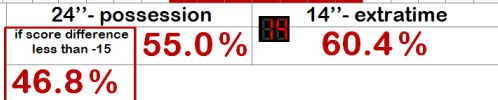
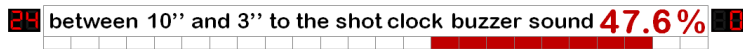
Sara Errani: First round



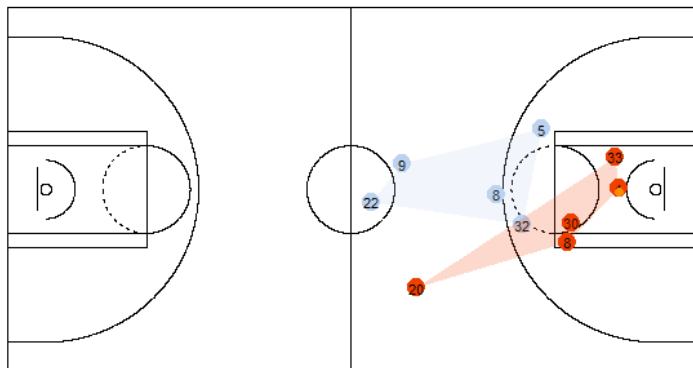
# Sport e Statistica - basket

[http://www.fiba.com/olympics/2016/1708/USA-Argentina#tab=shot\\_chart](http://www.fiba.com/olympics/2016/1708/USA-Argentina#tab=shot_chart)

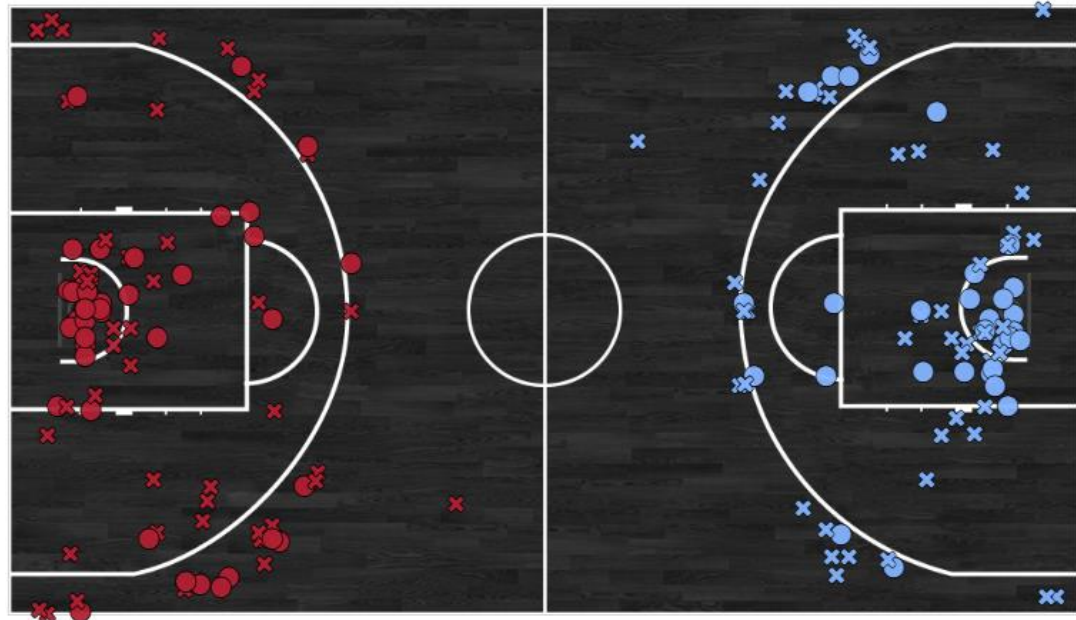
## 2-Point Field Goal Percentages



Analysis carried out on 33682 2-point field goal attempts - 2015/2016 Italian A2 Championship

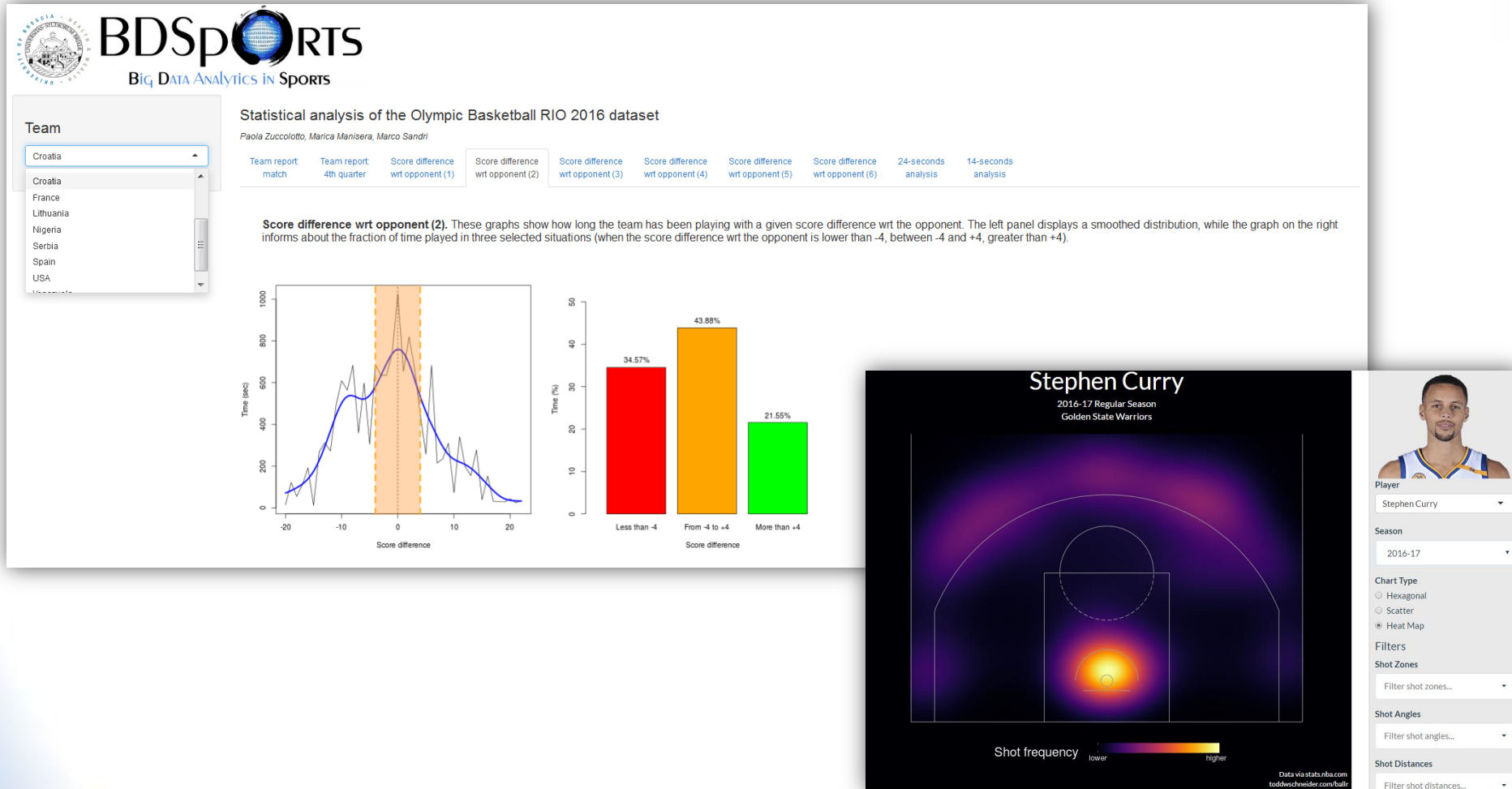


Share this



# Sport e Statistica - basket

<http://bodai.unibs.it/BDSports/basketball.htm>



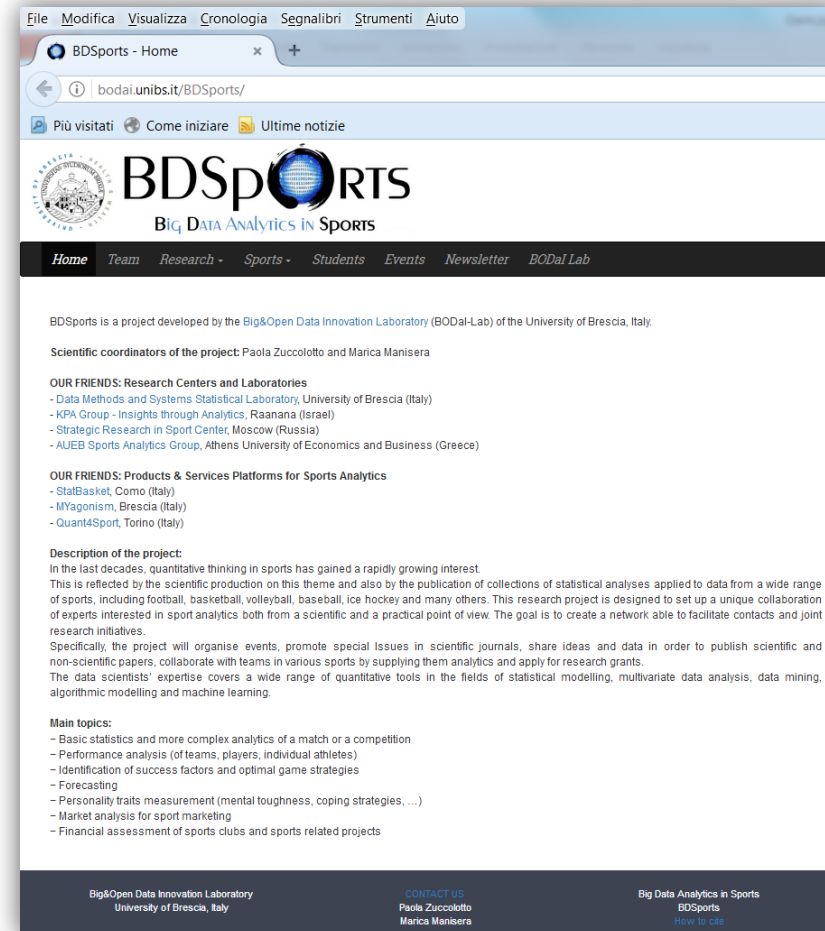
# BODal-Lab e BDSports



In questo quadro è nato il progetto internazionale  
**Big Data Analytics in Sports (BDsports)**

[bodai.unibs.it/bdsports](http://bodai.unibs.it/bdsports)

nell'ambito delle attività del  
**Big&Open Data Innovation Laboratory**  
dell'Università di Brescia



Paola Zuccolotto e Marica Manisera - Big&Open Data Innovation Laboratory





# Il Team di BDSports

Più di 50 partecipanti provenienti da Italia, Germania, Paesi Bassi, Belgio, Russia, Stati Uniti.

Suddivisi per competenze in 4 sezioni.

### DATA SCIENCE

Conoscenza ed implementazione informatica di strumenti quantitativi nel campo dei modelli statistici, dell'analisi multivariata, del data mining, degli algoritmi di intelligenza artificiale e apprendimento automatico, con particolare – ma non esclusivo - riferimento all'analisi dei big data.

### TECHNICAL EXPERTS & COACHING

Proposta di tematiche di ricerca, richiesta di sviluppo di analisi per necessità specifiche, supporto nell'implementazione dei modelli che richiedono conoscenza tecnica dell'ambiente di applicazione, interpretazione dei risultati e feed-back.

### SPORTS ANALYTICS SERVICES & DATA MANAGEMENT

Condivisione di idee di lavoro e dati, implementazione su supporti user-friendly e standardizzazione dei modelli statistici sviluppati.

### SPORT MANAGEMENT, MARKETING AND FINANCE

Analisi di mercato, strategie di management, valutazioni finanziarie.

# Le 4 anime di BDSports

Opera in 4 canali di azione.

## RICERCA SCIENTIFICA

Pubblicazione di articoli su riviste scientifiche internazionali; organizzazione di workshop, seminari, sessioni a conferenze; guest editing di numeri speciali di riviste dedicati al tema delle Sport Analytics; partecipazione a comitati scientifici di riviste; predisposizione e condivisione database e software specifici (pagina «Research» su [bodai.unibs.it/BDSports](http://bodai.unibs.it/BDSports)).

## APPLICAZIONE SUL CAMPO

Confronto con specialisti del settore per individuare le linee di ricerca; implementazione software user-friendly; utilizzo modelli su dati reali (pagina «Sports» su [bodai.unibs.it/BDSports](http://bodai.unibs.it/BDSports)).

## DIDATTICA

Predisposizione progetti didattici per la scuola e l'università; supervisione di tesi di laurea, master e dottorato; tutoraggio di stage presso aziende (pagina «Students» su [bodai.unibs.it/BDSports](http://bodai.unibs.it/BDSports)).

## DIVULGAZIONE

Partecipazione a eventi per il pubblico (pagina «Events» su [bodai.unibs.it/BDSports](http://bodai.unibs.it/BDSports)); articoli su giornali e riviste orientate alla divulgazione .

# Dati



## Dati rilevati manualmente

event-log, statistiche, resoconti di sessioni di allenamento, valutazioni di esperti, ...



## Dati rilevati mediante dispositivi tecnologici

GPS, sensori, videocamere, pedane, tecnologie wearable, ...



## Dati da questionari

questionari psicometrici per la valutazione di tratti latenti della personalità



## Altri dati

analisi di mercato, dati di social networks, Google trends, ...



# Domande



## DATA SCIENCE:

Statistica

Analisi dei Dati

Data Mining

Machine Learning

Artificial Intelligence

# Interpretazione dei risultati



- Analisi di una gara, un match, un torneo, un campionato
- Analisi della Performance (di squadre, giocatori, atleti individuali)
- Identificazione dei fattori di successo e delle strategie di gioco ottimali
- Previsione
- Misurazione dei tratti della personalità (*mental toughness, coping strategies, ...*)
- Analisi di mercato per il marketing management sportivo
- Analisi finanziaria



# Sports Analytics – i dati

## Dati

**Dati rilevati manualmente**  
event-log, statistiche, resoconti di sessioni di allenamento, valutazioni di esperti, ...

**Dati rilevati mediante dispositivi tecnologici**  
GPS, sensori, videocamere, pedane, tecnologie wearable, ...

**Dati da questionari**  
questionari psicometrici per la valutazione di tratti latenti della personalità

**Altri dati**  
analisi di mercato, dati di social networks, Google trends, ...

**Stats**

**event-log  
(play-by-play)**

**tecnologie  
wearable**

**Psychometrics**

**sensor data  
GPS**

**Market Analysis**

**Sentiment Analysis**

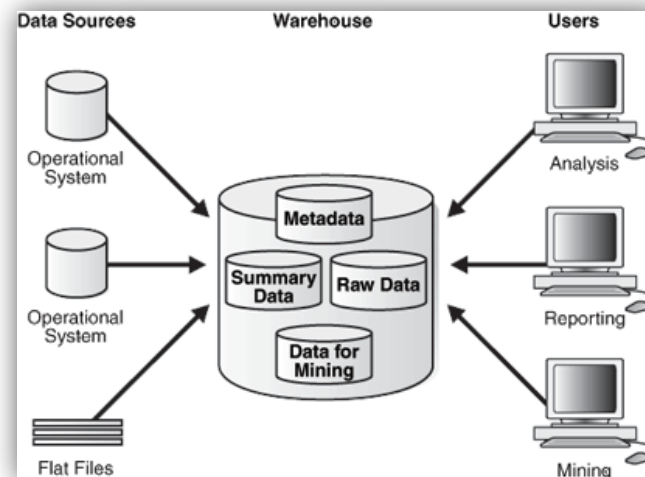
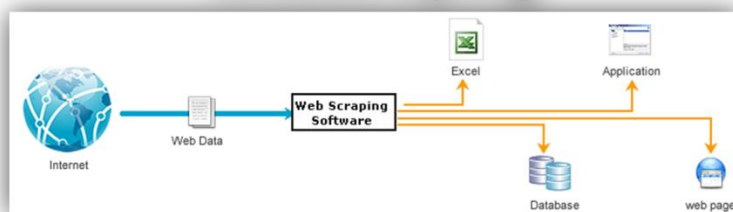
**Data**

**Big Data**

# Sports Analytics – i dati



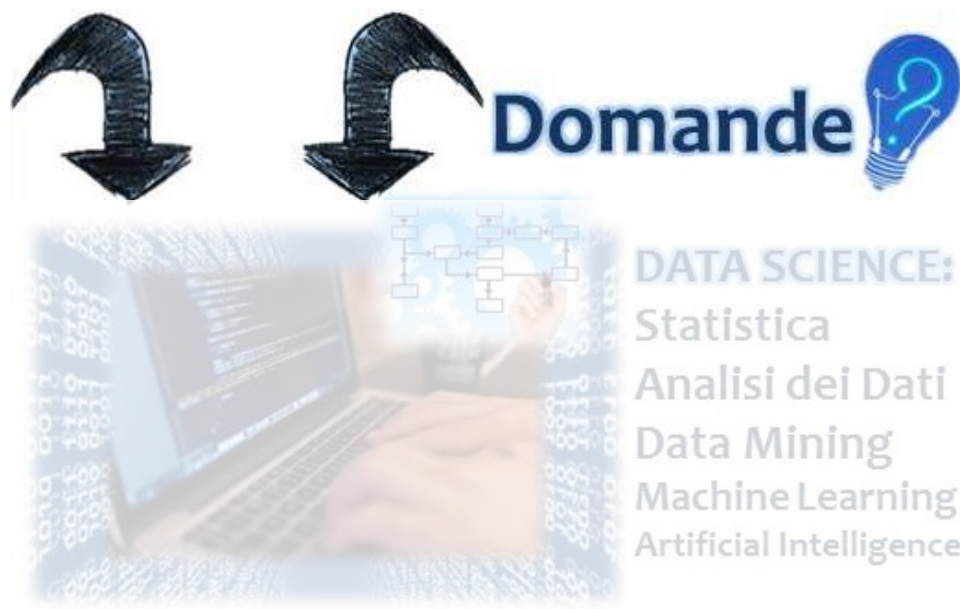
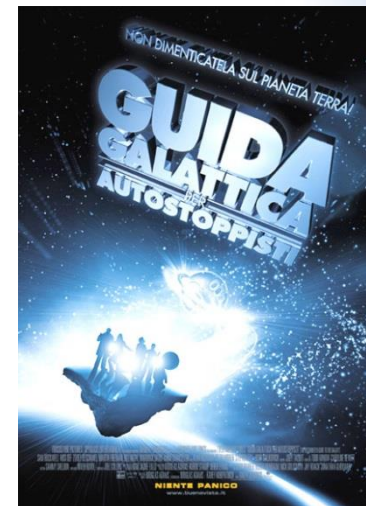
## Webscrapping



# Sports Analytics – le domande

«Quarantadue!» urlò Loonquawl. «Questo è tutto ciò che sai dire dopo un lavoro di sette milioni e mezzo di anni?»

«Ho controllato molto approfonditamente, » disse il computer, «e questa è sicuramente la risposta. Ad essere sinceri, penso che il problema sia che voi non abbiate mai saputo veramente qual è la domanda. » (Douglas Adams, Guida galattica per gli autostoppisti)

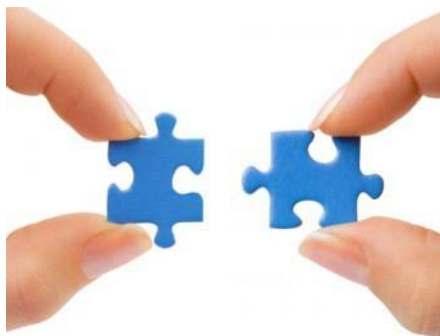




# Sports Analytics – le domande



# Sports Analytics – le domande



# Sports Analytics – le domande



Misurare l'abilità di un giocatore di tennis

ESEMPIO



# Sports Analytics – le domande

Misurare l'abilità di un giocatore di tennis



## DEFINIRE ABILITÀ

- esecuzione delle varie tipologie di colpi (servizio, dritto, rovescio, ...);
- modo in cui essi vengono effettuati (forza, precisione, ...);
- scelta di come e dove tirare in relazione al momento di gioco e all'avversario;
- ...;
- capacità di rimanere lucidi e freddi nei momenti cruciali;
- capacità di tenere costante il livello anche quando il match è lungo e la stanchezza avanza

# Sports Analytics – le domande

**Misurare l'abilità di un giocatore di tennis**



## **DEFINIRE ABILITÀ**

- esecuzione delle varie tipologie di colpi (servizio, dritto, rovescio, ...);
- modo in cui essi vengono effettuati (forza, precisione, ...);
- scelta di come e dove tirare in relazione al momento di gioco e all'avversario;
- ...;
- capacità di rimanere lucidi e freddi nei momenti cruciali;
- capacità di tenere costante il livello anche quando il match è lungo e la stanchezza avanza



**Scegliere aspetti su cui concentrare attenzione**  
**Stabilire quali dati li descrivono**

# Sports Analytics – le domande

## Misurare l'abilità di un giocatore di tennis



### DEFINIRE ABILITÀ

- esecuzione delle varie tipologie di colpi (servizio, dritto, rovescio, ...);
- modo in cui essi vengono effettuati (forza, precisione, ...);
- scelta di come e dove tirare in relazione al momento di gioco e all'avversario;
- ...;
- capacità di rimanere lucidi e freddi nei momenti cruciali;
- capacità di tenere costante il livello anche quando il match è lungo e la stanchezza avanza



**Scegliere aspetti su cui concentrare attenzione**  
**Stabilire quali dati li descrivono**



**DOMANDA:** valutare il trade-off tra forza e precisione dei tiri - separatamente per servizio, dritto e rovescio - misurando l'efficacia dei tiri in termini di difficoltà causata all'avversario, quest'ultima valutata in una scala da 1 a 5 da un team di tecnici che effettueranno l'analisi video di un dato numero di match giocati contro lo stesso avversario



# Sports Analytics – DATA SCIENCE



**Statistica tradizionale**

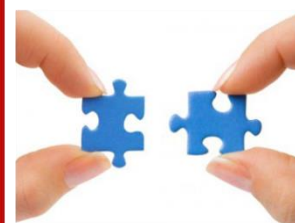
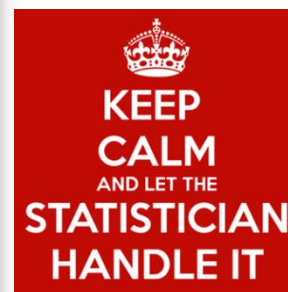


**Statistica moderna**

Machine learning  
Algorithmic modelling  
Data Mining



# Sports Analytics – interpretazione dei risultati



- Analisi di una gara, un match, un torneo, un campionato
- Analisi della Performance (di squadre, giocatori, atleti individuali)
- Identificazione dei fattori di successo e delle strategie di gioco ottimali
- Previsione
- Misurazione dei tratti della personalità (*mental toughness, coping strategies, ...*)
- Analisi di mercato per il marketing management sportivo
- Analisi finanziaria

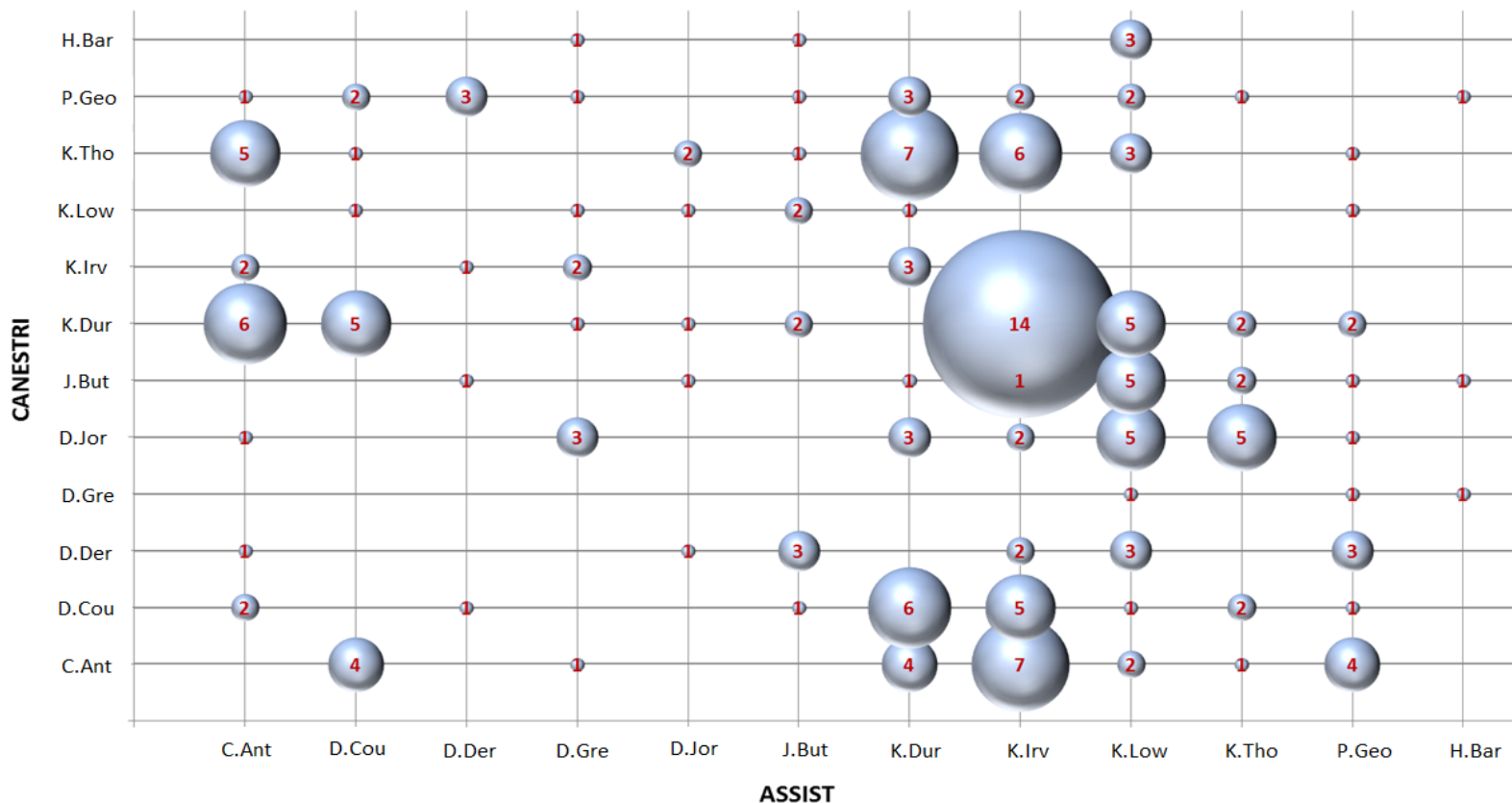
# Sports Analytics – esempi



Dataset contenente tutti i singoli momenti di gioco



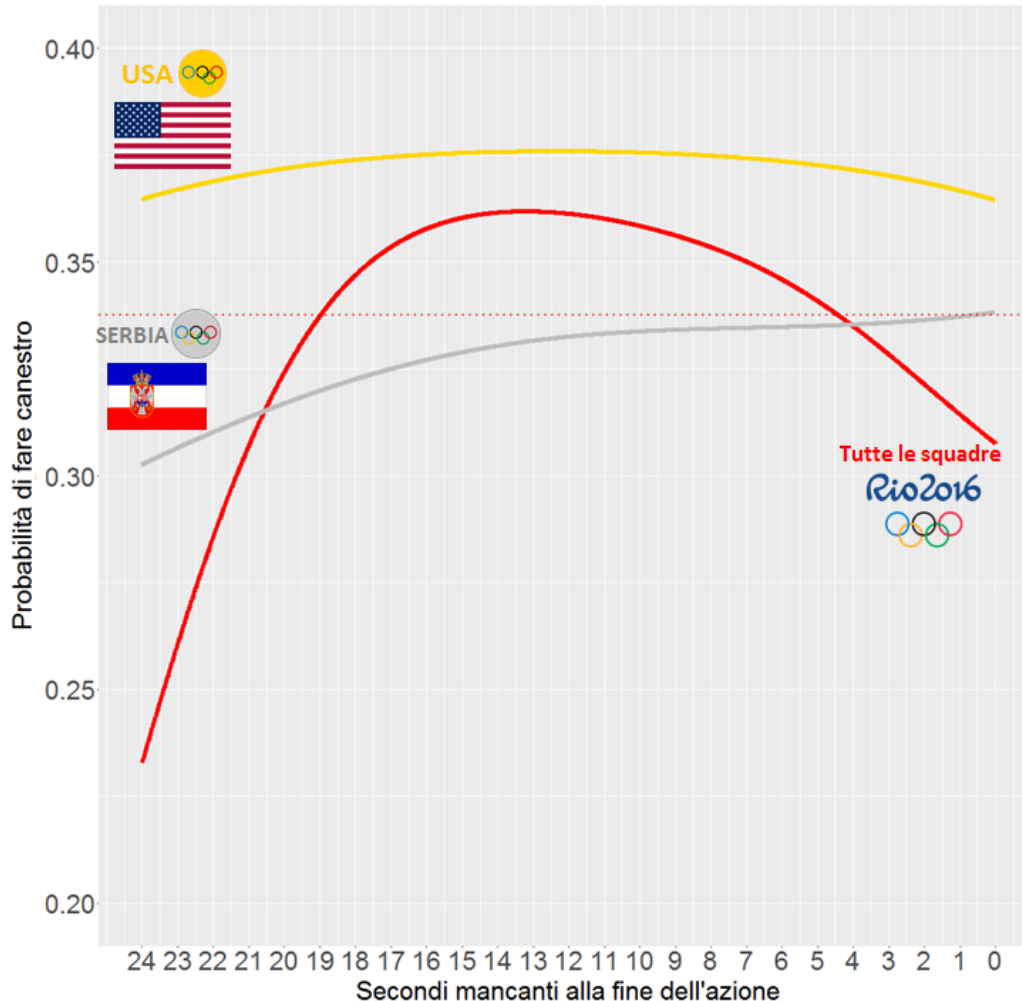
Grafico ASSIST-CANESTRI





# Sports Analytics – esempi

Analisi della probabilità di fare canestro rispetto ai secondi mancanti alla fine dell'azione - tiri da 3 punti



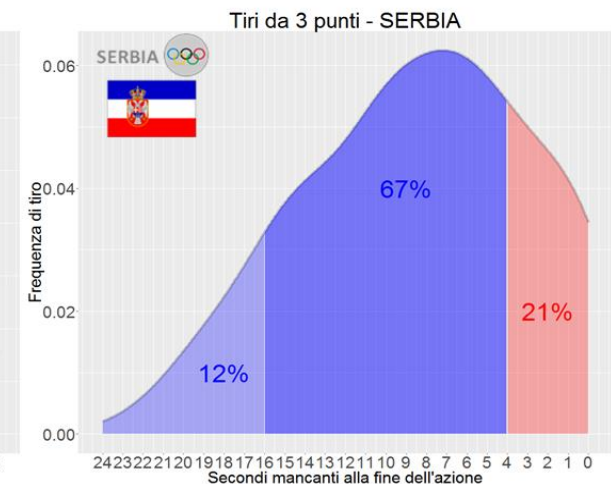
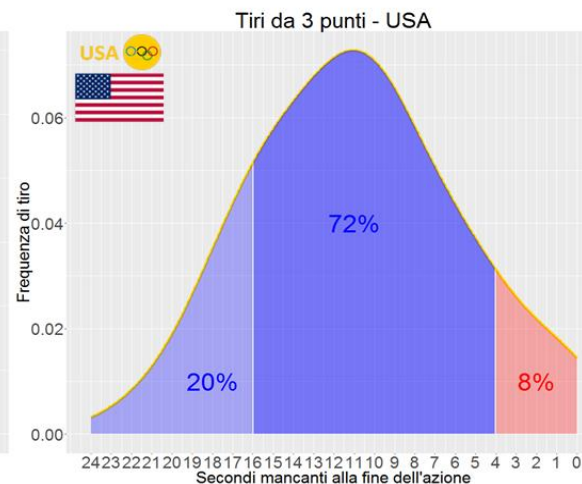
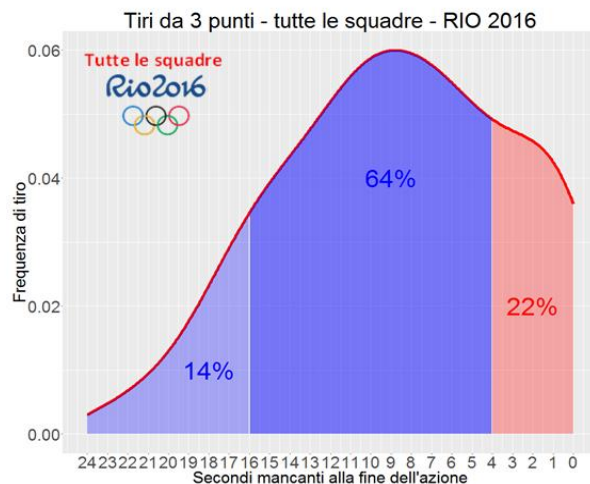
Dataset contenente tutti i singoli momenti di gioco (più di 6000 tiri a canestro!)

# Sports Analytics – esempi

Dataset contenente tutti i singoli momenti di gioco (più di 6000 tiri a canestro!)



## Analisi della frequenza di tiro rispetto ai secondi mancanti alla fine dell'azione - tiri da 3 punti

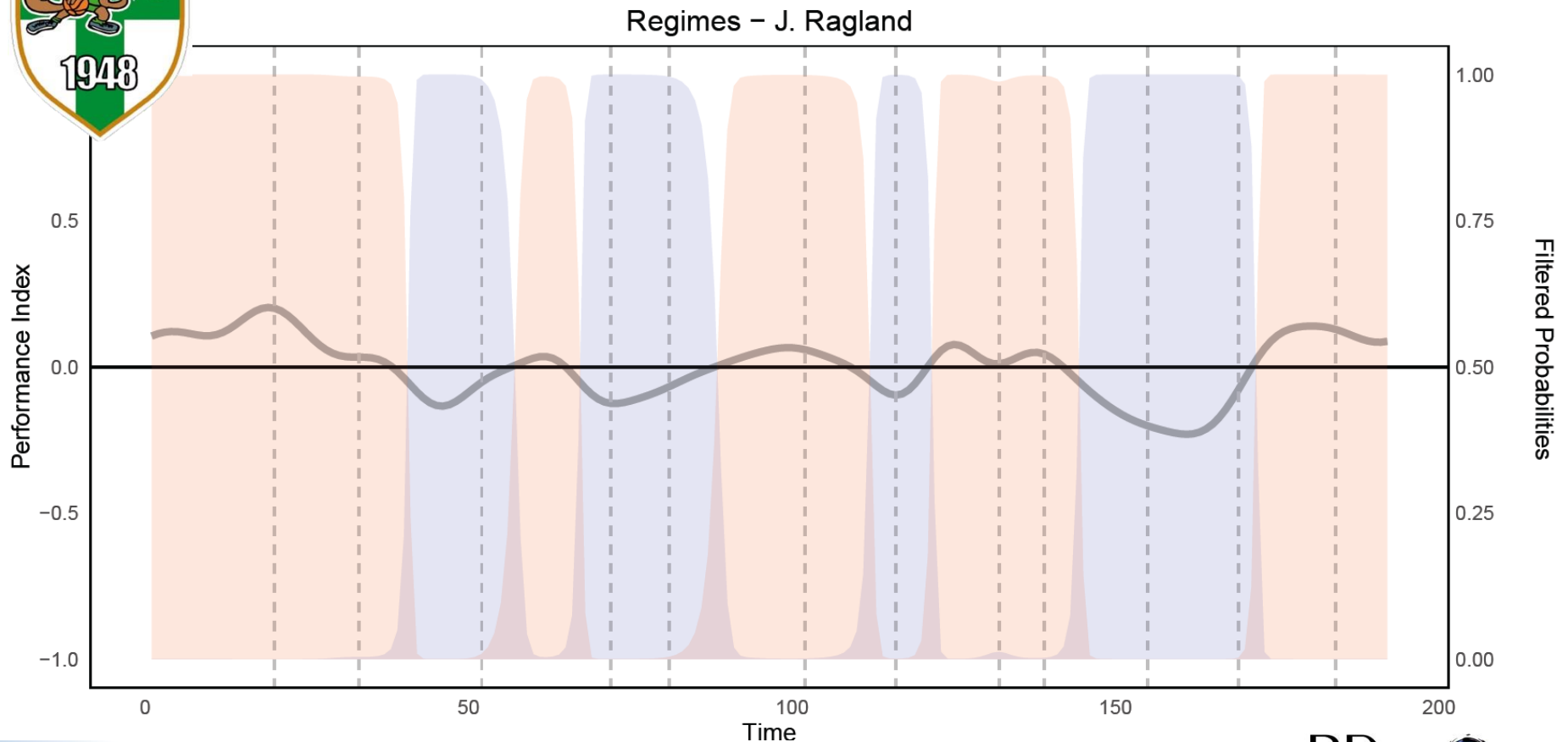


# Sports Analytics – esempi

Dataset contenente tutti i singoli momenti di gioco (più di 300 partite!!)



BASKETBALL  
CHAMPIONS  
LEAGUE



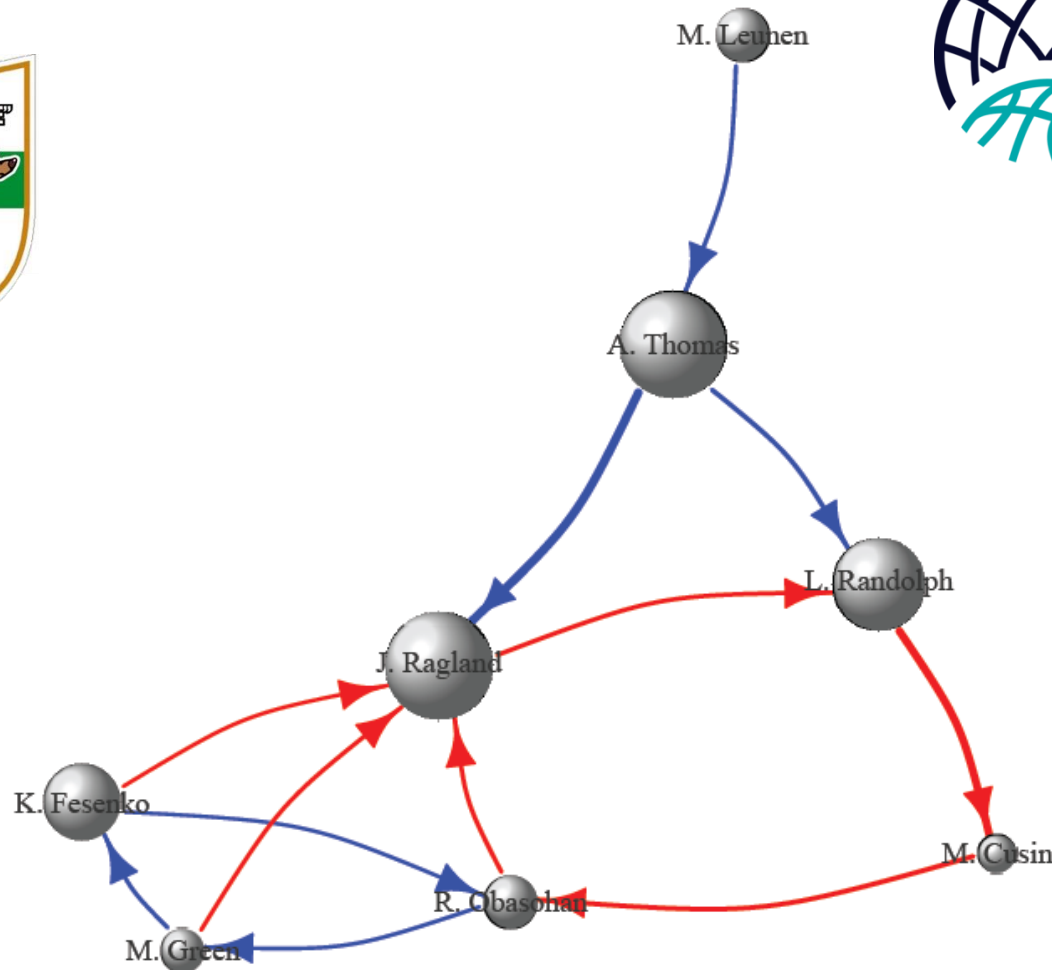


# Sports Analytics – esempi



BASKETBALL  
CHAMPIONS  
LEAGUE

Dataset contenente tutti i  
singoli momenti di gioco (più  
di 300 partite!!)



# Sports Analytics – esempi

**Strategia di gioco:** fattori che influenzano la probabilità di vittoria

Dataset contenente tutti i singoli momenti di gioco (più di 800 variabili!!) dal 2008 al 2012



- shot.attack.home ↑
- aerial.attack.home ↓
- defense.home ↓



- midfield-defense.counterattack.away ↑
- shot.attack.away ↑
- area-defense.away ↓

**Molti altri esempi nella sezione Sports del nostro sito web!**

**Grazie per l'attenzione**

