



UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DIPARTIMENTO DI ECONOMIA E MANAGEMENT

Corso di Laurea
in Economia e Gestione Delle Imprese

Relazione Finale

Sport e Data Analytics:
Indicatori di Performance nel Tennis

Relatore: Chiar.ma Prof. Zuccolotto Paola

Laureando: Persico Giulio

Matricola n. 711803

Anno Accademico 2019/2020

INDICE

INTRODUZIONE.....	4
1. SPORT E DATI: UN CONNUBIO VINCENTE.....	5
1.1 LA RILEVANZA ECONOMICA DEL SETTORE SPORTIVO.....	5
1.2 DATA ANALISI E INDUSTRIA SPORTIVA.....	7
1.3 DATA ANALYST E DATA SCIENTIST NELLO SPORT.....	9
1.4 PERFORMANCE ANALYSIS NELLO SPORT.....	10
2 IL CASO DEL TENNIS: UN POTENZIALE ANCORA INESPRESSO.....	12
2.1 TENNIS E DATI: UNO SGUARDO GENERALE.....	12
2.2 CLASSIFICAZIONE DELLE TIPOLOGIE DI DATI RACCOLTI NEL TENNIS.....	14
2.3 FONDAMENTI DI TENNIS DATA ANALYTICS.....	19
3 TENNIS DATA ANALYTICS: IL SERVIZIO.....	21
3.1 FATTORI CRITICI DI SUCCESSO NEL TENNIS.....	21
3.2 ANALISI DI UN FATTORE CRITICO: IL SERVIZIO.....	24
3.2.A. QUANTIFICARE L'IMPATTO DEL SERVIZIO NEL TENNIS.....	26
3.2.B INDICATORI DI PERFORMANCE E DI RATING IN USO ATTUALMENTE.....	28
3.2.C SCOMPORRE IL SERVIZIO IN FATTORI: ANALISI DEL COLPO A PARTIRE DAI DATI ANALITICI DI HAWK-EYE.....	31
3.2.D UN INDICATORE DIFFERENTE.....	39

3.2.E	CARATTERISTICHE FISICHE DEL GIOCATORE: LA STATURA.....	41
3.3.	LA STRATEGIA DELLE DUE PRIME.....	44
4.	TENNIS DATA ANALYTICS: IL PUNTEGGIO.....	49
4.1	PREMESSE.....	49
4.2	PRESENTAZIONE DELLE ANALISI EFFETTUATE.....	51
4.2.A	MATCH 2 SU 3 – ESITO DELLA PARTITA.....	52
4.2.B	MATCH 2 SU 3 – ANDAMENTO DELLA PARTITA.....	53
4.2.C	MATCH 3 SU 5.....	63
4.2.D	IL PUNTEGGIO DEI SET.....	65
4.3	CONSIDERAZIONI FINALI SUL PUNTEGGIO.....	70
5.	CONCLUSIONI.....	71
6.	RIFERIMENTI.....	74

INTRODUZIONE

Questo elaborato ha l'obiettivo di presentare, spiegare ed esemplificare le dinamiche che stanno alla base dei processi di analisi statistica applicati al mondo dello sport.

Il settore sportivo, oltre ad avere una certa importanza dal punto di vista economico, negli ultimi anni ha saputo modernizzarsi tramite l'introduzione dell'utilizzo di processi di Data Analytics per migliorare le prestazioni degli atleti e delle squadre e proprio questo particolare accostamento tra sport e statistica sarà l'oggetto di un iniziale approfondimento.

A seguire verrà offerta una panoramica delle tecniche di raccolta ed elaborazione dati di una singola disciplina sportiva, il Tennis, rispetto al quale credo di aver maturato le competenze necessarie per affrontare con cognizione di causa le varie tematiche proposte e rielaborarle con occhio critico.

Viene spontaneo durante l'esposizione offrire poi anche analogie con le tematiche trattate in alcuni corsi del percorso di economia e gestione delle imprese.

Infatti verranno fornite interpretazioni sia oggettive sia personali delle statistiche prese in esame ed inoltre saranno avanzate proposte per il miglioramento della qualità dell'elaborazione dei dati disponibili.

Infine sarà presentato, in appendice all'elaborato corrispondente alla relazione finale vera e propria, un estratto del lavoro svolto in qualità di progetto sostitutivo dello tirocinio curricolare, che ha previsto la costruzione di un database incentrato sulla Tennis Data Analytics, argomento che rappresenta il filo conduttore dell'intero lavoro, e la successiva analisi delle informazioni raccolte.

1. SPORT E DATI: UN CONNUBIO VINCENTE

In questo primo capitolo verrà innanzitutto raccontato il settore sportivo, argomento teatro degli approfondimenti di questo elaborato, dal punto di vista economico.

Sarà in seguito introdotta la tematica riguardante l'applicazione e l'utilizzo dei processi di Data Analytics nello sport, per mezzo dei quali, grazie al lavoro di figure professionali specializzate e alle nuove tecnologie, è possibile effettuare elaborazioni di cruciale importanza per il raggiungimento del successo competitivo nello sport.

1.1 LA RILEVANZA ECONOMICA DEL SETTORE SPORTIVO

La consuetudine vuole che, quando viene nominata la parola "sport", nell'immaginario collettivo avvenga spontaneamente l'associazione mentale con tematiche di tipo ludico-ricreativo, ma di scarsa rilevanza nell'economia reale.

La verità però è che lo sport, oltre al fatto oramai assodato di ricoprire un ruolo determinante nel creare benefici sociali e contribuire alla funzione formativa dei più giovani, si pone anche concretamente al centro dello scenario economico attuale, in particolar modo nei paesi più sviluppati¹.

Il settore dell'industria sportiva, ovvero quella che comprende tutti gli esercizi delle attività correlate alla pratica sportiva, dalla vendita di articoli sportivi ai servizi collegati alla pratica dell'attività in senso stretto, rappresenta infatti un settore florido, con grandi prospettive di crescita e che a livello numerico ricopre tutt'ora un ruolo di primaria rilevanza².

^{1,2} Il libro bianco dello sport Italiano, CONI, 2020, coni.it

Le imprese dello sport: i numeri del settore



Fig. 1 "I numeri del settore sportivo in Italia".

Fonte: "Sport e imprese: Work Together" (Comitato Leonardo, 2018)

Il settore economico dello sport rappresenta circa l'1% del Prodotto Interno Lordo italiano, e si colloca oggi tra le prime dieci industrie del nostro Paese a livello di fatturato. Come illustrato in Figura 1, tutto lo sport riesce a produrre complessivamente un valore 17,5 di miliardi e mezzo di euro nell'ambito del nostro Paese. Le persone che lavorano nel mondo dello sport, sia a titolo professionale, sia a titolo dilettantistico, sono oggi circa 118 mila.

Ciò dà chiaramente l'idea di quanto il fenomeno sportivo sia importante economicamente. A livello globale invece, secondo i dati riportati dal Sole 24 Ore³, nel 2018 lo sport ha prodotto un fatturato pari a 130 miliardi (imputabili in larga parte al commercio dei diritti tv, alle sponsorizzazioni, agli ingressi negli stadi/palazzetti e alla vendita del "merchandising") di euro, corrispondente al 21% del fatturato totale del settore dell'intrattenimento (616 miliardi, +34% rispetto al 2014), mercato in cui può essere inquadrato il settore sportivo.

Ritornando invece nel nostro paese e guardando all'interno di questo settore, il report "Sport e imprese: work together" (Comitato Leonardo, 2018) rileva che il comparto correlato allo sport più sviluppato è quello del commercio, che rappresenta poco meno della metà del valore della produzione e del numero di addetti. Il secondo è quello manifatturiero, che fa riferimento ai produttori di articoli sportivi, con circa 2.300 imprese e quasi 19 mila addetti, che generano 5,3 miliardi di euro di fatturato. Segue come terzo comparto quello dei servizi

³ "Industria dell'entertainment: solo lo sport fattura 130 miliardi" (M.Frisone, www.ilsole24ore.com, 2019)

sportivi, che incide per il 21% sul valore della produzione ma sfiora il 37% in termini occupazionali.

Parlando invece in termini di praticanti, negli ultimi anni, anche la diffusione dell'attività sportiva è notevolmente aumentata e tocca ormai più del 60% della popolazione (dati ISTAT, 2015) e le recenti stime, hanno calcolato che sono circa 20 milioni⁴ coloro che praticano attività sportiva continuativamente, infine le società sportive organizzate sono circa novantamila. Quello sportivo può così essere considerato il sistema più capillare del nostro Paese: esiste un punto sport ogni 604 abitanti, con un rapporto superiore a quello di scuole, parrocchie, edicole, bar, ristoranti⁵.

1.2 DATA ANALISI E INDUSTRIA SPORTIVA

Come anticipato, l'obiettivo principale di questo elaborato è illustrare ed esemplificare i processi di Data Analytics applicati al mondo dello sport: il ricorso alle tecniche più avanzate di analisi dei dati nell'attività di coaching può infatti essere considerato la più moderna frontiera del legame tra scienza e sport, il cui scopo è quello di contribuire ad alzare sempre più in alto l'asticella delle prestazioni e al contempo ridurre l'incertezza e la casualità delle dinamiche che gli atleti affrontano durante le competizioni sportive.

L'applicazione dei Big Data allo sport si basa sull'adozione di un preciso modello, denominato "data-driven", che si fonda sull'elaborare strategie a più livelli partendo dall'analisi dei dati e delle informazioni provenienti dalle più disparate fonti, permettendo di pianificare, di monitorare gli obiettivi e di massimizzare il rendimento delle organizzazioni sportive e degli atleti professionisti⁶.

La definizione di Data-driven rimanda quindi al concetto di "evidence-based decisions" (Gil Fried, 2017), ovvero decisioni effettuate grazie a evidenze tangibili, e descrive perciò contesti

⁴ "La pratica sportiva in Italia", ISTAT, 2015, istat.it

⁵ Tavolo Nazionale per la Governance nello Sport. Piano Nazionale per la Promozione dell'Attività Sportiva. 1° edizione. Roma, 2012, archivi.istruzione.it

⁶ Capgemini Research Institute, www.capgemini.com

in cui le scelte vengono basate su fatti oggettivi, e non su sensazioni personali o luoghi comuni, tramite l'utilizzo sistematico della tecnologia digitale e diffondendo la cultura del dato a tutti i livelli dell'organizzazione.

Il ricorso, sempre più frequente, ai Big Data nello sport, consente perciò di sviluppare modelli di allenamento estremamente personalizzati, che tengono quindi conto delle caratteristiche fisiche e tecniche proprie di ogni atleta, nonché di realizzare modelli che correlino i risultati della prestazione alle differenti strategie utilizzabili.

L'obiettivo perciò è duplice, poiché lo scopo è quello di ottenere sia un vantaggio prestazionale, ovvero un miglioramento dell'atleta, sia un vantaggio di tipo competitivo, che riguarda il miglioramento nell'efficacia nel confronto con i competitors del singolo o del collettivo.

La crescita esponenziale delle attività correlate all'industria sportiva testimonia il progressivo avvicinamento della tecnologia allo sport, che si manifesta in modalità differenti e con benefici per tutti gli attori della "filiera": atleti, squadre, tifosi e professionisti del settore.

Per quanto riguarda questa innovazione, il baseball può essere considerato probabilmente il primo sport a introdurre l'analisi di big data come fondamento del processo decisionale⁷: ogni scelta, dal reclutamento dei giocatori al marketing per il coinvolgimento dei tifosi viene effettuata secondo un metodo scientifico.

In molti sport le squadre dispongono di staff dedicati alla Data Analytics, che utilizza i risultati della Match Analysis per effettuare scelte sulla composizione della formazione e della rosa. Inoltre si fa sempre più affidamento sull'analisi dei dati di allenamento per ottimizzare le prestazioni degli atleti, sfruttando il fatto che le tecnologie di aggregazione dei dati e di monitoraggio stanno diventando sempre più efficienti. Tramite questi processi si mira ad ottenere un miglioramento della qualità dei programmi di preparazione atletica adottati sia durante sia al termine della stagione⁸.

L'efficienza delle tecniche di monitoraggio è dovuta al fatto che le prestazioni degli atleti sono registrate e analizzate attraverso telecamere, sensori e prodotti indossabili che tengono conto di ogni aspetto dello sport di riferimento. Queste tecnologie oggi sono in grado di fornire

^{7,8} LA CRESCENTE IMPORTANZA DELLA FOOTBALL ANALYTICS, Soccerment Research, soccerment.com

⁸ "Migliorare le prestazioni sportive, fotogramma per fotogramma", Technogym.it

anche i più piccoli dettagli riguardo le prestazioni del giocatore, segnalando allo staff tecnico non solo le informazioni essenziali ai fini dell'analisi della performance dell'atleta o all'analisi strategica di un match, ma anche quelle riguardanti lo stato fisico dell'atleta.

Negli ultimi anni, l'esponenziale velocità di miglioramento delle tecnologie a supporto della raccolta, archiviazione e analisi dei dati è andata di pari passo con un aumento esponenziale del capitale umano investito nella Sport Data Analytics, causando un netto miglioramento della qualità di questi processi, e determinando la diffusione di nuove posizioni professionali.

1.3 DATA ANALYST E DATA SCIENTIST NELLO SPORT

In questa nuova frontiera del connubio ormai consolidato fra scienza e sport si sono create in particolare due nuove (soprattutto in ambito sportivo) figure professionali: il Data Analyst e il Data Scientist⁹.

Queste figure sono coinvolte sempre in maggior numero nel lavoro quotidiano di team sportivi a livello professionistico. Qual è dunque la natura del loro lavoro?

Innanzitutto entrambi sono deputati alle operazioni di Data Analytics, che si può definire come tutto ciò che riguarda i processi di estrazione ed elaborazione di informazioni a partire da un determinato bacino di dati¹⁰.

Nello sport nel dettaglio il Data Analyst è la figura professionale preposta alla raccolta e rapida elaborazione dei dati di allenamento e gara, con l'obiettivo di fornire informazioni utili al processo decisionale di coaching. Fondamentali in queste operazioni sono la rilevanza delle informazioni elaborate per la prestazione e la loro validità.

La figura del Data Scientist, ovvero di colui che grazie alle sue conoscenze del processo di allenamento e dell'approccio empirico-scientifico elabora procedure valide e efficaci per sostenere la dinamica degli adattamenti (allenamento), nasce quindi per garantire validità e

⁹ *The Role of Data Science in Sports, mastersintdatascience.com*

¹⁰ "Data Scientist vs Data Engineer vs Data Analyst – What really differentiates them?", data-flair.traninig

rilevanza delle variabili analizzate per la prestazione. Idealmente il compito del Data Scientist è quello di strutturare l'esperienza di allenamento attraverso la validazione delle procedure mentre il Data Analyst si occupa del loro impiego e della rapidità e sostenibilità del processo¹¹. Questa definizione proposta dal CONI, fa riferimento più che altro a sport dove queste figure professionali incentrano il loro lavoro sul miglioramento della prestazione fisica, con competenze quindi più attinenti alla sfera medico-sportiva. Ci sono però sport, come il Tennis ad esempio, in cui invece Data Analysts e Data Scientists vengono impiegati principalmente per svolgere processi di Match Analysis, per cui le competenze richieste riguardano in particolar modo il mix tra scienze statistiche e conoscenze tecniche e tattiche dello sport in questione. In ogni caso sono fondamentali per queste figure competenze interdisciplinari, che toccano anche programmazione, machine learning (fondamenti delle tecniche di Intelligenza Artificiale e implementazione delle stesse) e data visualization (che riguarda la rappresentazione grafica dei risultati delle elaborazioni effettuate).

1.4 PERFORMANCE ANALYSIS NELLO SPORT

La Performance Analysis consiste nell'insieme dei processi che mirano a consentire l'analisi e la spiegazione accurata di un fenomeno motorio e sportivo. Essa non è legata esclusivamente alla prestazione sportiva, ma sicuramente trova ampio spazio in questo ambito di applicazione (Hughes, Lipoma, & Sibilio, 2015).

La Performance Analysis in generale (per esempio può riguardare le prestazioni di un'azienda) si basa sull'identificare KPI (Key performance indicators) ovvero "indicatori chiave di prestazione" che forniscono un parametro oggettivo di misurazione e un'indicazione circa alcune variabili della performance che si vuole monitorare¹².

L'indicatore chiave di prestazione (KPI) è una metrica che indica il livello di raggiungimento di un dato obiettivo da parte di un atleta o, allargando il discorso ad altri campi, di un reparto o

¹¹ Definizione CONI, scuoladellosport.coni.it

¹² "Cos'è la performance analysis?" bskilled.it

di un'azienda. I KPI che vengono individuati nei vari sport offrono un'indicazione sull'andamento delle performance individuali o di squadra, ma affinché siano veramente rilevanti nel sintetizzare correttamente la realtà è fondamentale che vengano individuati KPI realmente significativi dal punto di vista fisico, tecnico e tattico e strategico¹³.

Si tratta dunque non solo di creare e considerare indicatori di performance corretti, ma anche di determinare le aree critiche della prestazione a cui legare i KPI, identificando con particolare attenzione i cosiddetti "fattori critici di successo" (KSF), ovvero quel numero limitato di aree in cui i risultati, se sono soddisfacenti, assicurano all'organizzazione (o atleta) una performance competitiva di successo.

Per quanto riguarda lo sport, la performance analysis può consistere nell'analisi della performance atletica vera e propria o negli sport che prevedono il confronto diretto con l'avversario nell'ormai consueta "Match analysis", che può essere definita come l'analisi e visualizzazione di alcuni parametri osservabili in una competizione sportiva¹⁴ (per esempio percentuale di possesso palla, metri percorsi dai giocatori, percentuale di battute corrette). In molti sport la "Match Analysis", oltre che essere un uno strumento per agevolare la comprensione delle dinamiche dello sport in questione ai fruitori del prodotto televisivo, costituisce un elemento fondamentale per l'analisi e il miglioramento della prestazione dei singoli e della squadra.

¹³ "Key Performance Indicators" valentinomea.it

¹⁴ "La match analysis", scienzaemovimento.it

2. IL CASO DEL TENNIS: UN POTENZIALE ANCORA INESPRESSO

Dopo aver effettuato nel primo capitolo un discorso relativo allo sport in generale, in questa seconda parte verrà introdotto un discorso più specifico riguardo la Tennis Data Analytics.

Il Tennis è infatti lo sport scelto per l'approfondimento delle tematiche riguardanti l'applicazione dei processi di analisi statistica allo sport poiché caratterizzato da notevole una vastità e varietà di informazioni potenzialmente disponibili relative alla performance degli atleti.

2.1 TENNIS E DATI: UNO SGUARDO GENERALE

Il tennis è uno degli sport di più antica tradizione e tra quelli maggiormente praticati al mondo: infatti secondo il "Global Tennis Report" redatto dell'ITF, la federazione tennistica internazionale, nel 2019 e basato su un'indagine che comprende i dati di 195 nazioni forniti dalle rispettive federazioni nazionali, il numero di giocatori raggiungerebbe gli 87 milioni, vale a dire l'1,7% della popolazione mondiale. Gli appassionati di tennis si distribuiscono in 71.263 mila club in 32 nazioni e 489.135 campi da gioco nel mondo: per ogni club, ci sarebbero quindi circa 771 praticanti¹⁵.

Per quanto riguarda l'accostamento del tennis al mondo della Data Analytics, bisogna cominciare sottolineando che il contesto offre ottime premesse: il tennis infatti è uno sport che ben si presta all'analisi dei dati poiché il sistema di punteggio utilizzato è caratterizzato da unità elementari, i punti, gerarchicamente strutturati in aggregati di punteggio discreti (game e set) e inoltre l'esito finale dei match è binario (vittoria o sconfitta).

Paradossalmente però i dati solitamente disponibili pubblicamente sono solitamente incentrati sugli aggregati di punteggio elementari (punti, game e set vinti), i singoli "punti importanti" vinti (i cosiddetti "break points") e indicatori di performance non troppo approfonditi sull'unico colpo dettagliatamente tracciato (il servizio).

¹⁵ ITF Global Tennis Report, 2019, itftennis.com

Oltretutto per quanto riguarda le statistiche offerte ai fruitori del prodotto televisivo (i proventi derivati dai diritti tv sono strategici per quanto riguarda l'economicità dei circuiti professionistici), che dovrebbero contribuire ad elevarne la qualità, bisogna evidenziare che ci sono importanti margini di miglioramento, se si considerano le potenzialità di questo sport.

Ad ogni modo, nonostante un evidente ritardo se comparato ad altri sport, specialmente quelli appartenenti al mondo del professionismo americano (Basket e Baseball in primo luogo), anche il Tennis sta cercando di compiere passi in avanti verso un approccio sempre più rivolto alle filosofie data-driven, anche tramite l'inserimento delle figure professionali citate in precedenza nei teams degli atleti. L'uso di tecniche di analisi dei dati infatti sta diventando sempre più consueto fra i top player e sta dando vita a un cambiamento anche culturale, richiedendo la presenza fissa all'interno dei team degli esperti di analisi dei dati che possano in primo luogo analizzare le grandi quantità di dati potenzialmente a disposizione e riuscire inoltre ad offrire ai coach informazioni ed elaborazione di facile lettura. A partire da queste informazioni, il team tecnico di un atleta può costruire sia i piani tattici per i singoli match (la strategia), quanto sviluppare piani di allenamento che, nel lungo periodo, possono modificare e migliorare il suo modo di giocare

Nel circuito professionistico femminile (WTA) gli strumenti di analisi dei dati raccolti nei tornei sono attualmente forniti da SAP¹⁶ e consentono però ai soli allenatori di accedere temporaneamente alle informazioni più dettagliate sull'andamento dei match in tempo reale. Nel circuito maschile invece (ATP) un altro colosso dell'Information Technology come Infosys ha recentemente rinnovato l'accordo di collaborazione con ATP in qualità di Global Technology Services Partner and Digital Innovation Partner fino al 2023¹⁷.

¹⁶ "Wta-Sap, la tecnologia al servizio del tennis femminile", 2016, F.Pallone, repubblica.it

¹⁷ "ATP and Infosys Extend Digital Innovation Partnership", Infosys.com

2.2 CLASSIFICAZIONE DELLE TIPOLOGIE DI DATI RACCOLTI NEL TENNIS

Il lavoro di analisi statistica applicata allo sport prevede per prima cosa la raccolta dei dati correlati alla prestazione, che possono essere però di diversa natura anche all'interno della stessa disciplina.

Nel tennis i dati raccolti tramite le più varie modalità, manualmente o tramite sistemi tecnologici, possono essere classificati secondo modello presentato in Figura 2.

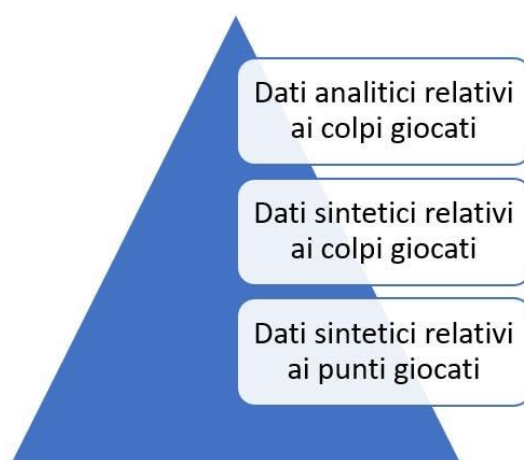


Fig.2 "Classificazione delle tipologie di dati potenzialmente raccolti ai fini dell'analisi statistica nel tennis"
Fonte: "Tennis e dati: ok, ma quali sono quelli disponibili? Dai più grezzi ai dati Hawkeye", F. Bertelli, 2020
Ubitennis.com

Secondo questa classificazione proposta dalla testata online specializzata Ubitennis¹⁸, si possono identificare tre diverse categorie di informazioni, che verranno citate in ordine crescente di complessità e per cui verrà effettuata di seguito una spiegazione dettagliata:

1. Dati sintetici relativi al punteggio
2. Dati sintetici relativi ai colpi giocati
3. Dati analitici relativi ai colpi giocati (dati Hawkeye)

¹⁸ "Tennis e dati: ok, ma quali sono quelli disponibili? Dai più grezzi ai dati Hawkeye", F. Bertelli, 2020, ubitennis.com

LA PRIMA CATEGORIA: DATI SINTETICI RELATIVI AL PUNTEGGIO

I dati sintetici relativi ai punti giocati sono dati “grezzi” ovvero non elaborati, che si riferiscono a tutto ciò che riguarda esclusivamente il punteggio.

Fanno parte di questa categoria, per esempio, i dati riguardanti la quantità di punti vinti da un giocatore, sia totali che percentuali in relazione alle principali fasi del gioco (servizio e risposta). L’oggetto di analisi di questa tipologia di dati sono le unità gerarchiche che fanno parte del sistema di punteggio tennistico ovvero i punti, “i games” e i “set”: questi elementi costituiscono dunque le unità base delle statistiche disponibili.

La maggior parte degli indicatori costruiti a partire da questa prima categoria è inerente alla fase di gioco di “servizio” (gli indicatori in questo possono essere per esempio percentuale di punti vinti al servizio o percentuale di turni di servizio vinti).

Le indicazioni ottenute da questi parametri sono tuttavia di scarsa profondità e rilevanza statistica, soprattutto se si vuole compiere analisi approfondite riguardo le performance dei tennisti e studiare le variabili tattiche che emergono nel gioco del tennis.

Nella realtà il punto può avere la composizione più variegata possibile dal punto di vista del gioco: lo svolgimento di un punto può infatti svilupparsi secondo infinite modalità, come combinazione delle innumerevoli variabili che potenzialmente si possono affrontare durante il gioco.

Partendo da dati di questo livello analizzare e interpretare un match di Tennis diventa di conseguenza piuttosto superficiale.

LA SECONDA CATEGORIA: DATI SINTETICI RELATIVI AI COLPI GIOCATI

La seconda categoria di dati è quella relativa ai dati di sintesi su singoli colpi giocati.

Come affermato in precedenza ogni punto giocato durante un match di tennis può svolgersi nei modi più differenti possibili: in questo sport i “colpi” possono essere considerati come le unità fondamentali, tramite il cui susseguirsi (alternando un colpo per giocatore) si articola lo svolgimento del punto.

Prendendo quindi in considerazione i colpi come prospettiva di analisi, si possono ottenere molte più chiavi di lettura per interpretare e comprendere il gioco del tennis, ma solitamente già questa categoria di informazioni non è di semplice reperibilità.

I dati di sintesi dei colpi giocati comprendono per esempio le informazioni riguardo lunghezza degli scambi (numero di colpi totali per punto), il numero di colpi vincenti (colpo che risolve lo scambio a favore del giocatore che lo effettua), il numero di errori forzati (provocati dall'avversario), e la fondamentale distinzione tra servizio, dritto, rovescio e colpi di volo giocati. Grazie a questa tipologia di informazioni perciò, oltre a conoscere se il punto si è giocato sul servizio del giocatore A o del giocatore B o da chi è stato vinto, si può dunque cominciare a conoscere quanto è stato lungo, che tipi di colpi sono stati giocati e come si è concluso il punto, tutte variabili determinanti ai fini della comprensione del gioco.

Le informazioni citate vengono raccolte dai tornei tramite i partner digitali (come Infosys, SAP, e IBM), ma solitamente non ne viene consentita l'accessibilità al pubblico, o quantomeno rimangono di proprietà dei tornei. Bisogna inoltre sottolineare come questi dati, qualora siano presentati dalle statistiche ufficiali, sono come detto mostrati sotto forma di aggregati (numero totale di dritti vincenti, numero totale di errori di rovescio) quindi gli schemi tattici che determinano un match rimangono di difficile interpretazione.

Esistono però progetti di Match Charting, che si basano sulla catalogazione colpo per colpo (ovvero effettuando operazioni di "Tagging manuale") di informazioni di questa categoria e sulla successiva condivisione dei risultati su piattaforme di crowdsourcing, in modo che essi possano assumere rilevanza in ambito di elaborazione e consentire di poterne ricavare le analisi delle trame tattiche di un match.

LA TERZA CATEGORIA: DATI ANALITICI RELATIVI AI COLPI GIOCATI (DATI HAWK-EYE)

La terza categoria infine riguarda i dati analitici relativi ai singoli colpi giocati, ovvero i dati raccolti durante i match dal sistema Hawk-eye (o da analoghi impianti tecnologici installati sul campo di gioco), il cui accesso è però concesso solo a poche organizzazioni, come le aziende principali di Sport Analytics, che riescono pertanto, dietro compenso, a fornire report avanzati ai tennisti che lo richiedono.

Come illustrato in Figura 3, il sistema Hawk-eye è in grado di fornire una quantità di dati impressionante per varietà e accuratezza: per ogni colpo giocato il sistema fornisce infatti tutte le informazioni riguardanti velocità, traiettoria, livello di rotazione, altezza sopra la rete, punto di rimbalzo della pallina. Inoltre, per quanto riguarda i giocatori, viene registrato il posizionamento al momento dell'impatto e i dati sui movimenti effettuati.

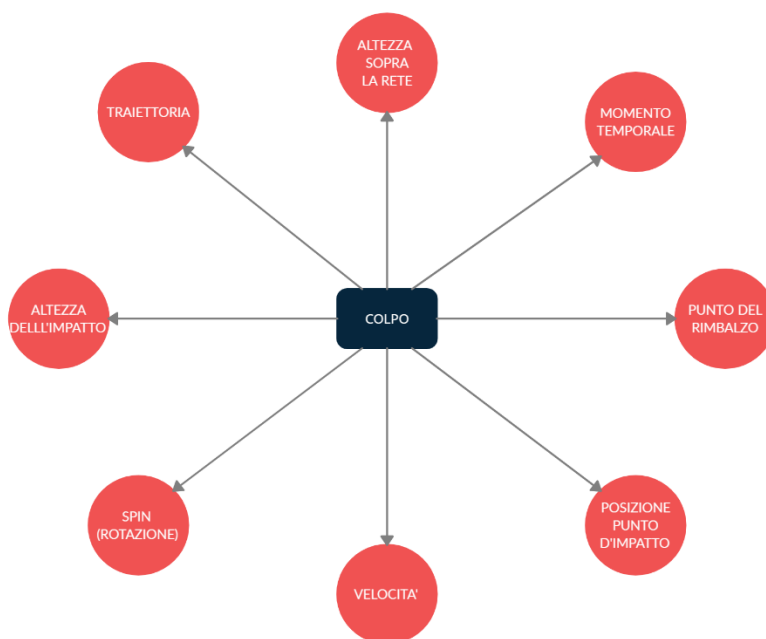


Fig.3 "Panoramica delle varie informazioni offerte dalla tecnologia Hawk-Eye"

E' chiaro che, proprio quest'ultima tipologia di dati, offre potenzialità di analisi pressoché infinite, poiché la varietà e le possibilità di correlazione delle informazioni raccolte è davvero significativa e l'accesso libero del pubblico ai dati Hawk-eye potrebbe rappresentare una svolta per quanto riguarda il mondo della Tennis Data Analytics poiché anche gli appassionati con competenze in merito potrebbero elaborare da sé i dati e offrire validi spunti ai fini della piena descrizione e comprensione delle dinamiche di questo sport.

Oltre a questo, permangono anche le potenzialità per migliorare ancora il prodotto televisivo (che contribuisce a portare centinaia di milioni di euro di introiti al circuito), poiché le statistiche spesso mostrate sono di scarsa rilevanza, anche se tuttavia, negli ultimi anni, si è venuta a creare una sensibilità più profonda rispetto alla materia, sia per cercare come detto

di rendere il prodotto più completo, sia per l'effetto traino dovuto ad altri sport come Basket e Baseball.

IL SISTEMA HAWK-EYE



Fig. 4 "Schema di funzionamento del sistema Hawk-Eye"

Fonte: "THE IMPACT OF THE HAWK-EYE SYSTEM IN TENNIS", trainingwithjames.wordpress.com

Si è menzionato già svariate volte il famigerato sistema Hawk-eye. Di che cosa si tratta?

Il sistema Hawk-Eye è stato concepito nel 2001 dagli ingegneri del Roke Manor Research Limited di Romsey¹⁹, in Inghilterra ed è stato utilizzato per la prima volta nel cricket nel Maggio dello stesso anno in un test match. Hawk-eye è stato poi introdotto nel tennis dove è diventato un sistema (nelle maggior parte dei tornei) di arbitraggio ufficiale a partire dal 2005. Nel 2011 la Hawk-eye Innovations Ltd., società proprietaria del brevetto e del marchio Hawk-eye è stata acquisita da Sony Professional Solutions²⁰.

Nelle competizioni tennistiche di primo livello (ATP/WTa/ITF) il sistema Hawk-eye viene utilizzato principalmente come strumento di arbitraggio elettronico, ma non solo: il sistema infatti è in grado di fornire anche tutte le informazioni riguardanti ogni singolo colpo effettuato dai giocatori in campo, diventando dunque uno strumento di primaria importanza per la raccolta di Big Data nel Tennis.

¹⁹ "Hawk-Eye", wikipedia.it

²⁰ "Hawk-Eye Innovations Ltd has been acquired by Sony Professional Solutions", oaklinks.com

Essenzialmente, il sistema di tracciamento si basa sui principi della triangolazione e utilizza immagini visive e dati di temporizzazione, fornite dalle telecamere ad alta velocità installate intorno allo stadio (Figura 4), in genere una decina, che vengono calibrate e sincronizzate prima di ogni evento. Le telecamere sono solitamente posizionate in una posizione elevata al di sopra dei campi da gioco in modo tale da poter catturare la traiettoria delle palline²¹.

2.3 FONDAMENTI DI TENNIS DATA ANALYTICS

I processi di analisi ed elaborazione delle informazioni ottenute nell'ambito dello svolgimento della pratica dello sport del Tennis possono essere raccolti sotto la denominazione di "Tennis Data Analytics".

Si possono individuare tre componenti fondamentali nel lavoro di Data Analytics applicato al Tennis: la prima riguarda la raccolta dei dati, che può essere effettuata manualmente (grazie ai più disparati sistemi di raccolta dati, dai sensori wearable, ai sistemi di tagging video manuali), tramite fonti ufficiali o a partire dai progetti di crowdsourcing; la seconda consiste nell'analizzare le informazioni raccolte, elaborarle e trovare eventuali correlazioni tra esse. La terza invece riguarda la rappresentazione e la traduzione in termini applicabili al lavoro quotidiano di allenamento o di preparazione al match delle informazioni estratte per gli utenti finali²².

Per fare ciò, viene creato ed utilizzato come database dai Data Analysts un "Data Lake", una sorta di magazzino di dati centralizzato nel quale sono raccolti dati strutturati e non provenienti dalle più svariate fonti²³. Le società meglio strutturate che offrono questi servizi si avvalgono anche di sistemi di intelligenza artificiale per automatizzare e rendere più efficiente il processo di raccolta dati e di analisi degli stessi.

²¹ "Tennis e dati: come si raccolgono i dati sul campo (e quanti soldi servono)", *Ubitennis.com*

²² Estratto di intervista a Shane Liyange, *Tennis Data Analyst, ubitennis.com*

²³ "Che cos'è un data lake?", *neodatagroup.com*

Vista la generale difficoltà nel reperire elaborazioni statistiche ufficiali e l'interesse crescente di giocatori e coach anche a livelli non di eccellenza assoluta, stanno nascendo diverse realtà più piccole, start-up o liberi professionisti che offrono report basati invece sul Match Charting svolto manualmente, con l'obiettivo di costruire una base dati sufficiente per poter effettuare rilevazioni sui singoli giocatori e sulle loro caratteristiche.

L'espressione "Match Charting" deriva dal Match Charting Project, ovvero il maggiore progetto indipendente di crowdsourcing per raccogliere dati punto per punto delle partite di tennis professionistico. Viene definito dai suoi stessi ideatori²⁴ come un sistema di codifica di informazioni che chiunque può utilizzare per registrare ogni colpo di una partita e nel dettaglio le informazioni riguardanti la direzione del servizio, la direzione e profondità della risposta, la direzione dei colpi di scambio, la tipologia e la direzione degli errori. Lo scopo del progetto è quello di creare un'aggregazione di dati di elevata dimensione per favorire i processi di analisi dei dati ed elaborazione degli stessi. Infatti ogni partita aggiunta al database, contribuisce ad ottenere una maggiore potenzialità di analisi e di comprensione delle dinamiche che intervengono nel tennis professionistico.

Gli obiettivi della Tennis Data Analytics sono principalmente due: il primo è raccogliere dati sul giocatore in oggetto e sui suoi schemi strategici, così da aiutare lui e il suo team a capire quali siano quelli più redditizi, quelli da migliorare o quelli che necessitano di cambiamenti; il secondo, è lo studio dell'avversario per ogni singolo match, in modo da sviluppare così un piano di gioco vincente che tenga conto dei punti di forza e dei punti di debolezza sia propri, sia dell'avversario. Si effettua perciò una sorta di analisi dei punti di forza-punti di debolezza dei concorrenti, analogamente a quanto avviene per le imprese, le quali, in sede di pianificazione del marketing, svolgono indagini di mercato per arrivare a definire quale possa essere la strategia competitiva più redditizia.

²⁴ "Match Charting Project", settesei.it

3. TENNIS DATA ANALYTICS: IL SERVIZIO

Nel terzo capitolo saranno effettuate, riportate, spiegate e commentate nel dettaglio varie elaborazioni statistiche che si configurano come veri e propri esempi di processi di Data Analytics applicati al Tennis. In questo capitolo le analisi vertono esclusivamente sul colpo di servizio, “fondamentale tecnico” che nel tennis moderno rappresenta l’area critica probabilmente più importante nel determinare i risultati conseguiti dagli atleti nelle competizioni professionistiche.

3.1 FATTORI CRITICI DI SUCCESSO NEL TENNIS

Una delle finalità della raccolta ed elaborazione dei dati nello sport è quella di valutare globalmente la performance dell’atleta sotto un particolare profilo tecnico o nell’arco di un determinato periodo di tempo. Le valutazioni statistiche che avvengono di conseguenza si effettuano mediante l’individuazione di indicatori di performance come metro di giudizio.

Introducendo la tematica degli indicatori di performance nel tennis, come nello sport in generale, si può proporre una utile analogia con le scienze economiche e citare dunque il concetto di “fattori critici di successo”.

I fattori critici di successo sono quel numero limitato di aree, su cui un’organizzazione deve focalizzarsi allo scopo di raggiungere risultati positivi per assicurare all’azienda un rendimento competitivo positivo (Rockart, 1979). Allo stesso modo un atleta/squadra deve cercare di ottenere performance brillanti nelle componenti fondamentali che caratterizzano lo sport in considerazione e che consentono di raggiungere livelli di eccellenza.

Per monitorare l’efficienza dei vari fattori che incidono sul risultato finale vengono utilizzati allora degli indicatori di performance che valutano il livello di rendimento del fattore critico, che deve essere in linea con gli obiettivi stabiliti e con valori di riferimento esterni (benchmark) (Krajewski, Ritzman, Mlhotra, Grando, & Secchi, 2015).

I fattori critici nel tennis sono molteplici e soprattutto di varia natura, poiché si spazia dai fattori tecnici, ovvero la qualità dei colpi, alle qualità fisiche (velocità, flessibilità, resistenza, forza, esplosività), fino ad arrivare alla forza psicologica. Bisogna però sottolineare che l'area più importante si può ritenere quella relativa alla qualità dei colpi poiché si tratta della "conditio sine qua non" per emergere a livello professionistico: ci possono infatti essere atleti con colpi adeguati e lacune nella fisicità ma è piuttosto inverosimile il contrario.

Per quanto concerne l'ambito tecnico, possono dunque essere considerati come "aree critiche fondamentali" i 4 colpi principali (evidenziati in Figura 5) che caratterizzano il gioco, ovvero:

1. Servizio: colpo d'inizio gioco, ovvero primo colpo di ogni punto
2. Risposta: colpo (o fase di gioco) di risposta al servizio
3. Dritto
4. Rovescio

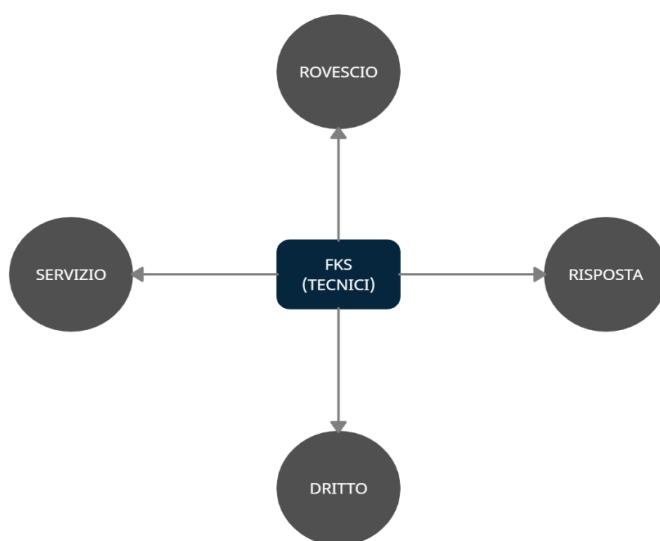


Fig. 5 "Fattori critici di successo a livello tecnico nel tennis"

Tra tutti, il colpo considerato a maggiore impatto sul risultato finale di un match di tennis è il servizio, che sarà dunque oggetto quindi di questa analisi, non solo per la sua incidenza sulla performance finale, ma anche perché rappresenta il comparto che, ai fini di un'analisi

statistica accurata, offre dati più “puliti” e con minori limiti tecnici per via della sua natura (discorso che sarà approfondito in seguito).

Verranno dunque citati gli indicatori di performance in uso attualmente, i fattori che determinano il mantenimento di alte prestazioni al servizio e verrà infine proposto uno strumento di rating del colpo, che secondo il sottoscritto presenterebbe meno lacune rispetto agli indicatori in uso attualmente.

3.2 ANALISI DI UN FATTORE CRITICO: IL SERVIZIO

Per prima cosa vengono effettuate di seguito delle premesse tecniche e regolamentari indispensabili ai fini della comprensione del lavoro presentato.

PREMESSE REGOLAMENTARI E GLOSSARIO

Il servizio nel tennis è il colpo di inizio gioco. I giocatori durante un match si alternano in turno di servizio (ogni turno detto game, si conclude a favore di un giocatore dopo almeno 4 punti vinti, a condizione che si siano ottenuti almeno due punti in più rispetto all'avversario) e turno alla risposta.

Per ogni punto il giocatore al servizio ha a disposizione due battute: in caso di errore (se il rimbalzo avviene fuori dal rettangolo di competenza) sulla prima (detta "prima di servizio" o semplicemente "prima") il giocatore può effettuarne un'altra, mentre in caso di errore anche sulla seconda ("seconda di servizio" o "seconda") il punto è perso e si parla di "doppio fallo". Il giocatore che non serve, si dice che "risponde" ed è definito in gergo "ribattitore".

PREMESSE TECNICHE

Si è parlato precedentemente di come il colpo meglio tracciato dalle statistiche ufficiali del tennis sia il servizio e di quanto proprio questo colpo sia quello considerato il più importante in questo sport.

Innanzitutto bisogna spiegare questa affermazione ovvero partire dal rispondere alla domanda: come mai è considerato un colpo così importante?

Il servizio è il colpo d'inizio gioco e in quanto tale è anche l'unico colpo che con certezza sarà giocato in ogni scambio (scambio è un termine per identificare punto), potenzialmente infatti una partita potrebbe infatti concludersi con soli servizi, se entrambi i giocatori riuscissero in ogni punto ad effettuare un colpo vincente (identificato del termine "Ace" nel caso del servizio) o costringere l'avversario ad una risposta errata con questo fondamentale.

Bisogna poi considerare che, svolgendo un lavoro di analisi statistica, la raccolta dati inerenti al servizio offriranno dati tutto sommato "puliti" e rilevanti a fini statistici in quanto è un colpo che l'atleta effettua da fermo, sempre nelle medesime condizioni e non è influenzato da altri

fattori come il livello dell'avversario, che invece impatta in qualsiasi altra situazione durante il gioco. Teoricamente ci sono altri fattori di tipo "ambientale" che potrebbero influenzare l'esecuzione del servizio, come orario di gioco, temperatura dell'ambiente, umidità, altitudine, caratteristiche fisiche delle palline e infine la tipologia (superficie) del campo: tranne quest'ultima variabile, che viene presa in considerazione in alcune delle analisi statistiche effettuate, le altre possono essere ritenute ininfluenti poiché considerate ad impatto minimo e inoltre difficilmente quantificabili in base ai dati solitamente disponibili.

Per quanto riguarda, come accennato, alle diverse superfici di gioco, queste possono essere divise in quattro categorie: citando in ordine decrescente di "velocità della superficie" troviamo erba, campi in cemento indoor, campi in cemento outdoor, e terra battuta²⁵.

Per velocità bisogna specificare che i campi non incidono sulla velocità del colpo in partenza dalla racchetta, (quella che fa fede a fini statistici poiché rilevata all'uscita del colpo dalla racchetta e quindi scevra da qualsiasi influenza esterna), ma solo sulla velocità finale (quella "percepita" dal ribattitore) dopo il rimbalzo nel campo dell'avversario, poiché alcune tipologie di campi di gioco rallentano la pallina (attrito) più di altre dopo il contatto con esse.

Per riassumere, la superficie influenza l'efficacia del colpo, ma non la sua velocità intrinseca. Infine bisogna aggiungere un ultimo presupposto: nel tennis moderno, per quanto riguarda il servizio, viene adottata solitamente dalla totalità dei giocatori una determinata strategia decisionale: la prima di servizio viene infatti eseguita in modo aggressivo per cercare il punto, mentre la seconda è eseguita in un modo più conservativo, in modo da scongiurare il rischio di doppio fallo.

Fatte le dovute premesse si possono introdurre gradualmente alcune elaborazioni statistiche di notevole importanza ai fini della comprensione della materia affrontata e dell'ultimo "assioma strategico" appena citato.

²⁵ *"Le superfici non sono diventate tutte uguali, lo dicono i dati", V. Vignoli, Ubitennis.com*

3.2.A QUANTIFICARE L'IMPATTO DEL SERVIZIO NEL TENNIS

Il punto di partenza è il concetto già espresso, ma non spiegato, che il servizio è il colpo probabilmente a maggior impatto nel tennis moderno. Perché si fa questa affermazione?

Dalle statistiche ufficiali ATP (e quindi le meno elaborate) risulta che in media circa due terzi dei punti vinti in ogni partita, sono conquistati durante i propri turni di battuta²⁶. La domanda a cui si vuole rispondere è però la seguente: quanti di questi punti conquistati durante i turni di battuta sono dovuti al colpo di servizio? Con questo si intende non solo i punti direttamente ottenuti con il primo colpo, ma la percentuale complessiva dei punti in cui non solo il servizio ha prodotto il punto ma anche quelli in cui è stato determinante ai fini della conquista dello stesso, poiché ha consentito al giocatore di porsi in una situazione di vantaggio nello scambio. A partire dai dati di Match Charting Project (citato in precedenza), è stato effettuato per TennisAbstract, portale online di tennis Data Analytics, da Jeff Sackmann (Data Analyst e sviluppatore di software per l'analisi statistica in ambito sportivo) uno studio²⁷ utilizzando come campione circa 1.200 partite maschili del circuito professionistico dal 2000 ad oggi.

Sackmann ha innanzitutto calcolato per ogni scambio (punto) da "x" numero di colpi effettuati (tranne che per i punti da 1 o 2 colpi, per cui si presume che siano nel 100% dei casi determinati dal servizio) quanti in percentuale terminano a favore del battitore.

Chiaramente le percentuali di vittoria da parte del battitore si abbassano man mano che si allunga lo scambio e ci si allontana dal servizio, fino a stabilizzarsi su un "base rate" (il "base rate" dovrebbe tendere al 50% ovvero probabilità eque per entrambi i giocatori) a partire dal quinto colpo del battitore, ovvero il momento in cui statisticamente i vantaggi acquisiti grazie al servizio nello scambio "svaniscono" e quindi il momento in cui le probabilità di vincere il punto per il battitore e l'avversario cominciano ad essere sostanzialmente pari.

Nella Figura 6 viene illustrato il "point ending rate" ovvero per scambio da "x" numero di colpi, quanti terminano a favore del battitore al colpo "x" (probabilità congiunte) e non le probabilità generale di vincere il punto del battitore (probabilità condizionata).

²⁶ ATP Leaderboards, Serve Stats, ATP.com

²⁷ "Measuring the Impact of the Serve in Men's Tennis", 2017, J. Sackmann, Tennisabstract.com

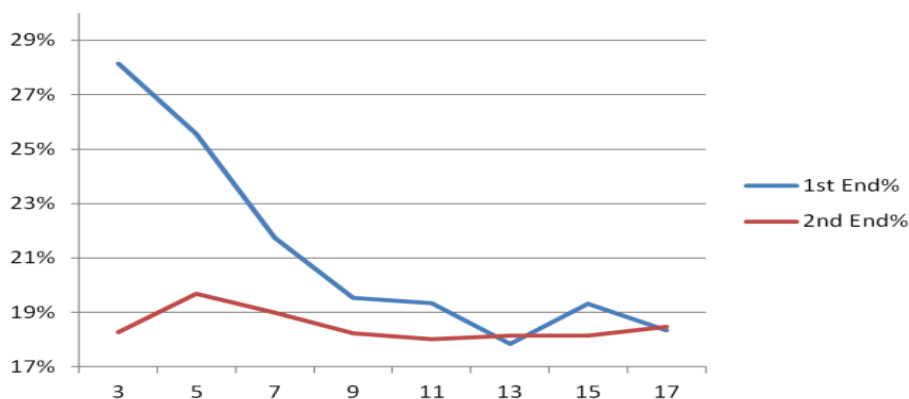


Fig. 6 “Probabilità del battitore di vincere il punto al colpo numero “x” dello scambio”. Sull’asse x viene indicato il numero di colpi complessivo, sull’asse y la percentuale di scambi che si concludono a favore del battitore in “x”. (In blu la probabilità in caso di punto giocato sulla prima di servizio, in rosso la probabilità in caso di punto giocato sulla seconda) Fonte: *Measuring the Impact of the Serve in Men’s Tennis*

A partire da questi risultati, che indicano come la probabilità di vincere il punto del battitore si assesti dopo il proprio quinto colpo e calcolando poi il differenziale tra “Winners rate per shot” (probabilità di effettuare un colpo vincente all’ “x” colpo) e il “Winners base rate” (probabilità base di effettuare un colpo vincente), Sackmann ha potuto sviluppare un indice denominato Serve Impact, “SI” (impatto del servizio sul punto), che determina in che percentuale il servizio è determinante nella conquista di un punto durante il proprio turno di battuta.

I risultati del suo studio, separatamente per superficie di gioco e prima o seconda di servizio, sono illustrati in Figura 7.

	1st SERVE IMPACT	2nd SERVE IMPACT	Total SERVE IMPACT
Tutte le superfici	63.4%	31.0%	53.6%
Cemento	64.6%	31.5%	54.4%
Terra	56.9%	27.0%	47.8%
Erba	70.8%	37.3%	61.5%

Fig.7 “Serve impact percentage, ovvero la quantità di punti vinti grazie al vantaggio procurato dal colpo di servizio rispetto ai punti totali effettivamente conquistati durante il proprio turno di servizio”
Elaborazione personale dei dati di “*Measuring the Impact of the Serve in Men’s Tennis*”, 2017, J. Sackmann, Tennisabstract.com

Risulta dunque, dopo un'analisi di dati sintetici di Match Charting, che il singolo colpo di servizio ha un impatto determinante sulla metà (53,6% la media di tutte le superfici) dei punti conquistati durante il turno di battuta e, conseguentemente a quanto affermato in precedenza, ovvero che quest'ultimi rappresentano più di due terzi dei punti vinti da un giocatore in ogni partita, si può affermare che più di un terzo dei punti che ogni giocatore conquista in un match, è ottenuto grazie al solo colpo di servizio.

E' un dato molto significativo poiché si può sostenere che gli altri due terzi dei punti ottenuti in una partita vengono ottenuti grazie ad un mix di molte variabili (altri colpi, strategia di gioco, tenuta mentale, tenuta atletica e via dicendo), oltretutto difficilmente esprimibili con precisione. Gli altri compartimenti indicati come fattori critici di successo principali infatti, durante lo scambio si sovrappongono, si sommano e vengono affiancati da altri fattori secondari solamente accennati in sede di introduzione.

3.2.B INDICATORI DI PERFORMANCE E DI RATING IN USO ATTUALMENTE

Si è già accennato come il servizio sia il colpo meglio tracciato dalle statistiche ufficiali, quelle che per esempio vengono offerte ai telespettatori durante/al termine di un match. E' dunque possibile tramite indicatori di performance e sistemi di rating parametrizzare il colpo e stilare un ranking sulla base della bravura dei giocatori in questo fondamentale?

L'ATP ha creato di recente un sistema di rating del servizio che aggrega e pondera le varie voci statistiche principali inerenti al colpo.

Di seguito viene mostrato il ranking dei 10 migliori battitori del circuito basato su questo sistema di rating offerto da ATP Leaderboards (sezione online dove sono raccolte le statistiche del circuito ATP), con le varie voci che lo determinano.

GIOCATORE	SERVE RATING	% PRIME IN CAMPO	%PUNTI VINTI SULLA PRIMA	%PUNTI VINTI SULLA SECONDA	%TURNI DI SERVIZIO VINTI	ACES PER MATCH	DOPPI FALLI PER MATCH
ISNER	317	71,1%	79,1%	55,9%	92,7%	20,5	2,3
RAONIC	309,6	62,5%	83,6%	56,5%	93,5%	17,2	3,7
OPELKA	302,5	64,0%	79,3%	53,3%	90,0%	19,0	3,1
KYRGIOS	299,1	67,7%	78,5%	50,5%	88,5%	16,8	2,9
FEDERER	295,9	64,8%	77,1%	58,2%	89,8%	7,7	1,7
BERRETTINI	289,2	61,4%	78,9%	54,1%	88,1%	8,9	2,2
NADAL	289,1	64,2%	75,2%	58,6%	88,6%	4,4	1,9
QUERREY	285,8	60,7%	80,1%	47,6%	84,7%	16,5	3,8
TSITSIPAS	285,2	64,3%	75,9%	54,4%	86,1%	6,7	2,2
TSONGA	284,7	61,1%	79,4%	50,9%	85,4%	10,3	2,4

*Fig.8 "Serve rating ATP index" Elaborazione personale dei dati di ATP Leaderboards.
(Vengono riportati solo i primi 10 giocatori per Serve Rating)*

Come notiamo in Figura 8, gli indicatori che concorrono alla formazione del punteggio di rating sono

1. Percentuale di prime di servizio in campo
2. Percentuale di punti vinti con la prima palla di servizio in campo
3. Percentuale di punti vinti con la seconda palla di servizio in campo
4. Percentuale di game di servizio vinti
5. Numero medio di aces per match
6. Numero medio di doppi falli per match

Di questi sei indicatori, però, la maggior parte presenta a mio avviso dei limiti evidenti a livello tecnico/statistico che cercherò di spiegare nel dettaglio:

I parametri **2. 3. e 4.** Sono indicatori che chiaramente variano all'aumentare della bontà del colpo ma allo stesso tempo si basano complessivamente sul numero punti vinti e quindi sono influenzati dall'efficacia del giocatore nelle altre aree del gioco.

I parametri **5. e 6.** (Numero medio di aces e doppi falli per match) andrebbero invece ponderati con la durata del match, poiché sono numeri che si alzano all'aumentare dei punti giocati. Sarebbe quindi più utile considerare la frequenza di aces e/o di doppi falli, dividendo il numero di aces e/o doppi falli per il numero di punti di servizio giocati (o addirittura nel caso dei doppi falli, dividendo per il numero di seconde di servizio giocate)

Per quanto riguarda il parametro **1.** Invece esso può essere considerato come indicatore tutto sommato corretto poiché indipendente da altre variabili.

Si può notare che le statistiche tradizionali presentano quindi dei limiti, ma cosa si potrebbe fare allora per offrire strumenti di valutazione più precisi?

LA MIA PROPOSTA: UN SISTEMA RATING BASATO ESCLUSIVAMENTE SUL VALORE ASSOLUTO DEL SINGOLO COLPO

Dal mio personale punto di vista, considerando quanto appena affermato, sarebbe più corretto introdurre degli indicatori basati più sulla bontà intrinseca del colpo, eliminando o lasciando come marginali quegli indicatori, invece influenzati da altre variabili, che rischiano quindi di offrire output ingannevoli.

Per comprendere questo concetto bisogna però per prima cosa analizzare i fattori determinanti che possono rendere il servizio un'arma importantissima per i giocatori, sviscerando le variabili intrinseche del colpo e in seguito le caratteristiche fisiche che agevolano alcune categorie di giocatori ad avere una resa del colpo maggiore, realizzando dunque quella che può essere vista come un "indagine di mercato", volta a comprendere quali sono le dinamiche che determinano il successo competitivo di questa tipologia di "prodotto".

3.2.C SCOMPORRE IL SERVIZIO IN FATTORI: ANALISI DEL COLPO A PARTIRE DAI DATI ANALITICI DI HAWK-EYE

PREMESSE

Come già anticipato la tecnologia Hawk-Eye offre una quantità di dati impressionante per ogni singolo colpo di ogni partita, consentendo di analizzarne le informazioni riguardanti soprattutto la velocità, il piazzamento (bisogna ora pensare al campo di gioco, come un piano cartesiano) e lo spin (ovvero le rotazioni per minuto che la pallina compie attorno al proprio asse). Bisogna ora fare una doverosa premessa riguardante lo spin e la velocità del colpo, in quanto tra queste due variabili per natura vi è un trade-off, ovvero più si imprime spin alla pallina, minore sarà la velocità della stessa²⁸.

Per analizzare statisticamente questo colpo e trarre delle conclusioni sulla sua reale efficacia a partire da questa tipologia di dati più elaborata, mi rifaccio a *“The serve impact in tennis: First large-scale study of big Hawk-Eye data: The Serve Impact in Tennis”*, (Mecheri, Rioult, Mantel, Kauffman, & Benguigui, 2016) che scompone sotto 3 profili principali citati il colpo:

- 1.VELOCITA: aumentare la velocità aumenta la resa del colpo
- 2.PIAZZAMENTO: un piazzamento più angolato aumenta la resa del colpo
- 3.SPIN: uno spin estremo aumenta la resa del colpo

Di seguito vengono riportate in breve le analisi dettagliate effettuate per ogni variabile, al fine di stabilire come interagiscono tra loro.

²⁸ P. O'Donoghue and A. Ballantyne, The impact of speed of service in Grand Slam singles tennis, In Science and Racket Sports III, A. Lees, J. F. Kahn, and I. Maynard, eds. London and New York, Routledge, 2004, 179–184

LA VELOCITA' DEL COLPO

Dal campione di è emerso che le velocità medie (e la relativa deviazione standard) rilevate per primo e secondo servizio nei tornei dei massimi circuiti professionistici maschili e femminili sono le seguenti (Figura 10):

		Service	Serve speed	SD
1	Men	First service	178.30 km/h	16.79
2	Men	Second service	145.79 km/h	12.97
3	Women	First service	152.73 km/h	13.32
4	Women	Second service	130.20 km/h	11.23

Fig.10 "Velocità medie e deviazioni standard del servizio in km/h" (distinzione per uomini e donne e per prima e seconda di servizio). Fonte: (Mecheri, Rioult, Mantel, Kauffman, & Benguigui, 2016)

Vi è una differenza tra la media di uomini e donne di circa 25 km/h per quanto riguarda la prima di servizio mentre invece di circa 15 km/h per quanto riguarda la seconda.

Ma quanto è importante la velocità? Per rispondere esaminiamo i dati che raggruppano le percentuali di punti vinti per le varie fasce di velocità del servizio a livello sia del circuito femminile che di quello maschile maschile (Fig. 11).

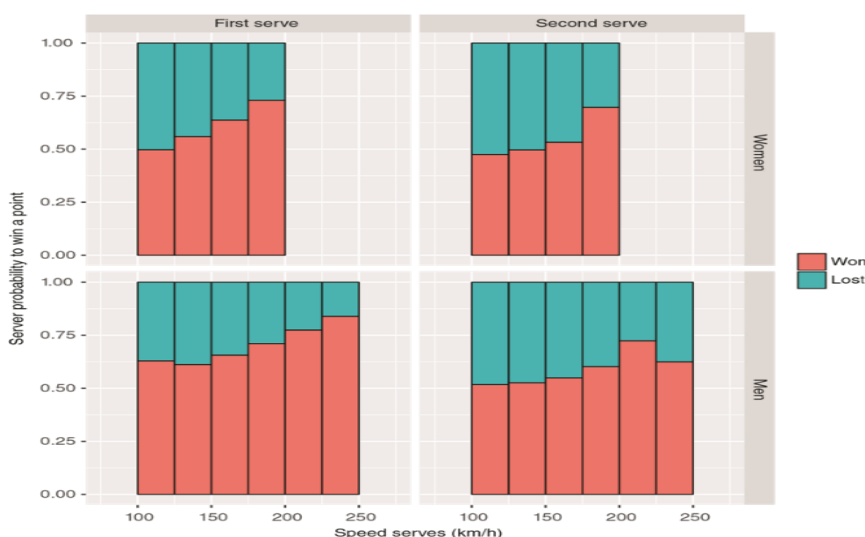


Fig.11 "Percentuale di punti vinti dal battitore in base alla fascia di velocità del suo colpo di servizio" (in rosso i punti vinti, in blu quelli persi).

Fonte: (Mecheri, Rioult, Mantel, Kauffman, & Benguigui, 2016)

Come si evince, vi è una chiara correlazione tra velocità del colpo e probabilità di vincere il punto. La probabilità di vincere il punto sale di molto al crescere della velocità, (fino ad arrivare a oltre l'80% per quanto riguarda i servizi oltre 225 km/h del settore maschile) sia per gli uomini che per le donne. Chiaramente gli uomini, avendo un vantaggio fisico, possono esprimere velocità maggiori e riusciranno a tenere mediamente una percentuale più alta di punti vinti.

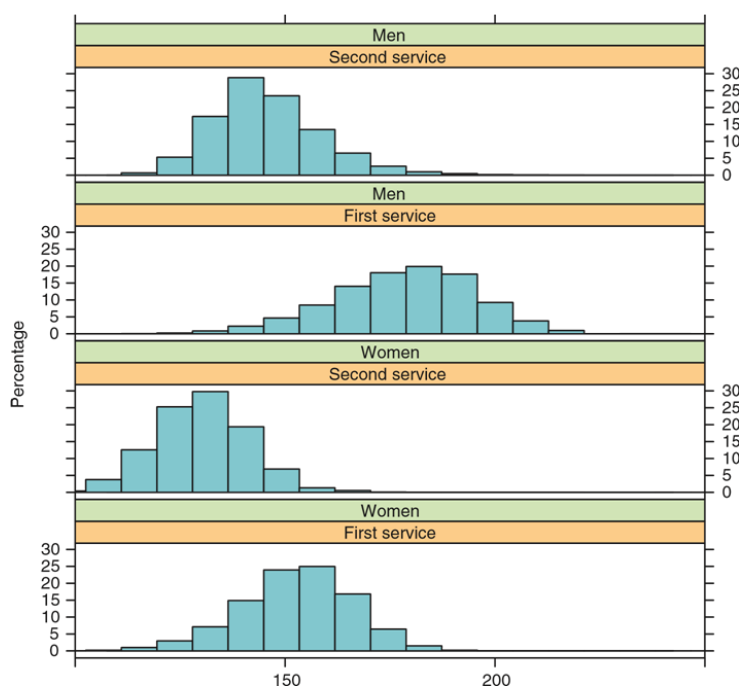


Fig.12 "Frequenza di colpi di servizio effettuati per fasce di velocità" (distinzione tra uomini, donne, prima e seconda di servizio). Fonte: (Mecheri, Rioult, Mantel, Kauffman, & Benguigui, 2016)

A partire dagli istogrammi di Figura 12, che esprimono invece la percentuale di servizi per ogni classe di velocità si può dedurre come in realtà la maggior parte dei giocatori serva con velocità in linea con i valori medi registrati, mentre sono solo pochi i giocatori che riescono a produrre colpi oltre le soglie di eccellenza necessarie per mantenere altissime le percentuali di punti vinti (soglia di eccellenza 80%).

IL PIAZZAMENTO DEL COLPO (PUNTO DEL RIMBALZO)

Si è dimostrato come la velocità del servizio abbia un ruolo fondamentale nell'aumentare la resa dello stesso colpo. E' però logico intuire che a parità di velocità, il piazzamento (in che punto del rettangolo di servizio, zona in cui si deve piazzare il colpo per non commettere un fallo, la pallina rimbalzerà) abbia un ruolo importante nell'efficacia del colpo: si può facilmente ipotizzare che più la pallina rimbalzerà vicina alle estremità (linee) del campo più questa sarà difficile da ribattere per il giocatore in risposta. Questa ipotesi viene confermata dai dati riportati in Figura 13.

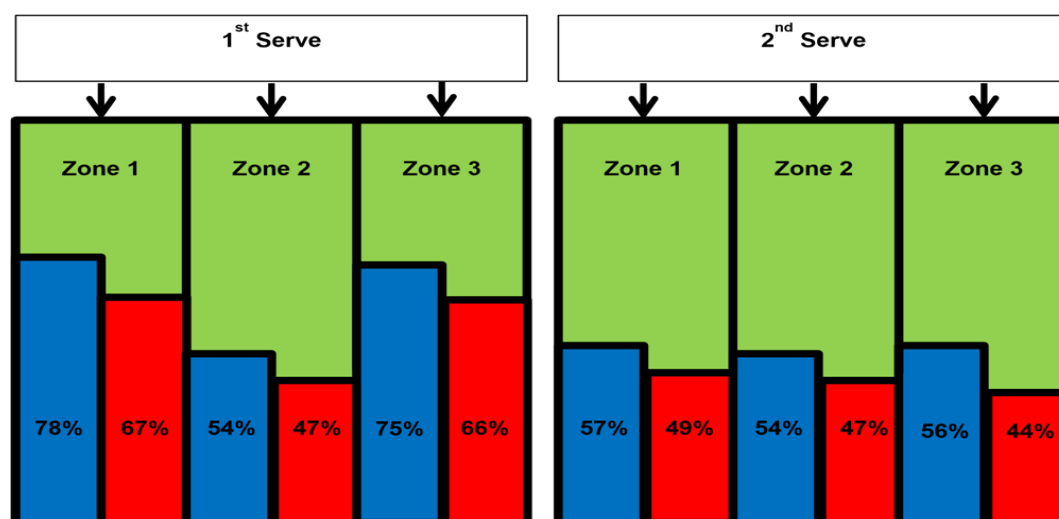


Fig. 13 "Probabilità da parte del battitore di vincere il punto in base alla zona di piazzamento del colpo" (in blu la probabilità per gli uomini, in rosso per le donne) Fonte: "Analysis of service placement within elite Tennis, E. Hurley, Cardiff school of sport

Dividendo ipoteticamente il rettangolo del servizio in tre ulteriori rettangoli uguali (Zona 1=esterno sinistro, Zona 2=Centrale, Zona 3=esterno destro) e analizzando le percentuali di vittoria del punto per zona di piazzamento del servizio, è possibile verificare due comportamenti molto diversi ma prevedibili per prima e seconda di servizio.

Se si serve la prima palla è infatti molto più probabile vincere il punto se il piazzamento avviene nelle zone esterne del rettangolo, mentre la probabilità di vincere il punto è simile in tutte le zone per la seconda.

Questo perché, come abbiamo già visto, i giocatori diminuiscono strategicamente la velocità tra prima e seconda di servizio e anche in caso di piazzamento angolato la velocità impressa alla seconda non è sufficiente per determinare un alto livello di efficacia.

E' dunque lecito affermare che il piazzamento sembra avere un impatto maggiore sulla resa soprattutto ad alte velocità, mentre a velocità basse avrà una minore influenza.

Ad ogni modo la scelta dei giocatori per quanto riguarda il piazzamento del servizio ricadrà pertanto principalmente sulle zone 1 e 3, più marcatamente sulla prima che sulla seconda come confermano anche gli istogrammi riportati nella figura 14, che rappresentano il numero di punti effettivamente giocati (in percentuale rispetto al totale) nel tour professionistico (settore maschile) per micro aree (fasce) di piazzamento del servizio, espresse come coordinate rispetto all'asse y, la larghezza, del campo di gioco (il rettangolo del servizio misura in larghezza circa 4 metri).

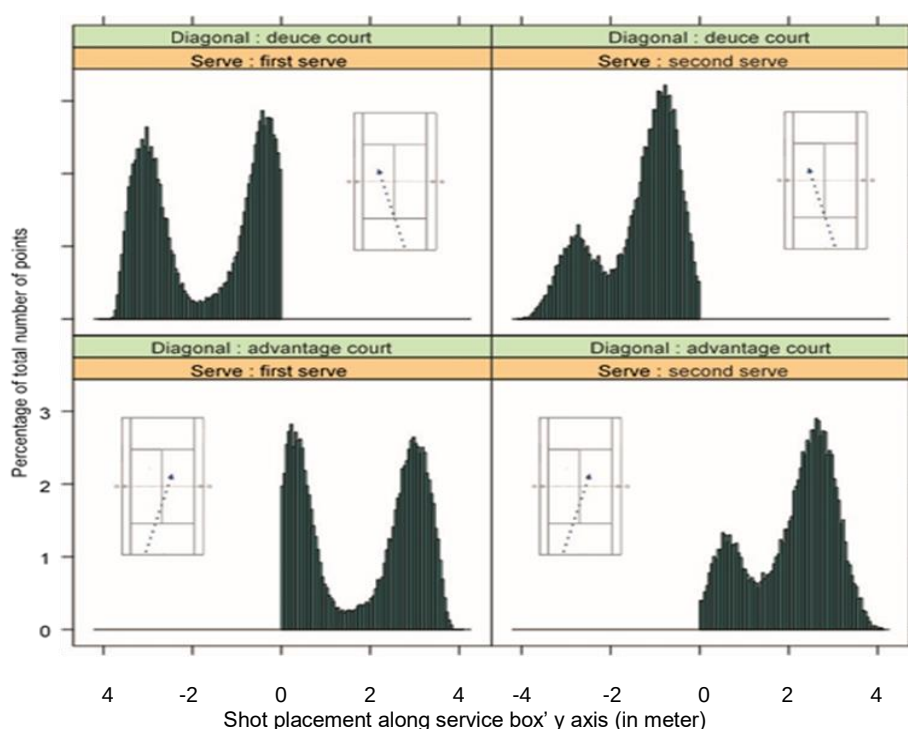


Fig.14 "Percentuale di servizi per fasce di piazzamento lungo l'asse y del rettangolo di servizio nel settore maschile". Fonte: (Mecheri, Rioult, Mantel, Kauffman, & Benguigui, 2016)

Sembrano infatti esserci due alternative principali per quanto riguarda la prima di servizio, ovvero cercare molto più frequentemente i lati esterni del rettangolo sulla prima, mentre per

quanto riguarda la seconda di servizio c'è un maggiore livellamento nelle scelte, anche se spicca la più alta concentrazione di colpi verso il lato sinistro del ribattitore: questo è dovuto ad un fattore tecnico, ovvero la ricerca del rovescio sul lato sinistro del ribattitore, ovvero il colpo generalmente meno efficace dei giocatori destrimani, che sono in netta maggioranza nel circuito professionistico.

LO SPIN

L'ultima variabile in analisi è lo spin, ovvero la rotazione della pallina attorno al proprio baricentro. Ci sono diversi tipi di spin a seconda dell'asse della pallina su cui avviene la rotazione, ma per rendere più semplice la comprensione si considera in questa sede un coefficiente di spin C'_L per indicare il livello di rotazione complessivo impresso alla pallina.

Si è detto che vi è un trade-off tra spin e velocità della pallina perciò, poiché si è dimostrato in precedenza che la velocità aumenta l'efficacia del colpo, ci si aspetterebbe che questa sia ricercata sempre e comunque. Bisogna però tornare al presupposto che sulla seconda palla non è strategicamente corretto cercare la velocità massima poiché il rischio di non colpire il rettangolo di servizio aumenta all'aumentare della velocità e, in caso di errore sulla seconda il punto è perso automaticamente.

I giocatori per aumentare l'efficacia (o quantomeno minimizzare l'inefficacia) del secondo servizio utilizzano generalmente una seconda battuta carica di spin. E' dunque la scelta corretta? Andiamo in tal senso ad analizzare le scelte dei giocatori e i risultati che ne conseguono (Figura 15)

	Serve	C'_L	Won	Lost	n
Women	1st	<0.07]	62.43	37.57	14279
Women	1st	(0.07,0.13]	61.78	38.22	13753
Women	1st	(0.13,0.18]	62.11	37.89	7187
Women	1st	>0.18	66.55	33.45	4359
Women	2nd	<0.07]	50.19	49.81	1574
Women	2nd	(0.07,0.13]	48.64	51.36	7010
Women	2nd	(0.13,0.18]	49.10	50.90	8913
Women	2nd	>0.18	51.22	48.78	5531
Men	1st	<0.07]	71.05	28.95	36960
Men	1st	(0.07,0.13]	68.73	31.27	24170
Men	1st	(0.13,0.18]	69.60	30.40	16748
Men	1st	>0.18	72.74	27.26	12606
Men	2nd	<0.07]	59.82	40.18	2103
Men	2nd	(0.07,0.13]	54.11	45.89	11212
Men	2nd	(0.13,0.18]	53.63	46.37	23270
Men	2nd	>0.18	54.06	45.94	20821

Fig.15 "Percentuale di punti vinti e persi e numero totale di punti giocati al servizio per fasce di spin".

Fonte: (Mecheri, Rioult, Mantel, Kauffman, & Benguigui, 2016)

I risultati in Figura 15 sembrano confermare le ipotesi fatte in precedenza: i giocatori tendenzialmente scelgono, di servire la prima di servizio cercando la massima velocità (in gergo si dice "servizio piatto") o comunque utilizzando un basso livello di spin (la percentuale di punti giocati in fascia minima di spin è nettamente maggiore rispetto alle altre fasce), mentre sulla seconda è quasi inutilizzato il servizio piatto mentre predomina la scelta di utilizzare un colpo carico di spin.

La resa percentuale del servizio aumenta tendenzialmente cercando di massimizzare una delle due componenti, quindi o massima velocità o alti livelli di spin (ma senza esagerare, abbassando la velocità sotto una certa soglia si ottiene l'effetto contrario). Compromessi intermedi sembrano diminuire la probabilità di vincere il punto o comunque non essere così efficaci.

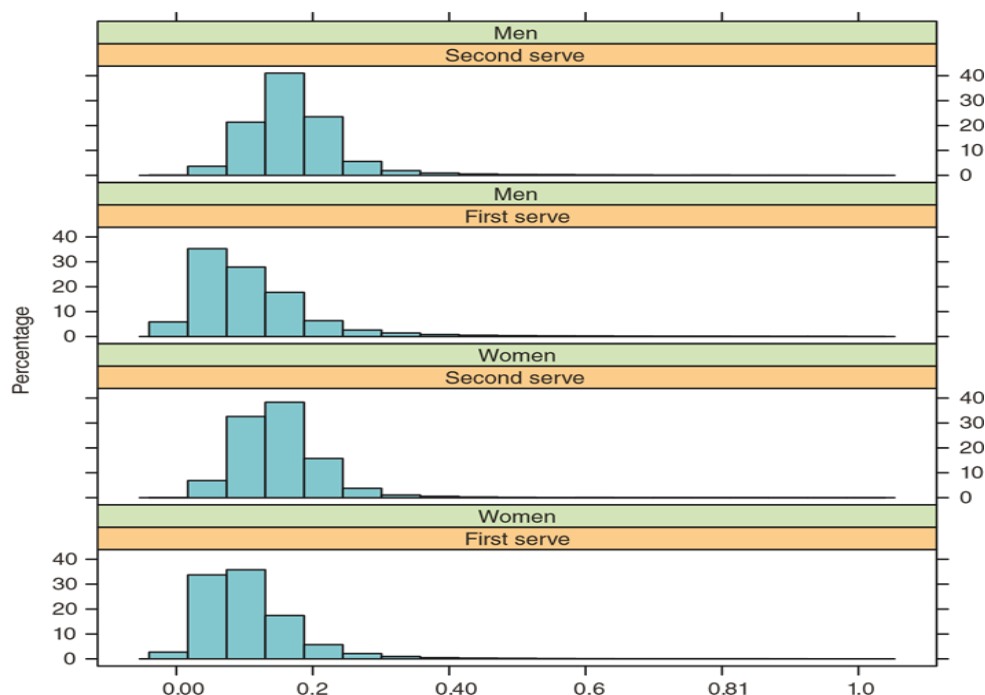


Fig.16 "Frequenza di punti giocati per fasce di spin del colpo di servizio". Sull'asse x il coefficiente di spin C'_L , sull'asse y la percentuale di punti effettuati.

Fonte: (Mecheri, Rioult, Mantel, Kauffman, & Benguigui, 2016)

Gli istogrammi di Figura 16 evidenziano la varietà comunque elevata di scelte riguardo alla quantità di spin utilizzata nel servizio (non tutti i giocatori usano il servizio piatto per la prima come non tutti massimizzano lo spin sulla seconda).

Viene spontaneo a questo punto chiedersi per quale motivo permanga questa varietà nelle scelte dei giocatori appena citate se si è appena concluso che tendenzialmente è più efficace utilizzare una delle due scelte estreme opposte.

Si possono dare risposte ipotetiche che provo a riassumere: innanzi tutto non tutti i giocatori riescono ad imprimere spin alti tali da massimizzare l'efficacia del colpo mantenendo comunque un livello di velocità minimo per competere ad alti livelli, quindi ci saranno players che si adatteranno su livelli di spin intermedi.

In secondo luogo anche l'imprevedibilità gioca il suo ruolo nell'aumentare l'efficacia dei colpi; questa è difficilmente quantificabile in questa sede di analisi, ma si può comunque affermare che è probabilmente più efficace variare di tanto in tanto il modo di colpire, utilizzando scelte sulla carta meno efficaci, in modo però da avere un apporto positivo dell'effetto sorpresa, necessario talvolta per accrescere le rese percentuali dei colpi.

3.2.D UN INDICATORE DIFFERENTE

Si è dunque capito che le variabili che determinano un buon colpo e quindi percentuali maggiori, sono velocità, piazzamento e spin. Riprendendo la volontà espressa in precedenza, quella di voler una proposta personale in merito all'adozione di un indicatore basato su queste variabili bisogna premettere che la costruzione e il calcolo di un indicatore di performance incentrato su queste caratteristiche sarebbe un compito piuttosto complesso, ma nemmeno impossibile pensando alle potenzialità delle tecnologie in uso e dei software di analisi.

Si potrebbe perciò pensare ad una valutazione sulla base di un punteggio per ognuna delle tre voci (costruendo un'apposita scala valori ovviamente), che assegni ad ogni servizio (con distinzione tra prima e seconda) un valore complessivo.

Mi spiego meglio. Le componenti di questo punteggio complessivo potrebbero essere:

1. **PUNTEGGIO VELOCITA' (Speed Index):** Andrebbe studiato come varia la velocità del colpo in relazione al piazzamento del colpo nel rettangolo del servizio, ponderando dunque l'effettiva velocità con un coefficiente che tenga conto delle aree dove è più difficile piazzare la palla mantenendo elevata la velocità (solitamente gli angoli stretti).
2. **PUNTEGGIO PIAZZAMENTO (Accuracy Index):** Come abbiamo visto, ci sono delle aree che se colpite, consentono con maggior probabilità di ottenere il punto. Si potrebbe dividere il rettangolo in aree ancora più piccole e assegnare un punteggio a ciascuna in relazione alla percentuale statistica di ottenere il punto colpendo quella "mattonella". Le aree ad elevata probabilità di punto offrirebbero in dote un punteggio maggiore. L'Accuracy Index del colpo sarebbe quindi determinato dal punteggio relativo all'area in cui viene piazzato il colpo. Potrebbe questo punteggio essere poi anche ponderato con la percentuale di servizi in campo.
3. **PUNTEGGIO SPIN (Spin Index):** Un semplice coefficiente di spin, come quello preso in considerazione in precedenza (basato sulle rivoluzioni per minuto, RPM, della pallina).

La somma di questi tre punteggi determinerebbe quindi un **PUNTEGGIO DEL SERVIZIO (Service Index)**, che verrebbe assegnato ad ogni singolo colpo e potrebbe costituire un valido

indicatore di performance dal momento che è oggettivo ed è svincolato da qualsiasi influenza di altre variabili.

Si potrebbe poi fare una media dei punteggi di ogni colpo per ogni giocatore, nell'arco di un solo match o per un periodo prolungato, consentendo dunque di conferire una certa oggettività a considerazioni tipiche del tennis come "il giocatore A in questo match ha servito male" oppure "il giocatore B ha un ottimo servizio".

3.2.E CARATTERISTICHE FISICHE DEL GIOCATORE: LA STATURA

Ora che si sono analizzate le variabili che determinano l'efficacia del colpo, viene da chiedersi se esistano precise caratteristiche fisiche dei giocatori che rivestano un ruolo chiave nel conferire efficacia al colpo.

E' abbastanza semplice ipotizzare come l'altezza dei giocatori possa costituire una variabile determinante nel consentire di sviluppare un colpo più efficace poichè maggiore è l'altezza del giocatore, maggiore sarà l'altezza del punto di impatto con la pallina da terra e di conseguenza sarà più vasto l'angolo "di visuale" del rettangolo di battuta, in modo da consentire una maggiore ricerca di velocità a parità di rischio.

Vengono dunque di seguito riportate le relazioni che intercorrono tra l'altezza dei giocatori e diverse statistiche sul servizio, rifacendomi ad un altro studio di Tennisabstract²⁹.

CORRELAZIONE TRA ALTEZZA E SERVIZIO

Le supposizioni appena fatte vengono confermate. Attraverso la rappresentazione grafica di Figura 17 su può capire come i giocatori più alti siano più efficaci al servizio, con un divario enorme, poichè si spazia dal 60% dei punti vinti al servizio per i giocatori più bassi (il limite del campione è 173 cm) fino a circa il 70% per il più alti (il limite superiore è 196 cm).

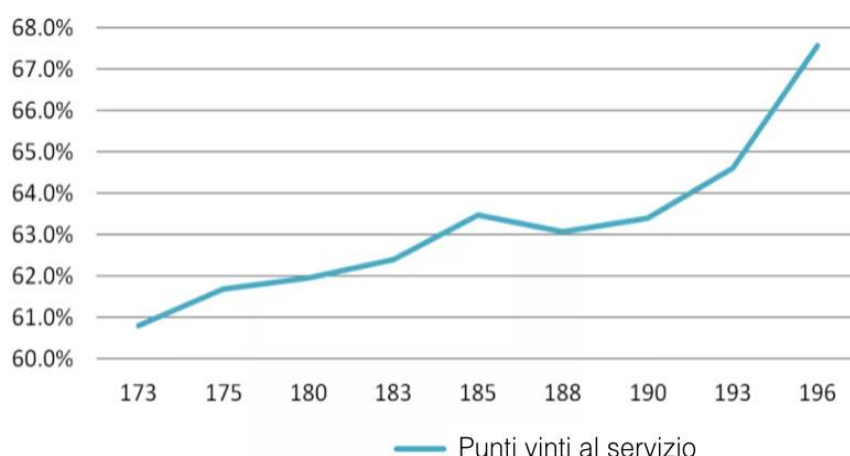


Fig. 17 "Relazione tra statura del giocatore (asse x) e percentuale di punti vinti al servizio (asse y)"
Fonte: How Much Does Height Matter in Men's Tennis?, 2017, J.Sackmann, Tennisabstract.com

²⁹ How Much Does Height Matter in Men's Tennis?, 2017, J.Sackmann, Tennisabstract.com

Per quanto si possa evidenziare una chiara relazione tra queste due variabili, la relazione tra altezza e frequenza di ace (numero di aces / numero di servizi totali) è ancora più forte (Figura 18), con un r^2 pari a 0.83 (coefficiente di determinazione):

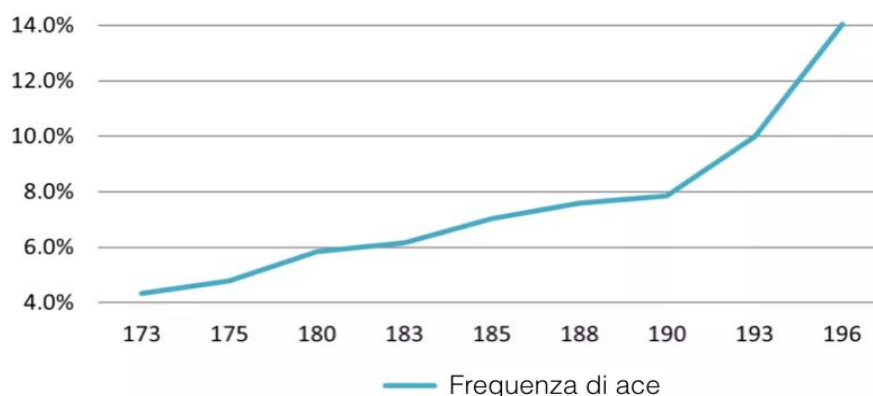


Fig. 18 "Relazione tra statura del giocatore (asse x) e frequenza di ace ottenuti (asse y)"
Fonte: "How Much Does Height Matter in Men's Tennis?", 2017, J.Sackmann, Tennisabstract.com

La frequenza di aces però non rappresenta la cartina di tornasole più accurata per questo tipo di analisi, poiché solitamente essi vengono cercati e ottenuti solo con la prima palla di servizio e sono comunque un campione limitato dei punti vinti in questo frangente. Infatti la statistica con la correlazione più forte rispetto all'altezza è il numero di punti vinti con la prima di servizio ($r^2 = 0.92$), come mostrato dal grafico sottostante (Figura 19).

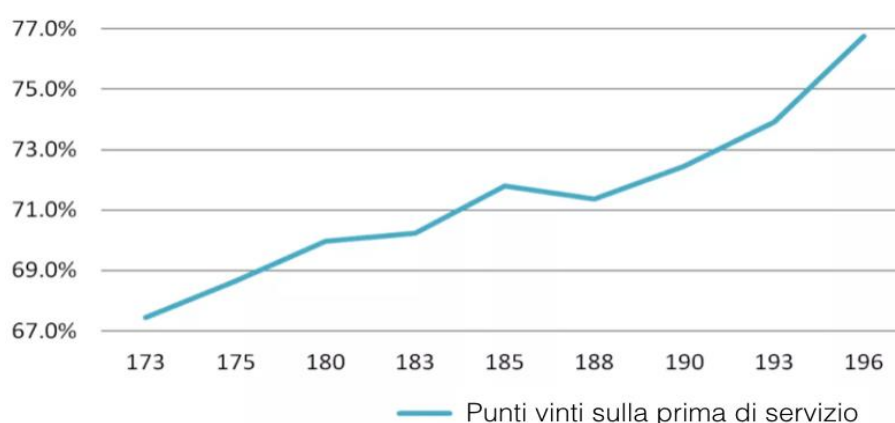


Fig 19. "Relazione tra statura del giocatore (asse x) e punti vinti sulla prima di servizio (asse y)"
Fonte: "How Much Does Height Matter in Men's Tennis?", 2017, J.Sackmann, Tennisabstract.com

Per quanto riguarda la seconda di servizio (Figura 20) vi è ancora una modesta correlazione con la statura ($r^2 = 0.18$), ma è la più debole tra tutte le statistiche presentate in quest'analisi.

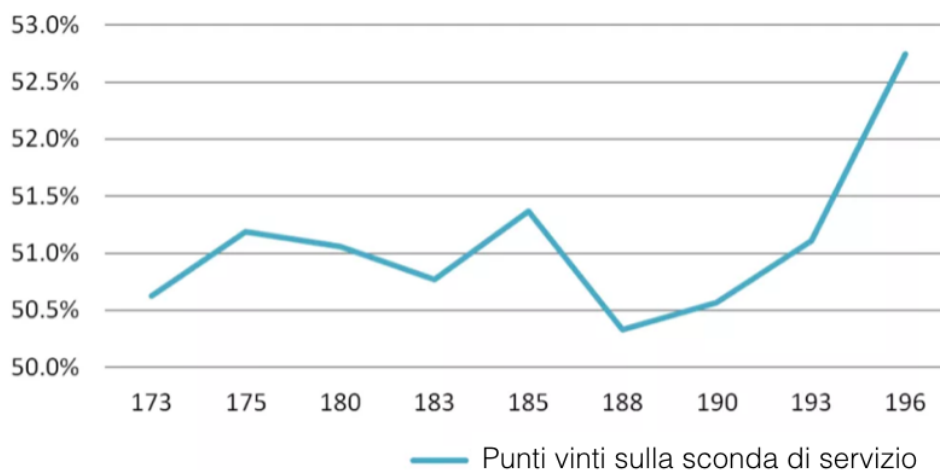


Fig 20 "Relazione tra statura del giocatore (asse x) e punti vinti sulla seconda di servizio (asse y)"
 Fonte: "How Much Does Height Matter in Men's Tennis?", 2017, J.Sackmann, Tennisabstract.com

Essere alti è dunque un vantaggio per quanto riguarda il servizio, anche se con importanti differenze tra prima e seconda di servizio.

Mentre sulla prima il vantaggio è indiscutibilmente evidente, sulla seconda palla di servizio le i risultati sono differenti: i giocatori alti 193 cm vincono infatti praticamente lo stesso numero di punti sulla seconda di servizio rispetto ai giocatori alti 175 cm.

Questo potrebbe significare che la seconda di servizio dei giocatori bassi a livello qualitativo si avvicina quella dei giocatori più alti, ma più probabilmente rispecchia il fatto che, anche con differenze marcate sul colpo in sé, i giocatori di statura inferiore possono generalmente fare affidamento su una predisposizione maggiore ad affrontare punti giocati sulla seconda di servizio, poichè normalmente hanno una durata maggiore e, durante lo scambio prolungato, emergono altre qualità. É però comunque difficile stabilirle quale delle due spiegazioni sia maggiormente pertinente.

In figura 20 si evidenzia come in ogni caso l'efficacia sulla seconda per i giocatori più alti del campione tenda comunque ad impennarsi ugualmente. Quest'ultimo fatto potrebbe conseguenza dell'utilizzo di una particolare strategia che verrà di seguito spiegata.

3.3 LA STRATEGIA DELLE DUE PRIME

A questo punto, riassumo quello che si è dimostrato e concluso finora

1. La strategia di giocare la prima di servizio ricercando la massima velocità e la seconda di servizio diminuendo la velocità, aumentando al contempo lo spin, è universalmente riconosciuta come la più corretta, nonché quella solitamente utilizzata per quanto riguarda il tennis professionistico (e non solo).
2. I giocatori alti tendono ad avere un vantaggio competitivo notevole rispetto ai colleghi di statura inferiore, per quanto riguarda le percentuali di vittoria del punto con la prima palla di servizio in campo, mentre per quanto riguarda la seconda di servizio questo vantaggio sembra svanire.

A partire da queste due considerazioni, sorgono spontanei quindi dei quesiti: in primis la strategia di differenziare prima e seconda e la più proficua? Determinate tipologie di giocatori potrebbero invece adottare diversi comportamenti strategici? Si può fare una previsione a partire dai dati in possesso sulla potenziale efficacia di una diversa strategia?

Vi è infatti un'alternativa a questo schema consolidato che consiste nel giocare sulle percentuali di resa e decidere quindi di giocare prima e seconda palla di servizio allo stesso modo, come se fossero entrambe una prima di servizio. Questa strategia viene denominata come “strategia delle due prime”.

In questo modo, giocando una seconda “aggressiva” aumenterà la percentuale di punti ottenuti, comportamento che però d'altro canto potrà essere compensato in negativo dall'aumentare della probabilità di commettere un “doppio fallo”, che consegnerà direttamente il punto all'avversario.

ESPRIMERE STATISTICAMENTE LA “STRATEGIA DELLE DUE PRIME”

Sulla seconda di servizio la percentuale complessiva di resa ottenuta con quella che verrà denominata “strategia tradizionale”, è determinata da:

$$\% \text{punti vinti sulla seconda (Z)} = \frac{(\% \text{ seconde in campo (X)} * \% \text{ punti con la seconda vinti quando la seconda è in campo (Y)})}{100}$$

Adottando la “strategia delle due prime”, la resa complessiva della seconda è però ora diversamente determinata. Facendo una sommaria analisi di sensitività (Figura 21), si intuisce come rispetto alla “strategia tradizionale” non si possa sapere con certezza se la resa complessiva sarà migliore o peggiore

	STRATEGIA TRADIZIONALE	STRATEGIA DUE PRIME
% SECONDE IN CAMPO	X	SCENDE
% PUNTI VINTI CON LA SECONDA IN CAMPO	Y	SALE
% PUNTI VINTI CON LA SECONDA	Z	?

Fig.21 Analisi di sensitività per le variabili che determinano la percentuale di punti vinti con la seconda di servizio per la “strategia delle due prime”

Bisogna dunque provare a costruire, a partire dai dati in possesso, la resa teorica dei giocatori sulla seconda utilizzando la “strategia delle due prime”.

Quello che differenzia le due strategie ovviamente non riguarda le percentuali di resa della prima (la percentuale di prime in campo e quella di punti vinti sulla prima di servizio rimangono le stesse), ma solo quelle sulla seconda, per cui avremo una nuova percentuale di resa complessiva. Questo dato non sarà più “a sé stante”, ma sulla carta potrebbe essere calcolato (a partire dai dati sulla prima) utilizzando la seguente formula:

$$\% \text{ punti vinti sulla seconda} = \frac{(\% \text{ punti vinti quando la prima è in campo} * \% \text{ prime in campo})}{100}$$

Sulla base di questa formula si può calcolare per ognuno dei giocatori nel database ATP Leaderboard menzionato in precedenza, quale sarebbe la resa teorica utilizzando la “strategia delle due prime”, confrontarla con la “strategia tradizionale e ottenere nell’ultima colonna a destra il differenziale ipotetico a favore (o contro) della strategia delle due prime (Figura 22).

GIOCATORE	% PRIME IN CAMPO	%PUNTI VINTI SULLA PRIMA	%PUNTI VINTI SULLA SECONDA	% PUNTI VINTI SULLA SECONDA "STRATEGIA DELLE DUE PRIME"	DIFFERENZIALE TRA STRATEGIE
ISNER	71,1%	79,1%	55,9%	56,2%	0,3%
RAONIC	62,5%	83,6%	56,5%	52,3%	-4,3%
OPELKA	64,0%	79,3%	53,3%	50,8%	-2,5%
KYRGIOS	67,7%	78,5%	50,5%	53,1%	2,6%
FEDERER	64,8%	77,1%	58,2%	50,0%	-8,2%
BERRETTINI	61,4%	78,9%	54,1%	48,4%	-5,7%
NADAL	64,2%	75,2%	58,6%	48,3%	-10,3%
QUERREY	60,7%	80,1%	47,6%	48,6%	1,0%
TSITSIPAS	64,3%	75,9%	54,4%	48,8%	-5,6%
TSONGA	61,1%	79,4%	50,9%	48,5%	-2,4%

Fig. 22 “Percentuali di punti vinti con la seconda di servizio con entrambe le strategie presentate e differenziale di resa tra le strategie”. (Vengono mostrati solo i primi 10 giocatori per “Serve rating ATP index”).
Elaborazione personale dei dati ATP leaderboard.

Ci sono dunque giocatori che sulla carta addirittura trarrebbero beneficio dall’utilizzo di questa strategia, migliorando la propria percentuale di punti vinti con la seconda, mentre la stragrande maggioranza peggiorerebbe anche piuttosto nettamente le proprie percentuali (più di 10 punti percentuali di resa in meno). Allargando a tutti gli altri giocatori nel database e non ai soli primi 10 del ranking basato sul rating del servizio si nota inoltre come solo 5 atleti su 87 trarrebbero un vantaggio mentre gli altri 82, vedrebbero le percentuali in diminuzione. La media complessiva del differenziale per atleta tra le due strategie è di -6,4%.

Questa elaborazione offre due chiavi di lettura:

1. É dunque confermato per la nettissima maggioranza dei giocatori che quello che abbiamo assunto come assioma, ovvero la strategia “tradizionale” di differenziare nell’esecuzione le due battute, è la decisione corretta
2. Vi è invece una minoranza di giocatori che sulla carta beneficerebbe della “strategia delle due prime”

La domanda che ci si pone a questo punto è: c’è qualche fattore che accomuna questi giocatori e che spiega questo assunto?

La risposta è anche piuttosto semplice e potrebbe infatti essere ricollegata a quanto esposto in precedenza, ovvero il fattore “statura”.

GIOCATORE	% PRIME IN CAMPO	%PUNTI VINTI SULLA PRIMA	%PUNTI VINTI SULLA SECONDA "STRATEGIA TRADIZIONALE"	%PUNTI VINTI SULLA SECONDA "STRATEGIA DELLE DUE PRIME"	DIFFERENZIALE TRA LE DUE STRATEGIE	ALTEZZA GIOCATORE
ZVEREV	67,6%	75,3%	43,7%	50,9%	7,2%	1,98
KYRGIOS	67,7%	78,5%	50,5%	53,1%	2,6%	1,93
HARRIS	64,7%	74,0%	46,5%	47,9%	1,4%	1,93
QUERREY	60,7%	80,1%	47,6%	48,6%	1,0%	1,98
ISNER	71,1%	79,1%	55,9%	56,2%	0,3%	2,08
MONFLIS	66,7%	71,1%	48,9%	47,4%	-1,5%	1,93
TSONGA	61,1%	79,4%	50,9%	48,5%	-2,4%	1,88
OPELKA	64,0%	79,3%	53,3%	50,8%	-2,5%	2,11
POSPISIL	60,6%	77,4%	49,6%	46,9%	-2,7%	1,91
GERASIMOV	65,2%	74,5%	51,5%	48,6%	-2,9%	1,93

Fig. 23. “I dieci giocatori più avvantaggiati (o meno svantaggiati) ipoteticamente dall’utilizzo della strategia presentata”

GIOCATORE	% PRIME IN CAMPO	%PUNTI VINTI SULLA PRIMA	%PUNTI VINTI SULLA SECONDA "STRATEGIA TRADIZIONALE"	%PUNTI VINTI SULLA SECONDA "STRATEGIA DELLE DUE PRIME"	DIFFERENZIALE TRA LE DUE STRATEGIE	ALTEZZA GIOCATORE
MILLMAN	59,4%	68,9%	53,9%	40,9%	-13,0%	1,83
LONDERO	60,9%	69,2%	55,0%	42,1%	-12,9%	1,80
NISHIOKA	66,7%	63,9%	55,1%	42,6%	-12,5%	1,70
POUILLE	57,4%	71,8%	53,2%	41,2%	-12,0%	1,85
PAUL	59,8%	69,0%	53,1%	41,3%	-11,8%	1,85
FOGNINI	58,9%	69,2%	51,9%	40,8%	-11,1%	1,78
KOHLSCHEIBER	60,8%	69,6%	52,9%	42,3%	-10,6%	1,78
CARUSO	60,2%	65,6%	50,1%	39,5%	-10,6%	1,83
BASILASHVILI	57,6%	67,4%	49,4%	38,8%	-10,6%	1,85
NADAL	64,2%	75,2%	58,6%	48,3%	-10,3%	1,85

Fig.24 “I dieci giocatori più svantaggiati ipoteticamente dall’utilizzo della strategia presentata”

Stilata infatti una classifica dei giocatori per differenziale tra le due strategie (a favore della “strategia delle due prime”) e presi in considerazione i primi 10 (Figura 23) e gli ultimi 10 (Figura 24) giocatori in questa classifica si può notare che:

1. La media di statura dei dieci giocatori con più benefici (o minor svantaggi) ipotetici è 196 cm.
2. Al contrario la media di statura dei dieci giocatori con più svantaggio ipotetico è 181 cm.

Se si paragonano queste due valori alla media della statura del campione in considerazione che è di 187 cm, sembrerebbe dunque confermarsi l'ipotesi che chi ha una minima possibilità di impiegare con successo questa strategia è proprio la categoria dei giocatori di maggior statura. Inoltre il coefficiente di correlazione lineare tra le due variabili ammonta a 0,62, un valore che sottolinea inconfutabilmente come esista un legame positivo tra le due variabili: la resa ipotetica della "strategia delle due prime" aumenterebbe chiaramente all'aumentare dell'altezza dei giocatori.

Questo perché gli atleti appartenenti alla suddetta categoria, in virtù del solitamente (confermato dalle statistiche del paragrafo 3.2) elevato differenziale di rendimento tra prima e seconda palla di servizio, lasciano una porta aperta ad un potenziale miglioramento delle rese sulla seconda palla, non tramite il miglioramento qualitativo del colpo in sé, ma semplicemente utilizzando una differente strategia.

Dunque solo i giocatori che presentano mediamente un grande gap di efficacia tra prima e seconda palla possono trarre beneficio da questa diversa scelta, ed è proprio questa la motivazione dell'impennata della resa sulla seconda di servizio evidenziata in Figura 20. I giocatori più alti infatti, consapevoli di quanto appena detto, nella realtà utilizzano già in parte (solo in determinate situazioni) questa strategia ed è questo il motivo per cui può esserci un differenziale teorico a favore della "strategia delle due prime", ma è di rado elevato (la resa teorica sarà sempre più alta di quella ottenuta all'atto pratico). L'utilizzo della Data Analytics sta infatti creando consapevolezza tra i giocatori, donando gli strumenti per effettuare valutazioni sulle strategie e sulle scelte da effettuare in campo.

In un certo senso si è operata, grazie all'approccio data-driven, una "segmentazione" dei giocatori per arrivare a stabilire quale possa essere in linea di massima la strategia migliore per ognuno dei due segmenti.

4.TENNIS DATA ANALYTICS: IL PUNTEGGIO

4.1 PREMESSE

Questo capitolo è un estratto dell'elaborato redatto in qualità di progetto sostitutivo dello stage curricolare.

Il lavoro che ho svolto ha implicato la creazione di un database in Microsoft Excel contenente la totalità dei match giocati al massimo livello professionistico (ATP/ITF per gli uomini e WTA/ITF per le donne) durante l'anno solare 2020.

Nel dettaglio si tratta della raccolta delle informazioni di livello base dei 3754 incontri svoltisi durante la suddetta stagione. Per livello base si intende che, per ogni match, è stato esplicitato il sesso dei giocatori, la superficie di gioco, il cosiddetto turno (fase dello stato di avanzamento del torneo in cui è stata giocata la partita), i nomi dei giocatori e la loro relativa nazionalità, il punteggio (diviso per set) e il luogo geografico in cui si è svolto il torneo.

L'obiettivo è stato quello di avere una base di dati sufficientemente ampia al fine di garantire analisi statistiche descrittive abbastanza accurate sui punteggi dei match presi in considerazione, in modo da carpire le peculiarità generali dell'andamento di un match di tennis.

L'OGGETTO PRINCIPALE DELL'ANALISI: IL PUNTEGGIO

Si è affrontato nel capitolo 3 il discorso degli indicatori di performance correlati ai fattori critici nel tennis. Per introdurre la tematica del punteggio di un match di tennis, oggetto di analisi di questo progetto, si può affermare che proprio il punteggio si può appunto considerare come un ulteriore indicatore di performance, questa volta non correlato necessariamente ad un solo fattore critico, ma più che altro alla prestazione espressa globalmente in un determinato match.

La problematica del considerare il punteggio un indicatore di performance risiede però nel fatto che più di ogni altro parametro, esso è influenzato, se ci si pone nella prospettiva di un

giocatore, dalla bravura, dall'efficacia nel contesto di gioco considerato e dalla prestazione dell'avversario.

Inoltre poiché il punteggio finale è soltanto l'apice della gerarchia del sistema di punteggio tennistico, spesso non consente di determinare la bontà della prestazione dei giocatori in campo con accuratezza. Oltretutto è addirittura possibile che un giocatore vinca un incontro di Tennis ottenendo meno punti e meno games rispetto all'avversario (mentre non è possibile vincere meno set, poiché sui set si basa il punteggio finale del match).

Il punteggio ad ogni modo è l'indicatore che probabilmente più si presta ad indagini statistiche senza dover per forza scomodare competenze tecniche e la conoscenza del gioco, perciò anche questo argomento offre potenzialmente spunti molto interessanti.

Facendo un parallelismo alla gestione delle imprese, si potrebbe accostare il punteggio al concetto di fatturato, il quale consente di determinare il giro di affari di un'impresa ma è da porre in relazione con altre variabili quali il settore di riferimento in cui l'azienda opera, gli investimenti sostenuti o le eventuali perdite derivanti dalla gestione operativa.

4.2 PRESENTAZIONE DELLE ANALISI EFFETTUATE

Nell'anno solare 2020 sono stati giocati 58 tornei, 34 per il settore maschile e 24 per quello femminile, per un totale di 3754 match, di cui 3679 conclusi regolarmente, 74 conclusi per ritiro di un giocatore e un solo match concluso per squalifica di uno dei due atleti in campo. Le superfici di gioco utilizzate durante la stagione sono state 3 (abituamente, in anni non pandemici sarebbero state invece 4): Terra battuta, Cemento Outdoor e Cemento Indoor.

CONSIDERAZIONI PRELIMINARI: DIVERSE FORMULE REGOLAMENTARI

Nel tennis professionistico esistono due differenti formule regolamentari per quanto riguarda il punteggio: una partita si può infatti concludere al "al meglio dei 3 set" (definita in gergo 2 su 3), ovvero vince il primo giocatore che conquista 3 set, o "al meglio dei 5 set" (in gergo 3 su 5), dove vince il primo giocatore che ottiene 3 set.

Nella totalità delle partite del settore femminile e nella larga maggioranza di quelle del settore maschile viene utilizzata la formula "2 su 3", mentre in una minoranza dei tornei maschili, quelli di maggiore importanza (i cosiddetti major o tornei del grande slam), viene utilizzata la "3 su 5".

In particolare nel 2020 sono state giocate e terminate per il settore maschile 1563 partite con la formula 2 su 3 e 373 partite con la formula 3 su 5.

4.2.A MATCH 2 SU 3 – ESITO DELLA PARTITA

Per quanto riguarda la formula 2 su 3, si possono identificare logicamente solamente due possibili risultati dell'incontro dal punto di vista del vincitore: due set a zero ("2-0") o due set a uno ("2-1"). Si può quindi affermare che in una partita 2 su 3 saranno giocati con certezza i primi due set: in caso di vittoria di entrambi da parte di un giocatore la partita è conclusa, mentre in caso di punteggio di 1-1 si giocherà il terzo e decisivo set.

E' implicito inoltre che dal punto di vista del perdente esistano anche le combinazioni di punteggio "speculari" 0-2 e 1-2, ma dal momento che non viene considerato il punto di vista specifico di un solo giocatore, ma solamente l'esito dell'incontro in sè, sarebbe una ripetizione considerarle in sede di analisi.

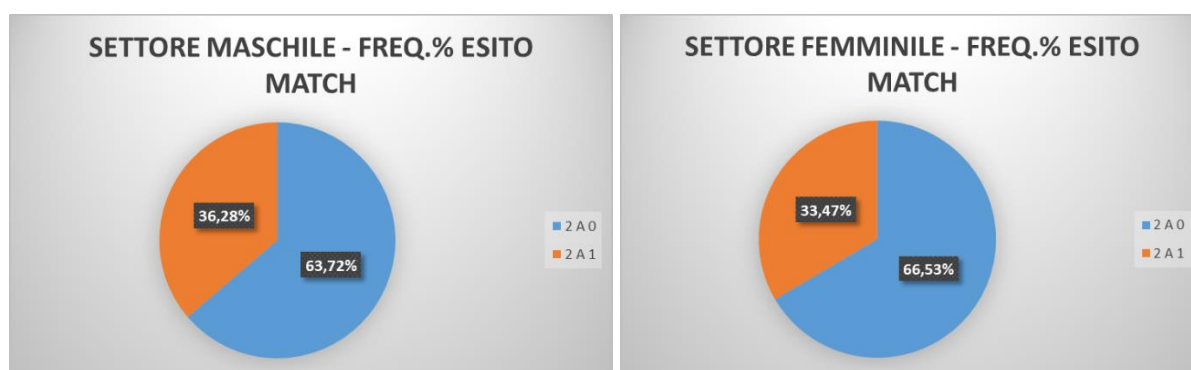


Fig.25 (a destra) e Fig.26 (a sinistra): "Frequenze percentuali dei possibili esiti di un match 2 su 3 nell'anno solare 2020"

Per quanto riguarda il settore maschile (Figura 25) il 63,72% dei match si è concluso con un punteggio di due set a zero, mentre il 36,28% con un punteggio di due set a uno. Nel settore femminile invece (Figura 26) nel 66,53% dei casi si sono svolti match da 2-0 mentre nel 33,47% dei casi la partita è terminata al terzo e decisivo set.

Sembra quindi esserci una frequenza maggiore di partite combattute e quindi un maggiore equilibrio dei valori in campo, terminate per 2 a 1 nel settore maschile (36,3% vs 33,4%), mentre un maggior numero di partite vinte con un punteggio netto (2 a 0) in quello femminile, che significa generalmente differenze più marcate tra le giocatrici in campo. La differenza tra maschi e femmine da questo punto di vista è comunque piuttosto ridotta.

4.2.B MATCH 2 SU 3 – ANDAMENTO DELLA PARTITA

Detto dei possibili esiti finali di un match 2 su 3, si possono ora compiere delle distinzioni in base a in che modo è maturato il risultato, ovvero considerare l'andamento dell'incontro.

Una vittoria per 2 a 0 può infatti essere ottenuta logicamente soltanto conquistando due set di fila, mentre una vittoria per 2 a 1 può essere conquistata in due modi differenti: Identificando con la sigla "W" i set vinti e "L" i set persi, si individuano dunque i 3 possibili andamenti di una partita vinta nella formula 2 su 3:

1. W – W (Corrispondente ad un punteggio di 2-0)
2. W – L – W (Corrispondente ad un punteggio di 2-1)
3. L – W – W (Corrispondente ad un punteggio di 2-1)

Analogamente a quanto affermato in precedenza se ci si ponesse nella prospettiva del giocatore perdente, si potrebbero considerare i 3 andamenti "speculari" con cui un incontro può essere perso:

1. L – L (0-2)
2. L – W – L (1-2)
3. W – L – L (1-2)

Essendo per natura due facce della stessa medaglia, di nuovo ci si porrà solamente nella prospettiva del vincitore per calcolare le frequenze percentuali con cui si verificano le citate situazioni.

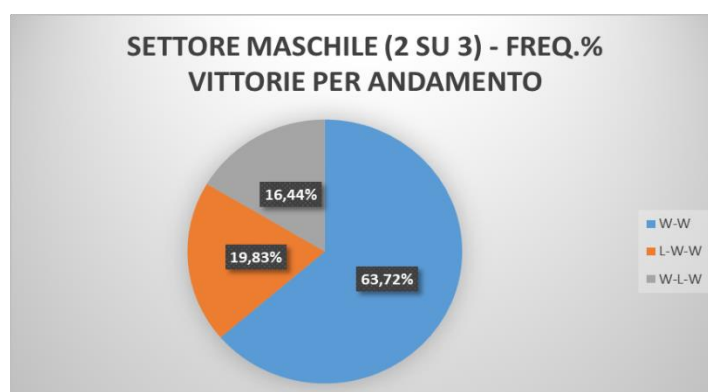


Fig.27: "Frequenze percentuali con cui si verificano i possibili andamenti di un match giocato con formula 2 su 3 nel settore maschile"

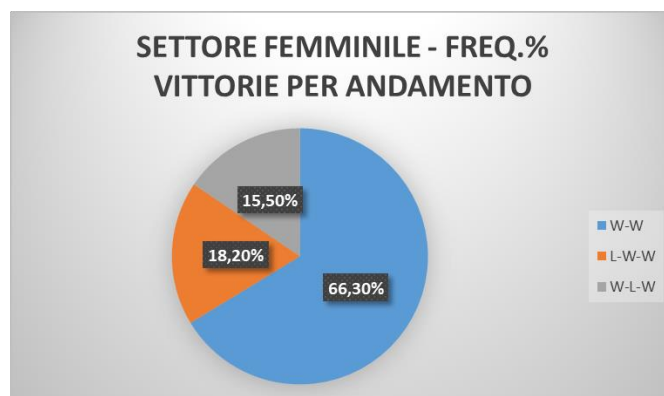


Fig.28: "Frequenze percentuali con cui si verificano i possibili andamenti di un match nel settore femminile"

Come si evidenzia dai grafici riportati in Figura 27 (settore maschile) e Figura 28 (settore femminile) la sequenza W-W, ovvero l'unica che corrisponde ad una vittoria per due set a zero è logicamente rappresentata dalle medesime percentuali del punteggio che rappresenta. Per quanto riguarda invece le partite concluse al terzo set, su può notare che la sequenza L-W-W (vittoria due set a uno avvenuta in rimonta), si verifichi più frequentemente che una vittoria ottenuta con un andamento W-L-W (18,2 % contro 15,5 % nelle donne e 19,83 % contro 16,44 % negli uomini).

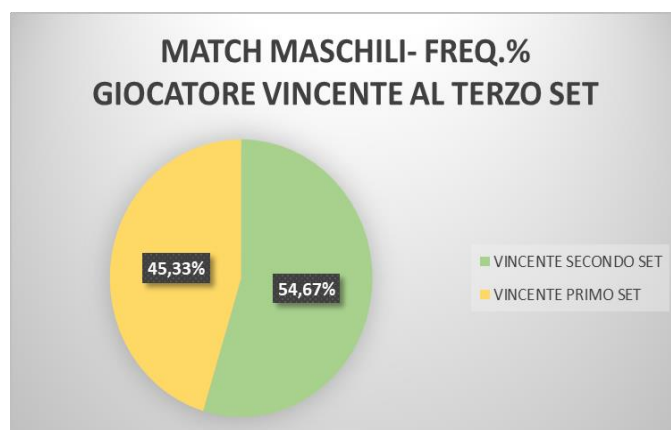


Fig. 29: "Frequenze percentuali di vittorie al terzo set per andamento della partita nel settore maschile"

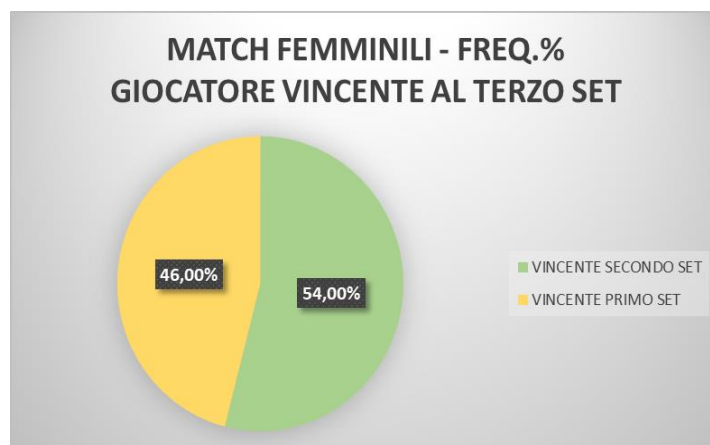


Fig.30: "Frequenze percentuali di vittorie al terzo set per andamento della partita nel settore femminile"

Nei grafici a torta in Figura 29 (settore maschile) e Figura 30 (settore femminile) si evidenzia come, conseguentemente a quanto appena affermato, in caso di partite decise al terzo set si verifichino leggermente più frequentemente vittorie sia nel settore maschile (54,67%) che in quello femminile (54,0%), del giocatore in rimonta che, grazie alla conquista del secondo parziale dopo aver perso il primo, allunga la partita al terzo set.

Risulta invece in entrambi i settori più difficile invertire nuovamente la tendenza dell'incontro e vincere il terzo set e quindi la partita dopo aver perso il secondo (45,33% uomini e 46,0%). Si può quindi affermare che in entrambi i settori, in caso di partita al terzo set, il favorito risulta essere il giocatore che si è aggiudicato il secondo set.

A. ANALISI VITTORIE W-W (2-0)

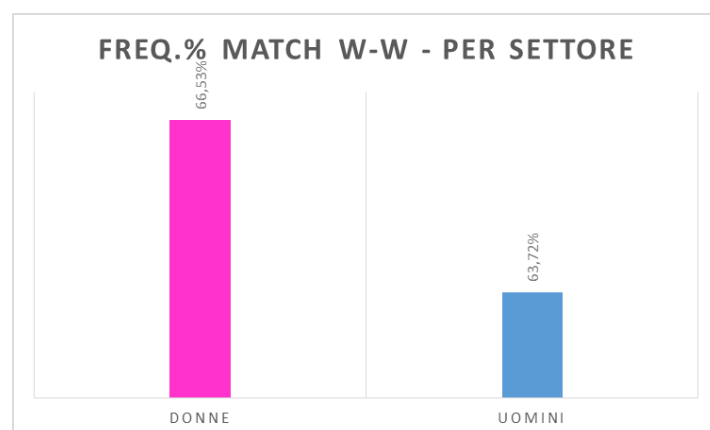


Fig.31: "Frequenze percentuali con cui si verifica l'andamento W-W nell'anno solare 2020"

Si è detto, come illustrato in Figura 31, che i match terminati per due set a zero (sequenza W-W) rappresentano circa due terzi dei match giocati (66,53% donne, 65,72% uomini).

Si sono a questo punto effettuate distinzioni per quanto riguarda sia la fase di gioco del torneo in cui sono avvenuti i match ("il turno"), che per quanto riguarda la superficie di gioco.

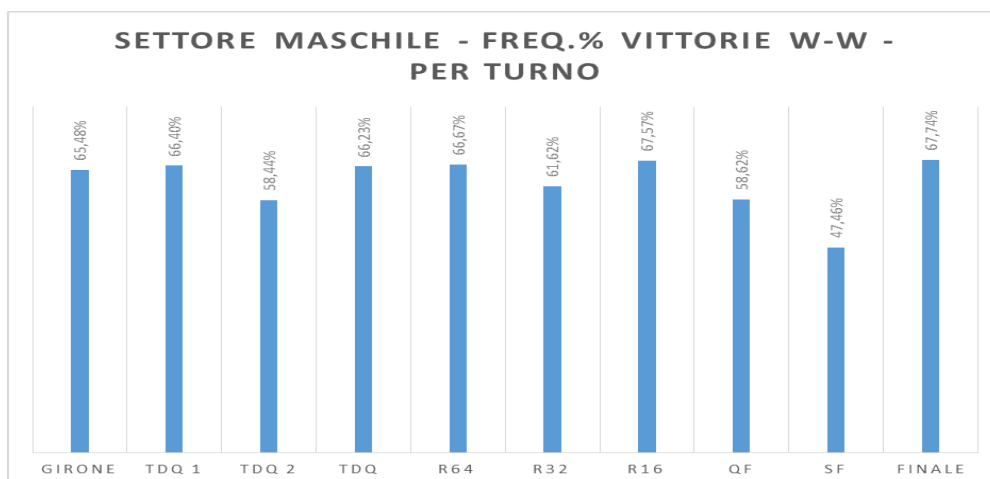


Fig.32 "Frequenze percentuali con cui si verifica l'andamento W-W nel settore maschile per turno"

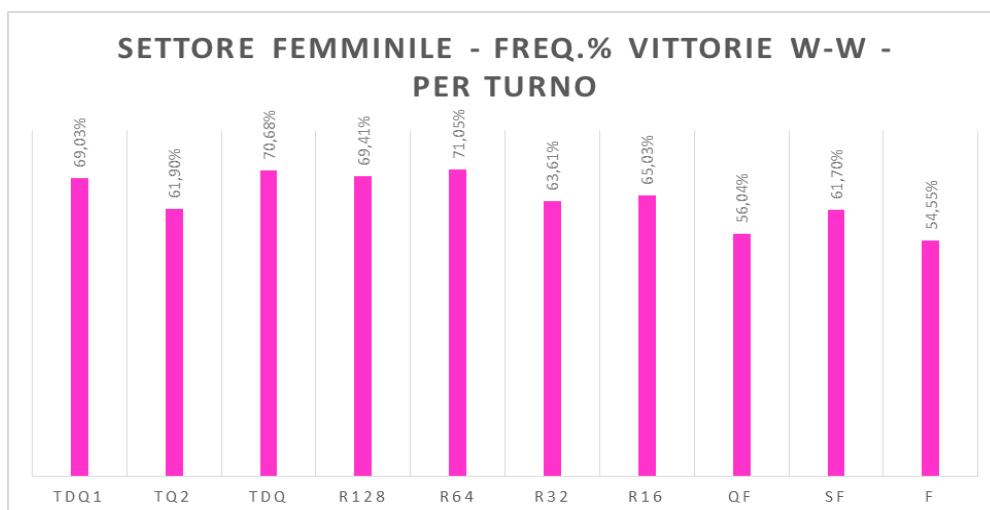


Fig.33 "Frequenze percentuali con cui si verifica l'andamento W-W nel settore femminile in base al turno"

Per quanto riguarda il settore maschile (Figura 32) non si rilevano tendenze particolari e il tasso di vittorie W-W (o anche vittorie per 2 set a 0) rimane sostanzialmente il medesimo,

anche se spicca il dato riguardo al turno di Semifinale (46,67%), particolarmente basso se confrontato soprattutto con quello delle Finali (67,74%), il più alto della serie.

Non è da escludere che il campione statistico dei turni finali, quantitativamente inferiore per natura man mano che si avanza nel tabellone (la struttura del tabellone tennistico è ad albero), abbia influito nel condurre a risultati non in linea con gli altri turni.

Per quanto riguarda il settore femminile (Figura 33) invece si può evidenziare una leggera tendenza al ribasso del tasso di vittorie per questa modalità all'avanzare nel tabellone. I tre dati inferiori riguardano infatti i turni di Quarti di Finale, Semifinale e Finale, il che potrebbe essere un sintomo di differenze molto più marcate nei turni iniziali e partite più equilibrate nei turni finali, o comunque di un generale livellamento per le giocatrici più forti.

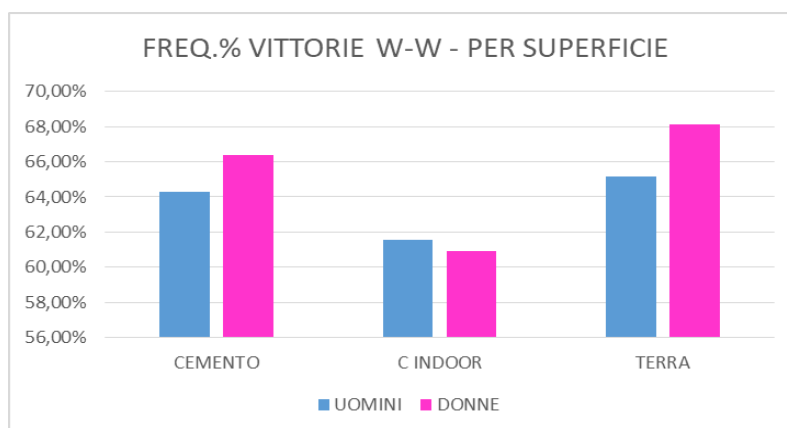


Fig.34 "Frequenze percentuali con cui si verifica l'andamento W-W in base alla superficie di gioco"

Scomponendo invece per superficie (Figura 34), in entrambi i settori si nota come la percentuale di match terminati 2-0 sia inversamente proporzionale alla "velocità" della superficie, citata nel capitolo 3 della relazione finale. Infatti il dato maggiore si verifica sulla terra (media tra uomini e donne 66,66%), a seguire il cemento (media 65,34%) e infine il cemento indoor (media pari a 61,22%).

Sembrerebbe quindi che al crescere della velocità della superficie, cresca anche la probabilità di partite più combattute e viceversa.

In ogni caso resta un punto fermo il fatto che sia costantemente più frequente per le donne ottenere un punteggio e un andamento di questo tipo su ogni superficie.

B. ANALISI MATCH L-W-W

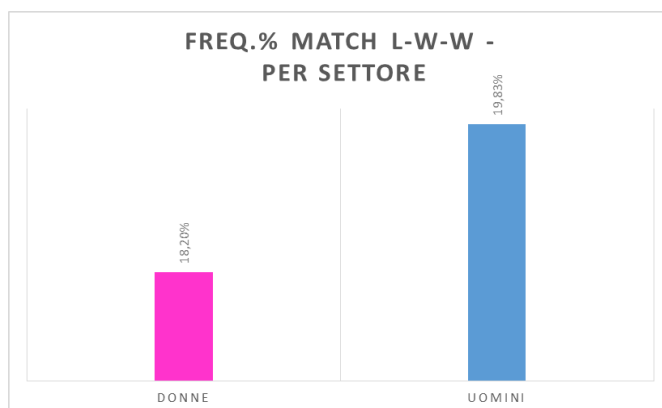


Fig.35: "Frequenze percentuali con cui si verifica l'andamento L-W-W nell'anno solare 2020"

Le vittorie uno avvenute in seguito ad un andamento L-W-W della partita (rimonta), rappresentano come detto la maggior parte delle vittorie avvenute al terzo set.

In generale rappresentano il 19,83% nel settore maschile ed il 18,20% nel settore femminile delle vittorie totali (Figura 35).

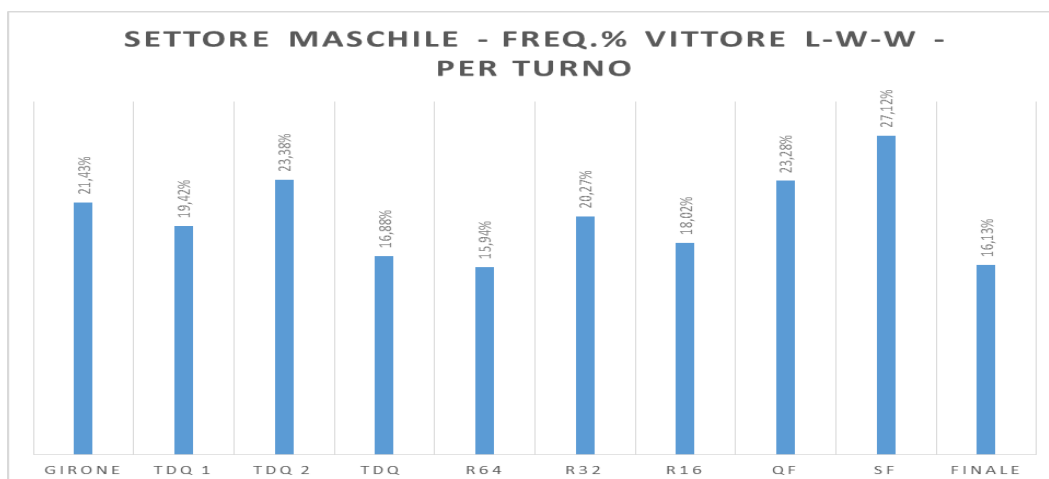


Fig.36: "Frequenze percentuali con cui si verifica l'andamento L-W-W nel settore maschile per turno"

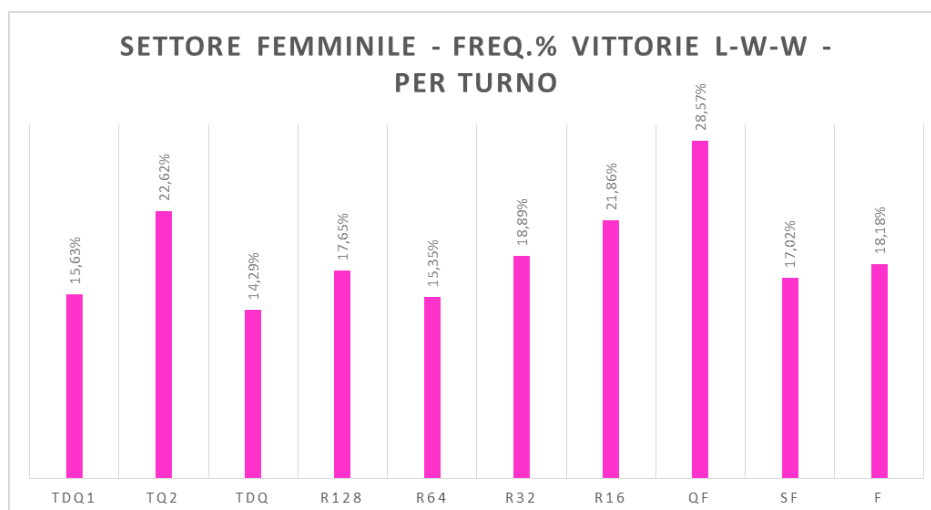


Fig.37: "Frequenze percentuali con cui si verifica l'andamento L-W-W nel settore femminile in base al turno di gioco"

Per quanto riguarda l'analisi per turni spicca ancora una volta il dato leggermente "fuori dal coro" delle semifinali maschili (27,12% contro una media tra i turni del 19,83%), ma si può affermare che non si evidenzino particolari tendenze (Figura 36).

Anche nel settore femminile (Figura 37) è difficile individuare correlazioni tra questo andamento e i vari turni di un tabellone tennistico.

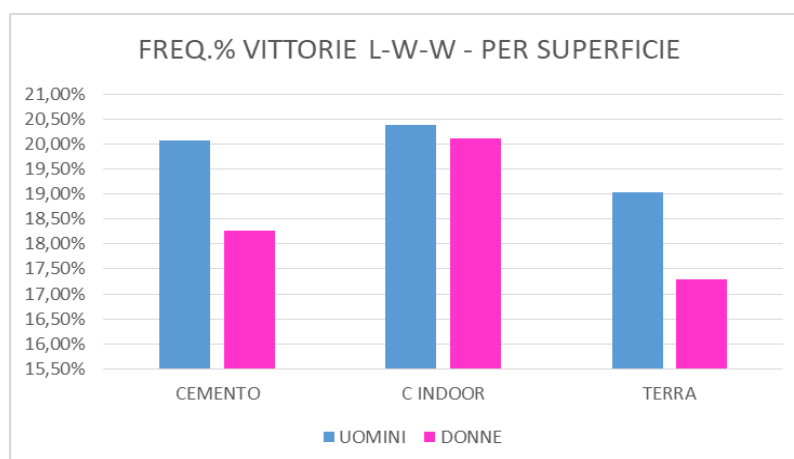


Fig.38 "Frequenze percentuali con cui si verifica l'andamento L-W-W in base alla superficie di gioco"

Per quanto concerne la distinzione per superfici di gioco (Figura 38), vale quanto detto in precedenza, ovvero il fatto che si può stabilire una tendenza relativa alla velocità di superficie.

In questo caso la probabilità di rimonte, o vittorie L-W-W, è direttamente proporzionale alla velocità della superficie di gioco (terra 18,16%, cemento outdoor 19,17% e cemento indoor 20,24% se si fa la media tra settore maschile e settore femminile). In ogni caso gli uomini presentano costantemente per tutte le superfici un dato maggiore di L-W-W.

C. ANALISI MATCH W-L-W

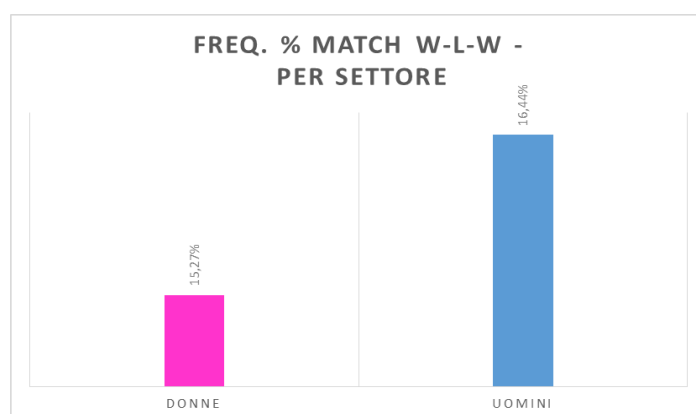


Fig.39: "Frequenze percentuali con cui si verifica l'andamento W-L-W nell'anno solare 2020"

L'andamento W-L-W (Figura 39) rappresenta la tipologia di vittorie meno frequente in un match di tennis, sia in generale, sia per quanto riguarda esclusivamente le partite concluse al terzo set, in particolare il 15,27% per le donne e il 16,44% delle vittorie del settore maschile.

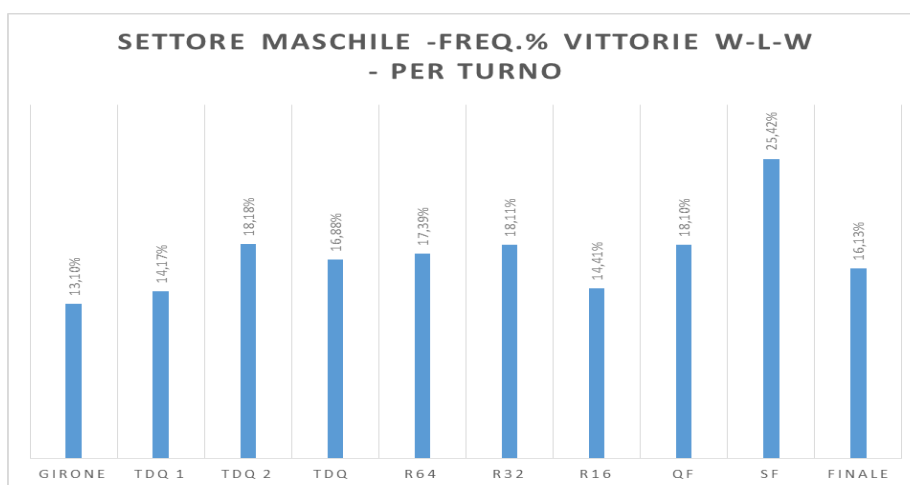


Fig.40: "Frequenze percentuali con cui si verifica l'andamento W-L-W nel settore maschile per turno"

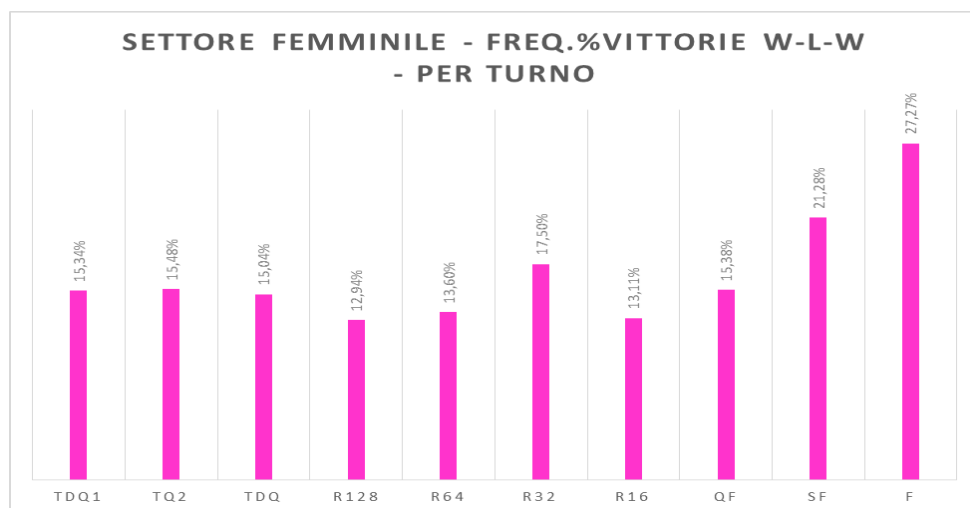


Fig.41: "Frequenze percentuali con cui si verifica l'andamento W-L-W nel settore femminile per turno"

Nel settore maschile (Figura 40) non si evidenziano ancora una volta tendenze particolari rispetto al turno di gioco (resta comunque dato anomalo il picco delle semifinali, 25,42%), mentre nel settore femminile (Figura 41) la percentuale di vittorie W-L-W sembra accentuarsi nei turni decisivi del torneo. E' complicato offrire una spiegazione a questo fatto, ma si potrebbe in tal senso far riferimento al maggior livellamento nelle fasi finali di un torneo femminile rispetto ai primi turni, senza però poter offrire alcuna certezza della correttezza dell'interpretazione.

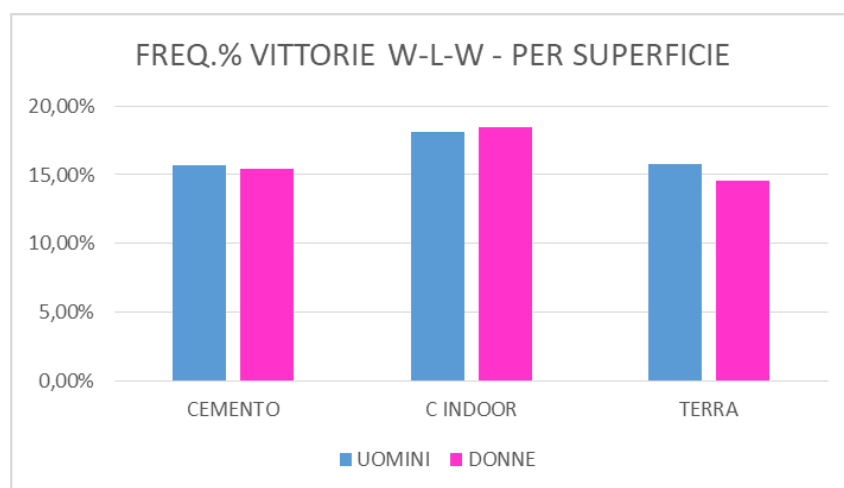


Fig.42: "Frequenze percentuali con cui si verifica l'andamento W-L-W in base alla superficie di gioco"

Ancora una volta (Figura 42) emerge una correlazione con la velocità della superficie, analoga a quella evidenziata per il caso L-W-W (aumento delle vittorie avvenute con questa modalità al crescere della velocità della superficie).

La spiegazione generale potrebbe essere la stessa, trattandosi entrambe della casistica di partite concluse in 3 set.

La correlazione tra superficie e punteggio si può esplicitare nel seguente modo: maggiore è la velocità della superficie, maggiori saranno presumibilmente le probabilità di incontri più combattuti in termine di punteggio.

4.2.C MATCH 3 SU 5 – ESITO DELLA PARTITA

La formula che prevede che il vincitore dell'incontro sia il giocatore che conquista per primo 3 set è detta "3 su 5", e viene utilizzata esclusivamente per il tabellone maschile dei 4 tornei major (i più importanti del circuito) che si disputano ogni anno.

Nel 2020 sono stati giocati solamente 3 dei 4 major in programma poiché il torneo di Wimbledon (unico major su erba) è stato cancellato per la pandemia di COVID-19.

In totale si sono perciò disputate 373 partite "3 su 5" su solamente due superfici differenti: cemento outdoor (248) e terra battuta (125).

Nel dettaglio gli esiti a livello di punteggio finale sono stati i seguenti (Figura 43):

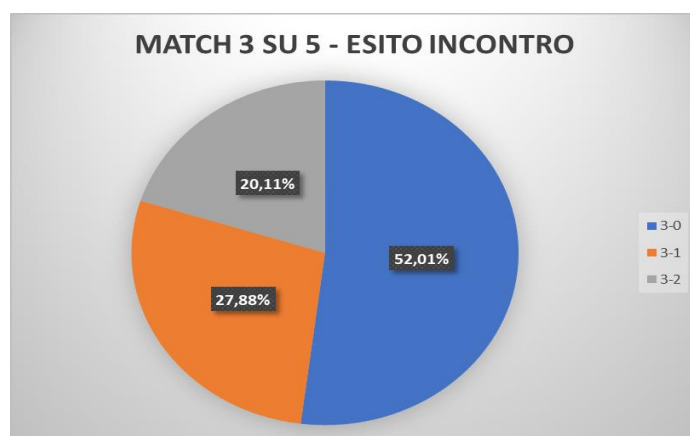


Fig.43: "Frequenze percentuali con cui si verificano gli esiti possibili di un match 3 su 5"

Più della metà delle partite sono state vinte con un punteggio "secco" di 3 set a 0 (52,01%), seguire per frequenza il punteggio di 3 a 1 (27,88%), infine troviamo i match conclusisi al quinto e decisivo set (20,11%)

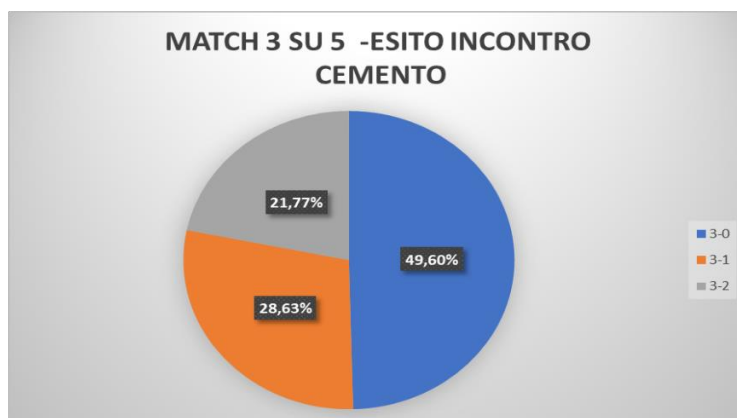


Fig.44: "Frequenze percentuali degli esiti possibili di un match 3 su 5 giocato su cemento outdoor"

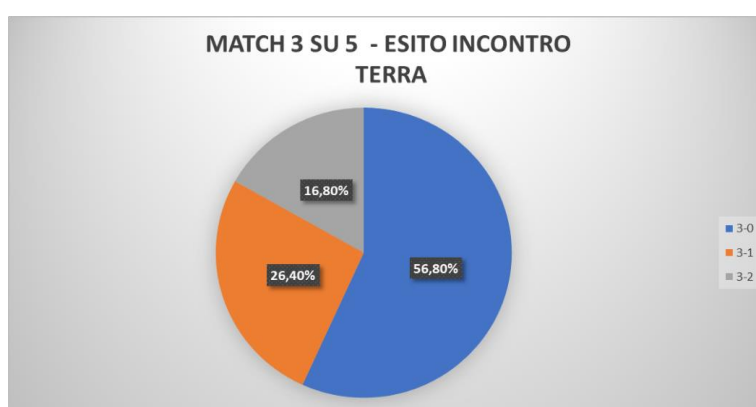


Fig.45: "Frequenze percentuali degli esiti possibili di un match 3 su 5 giocato su terra battuta"

Effettuando una distinzione per superfici, si può notare come venga confermata evidenziata nell'analisi dei match 2 su 3. Infatti la terra battuta (Figura 45) presenta una minore frequenza di match dal punteggio più combattuto rispetto al cemento (Figura 44). Soprattutto il punteggio di 3 a 2, che si verifica solo nel 16,8% sulla terra è molto più raro rispetto a quanto avviene sulla superficie più rapida (21,77%).

Per quanto riguarda l'analisi dell'andamento del match è invece difficile svolgere analisi con risultati facili da interpretare. Ci sono infatti 10 possibili combinazioni in cui è possibile che si svolga una partita al meglio dei 5 set, troppi per presentare elaborazioni qualitativamente soddisfacenti anche in relazione al campione di analisi, non troppo ampio per quanto concerne partite 3 su 5.

4.2.D IL PUNTEGGIO DEI SET

Un set di tennis presenta, analogamente al match, un numero discreto di combinazioni di punteggio possibile. Le unità fondamentali che costituiscono il punteggio dei set sono dette “games” o in italiano “giochi”.

Le sette combinazioni di punteggio possibile di un set (sempre esclusivamente dal punto di vista del giocatore che si aggiudica il parziale) sono dunque le seguenti:

- A. 6-0
- B. 6-1
- C. 6-2
- D. 6-3
- E. 6-4
- F. 7-5
- G. 7-6

Sono state effettuate delle analisi riguardanti la frequenza con cui si ripropongono durante i match presenti nel database i vari punteggi.

Per prima cosa viene riportato di seguito il grafico che raccoglie le frequenze con cui si verificano i vari punteggi per entrambi i settori (Figura 46).

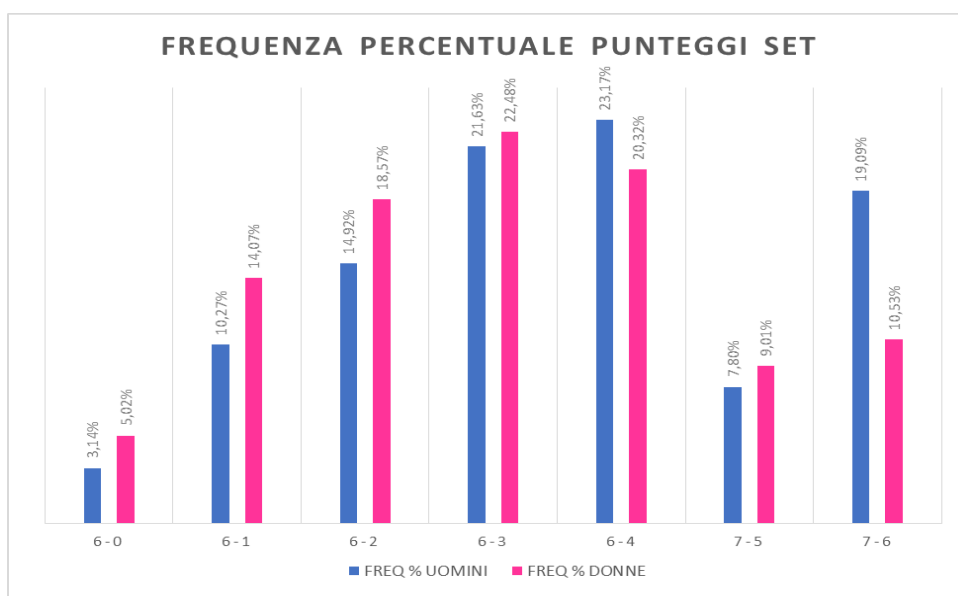


Fig.46: “Frequenze percentuali con cui si verificano i vari punteggi di un set per entrambi i settori”

Nell'illustrazione grafica si può notare come, ordinando i punteggi per punteggio crescente del giocatore sconfitto nel set, e tralasciando momentaneamente il punteggio di 7-5, sembra esserci una distribuzione normale dei valori.

Sia per gli uomini che per le donne vi è una distribuzione con frequenze maggiori in corrispondenza delle differenze più piccole: considerando i diversi possibili risultati, il picco (moda), coincide con il punteggio di 6-3 per le donne (22,48% dei set) mentre si trova in corrispondenza del 6-4 per gli uomini (23,17% sei set totali giocati).

Il fatto che il punteggio di 7-5 sia in netta controtendenza non è invece una sorpresa.

Infatti il regolamento prevede che un set possa essere vinto dal primo giocatore che si aggiudica 6 games, a patto che ce ne siano almeno due di scarto con l'avversario (ciò equivale a dire che ci sono 10 games a disposizione per arrivare a 6). In caso di 5 giochi pari però, si afferma che si procede "ad oltranza" ancora per due games, in modo da avere due situazioni possibili: la conquista del set in caso di 7 giochi a 5, o un tie-break (un game di spareggio), nella casistica di un punteggio di 6 pari.

A questo punto si può quindi iniziare a capire come mai il punteggio di 7-5 presenti una bassa probabilità di verificarsi: il suo ottenimento è infatti subordinato alla condizione che il punteggio dopo i primi dieci games sia di 5 pari, mentre la conquista di un set per 6-x, può essere ottenuta con qualsiasi andamento e senza alcuna condizione di altra probabilità sottostante.

Per effettuare un rilievo statistico corretto sarebbe quindi necessario sommare le probabilità di 7-5 e 7-6, unificandole in una sola voce, che verrà denominata nel grafico sottostante "5-5 - oltranza". Utilizzando questo espediente si ottengono i seguenti risultati (Figura 47).

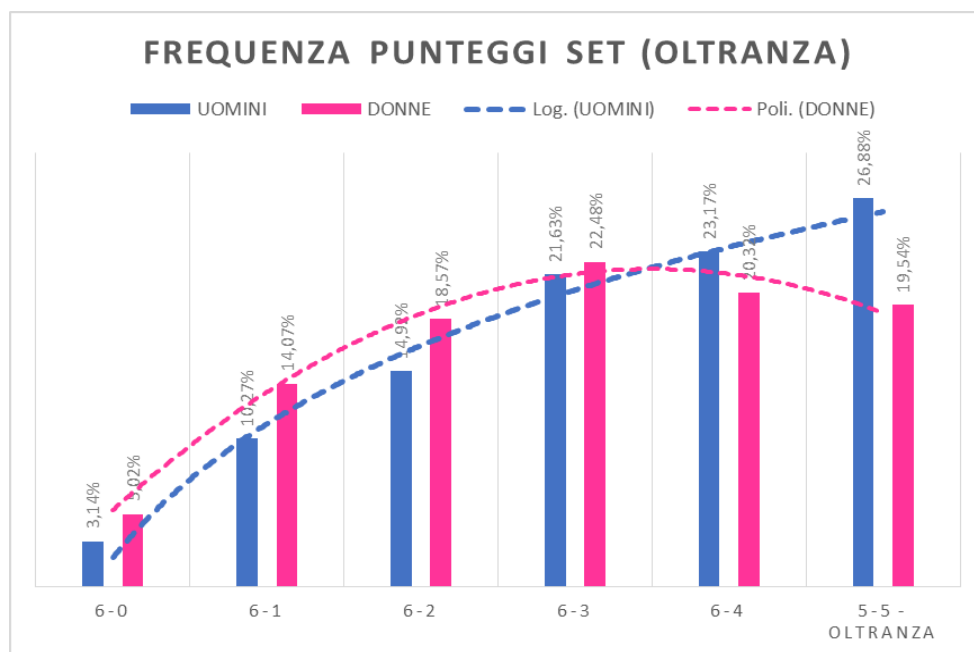


Fig.47: "Frequenze percentuali con cui si verificano i vari punteggi dei set, riviste con la considerazione "ad oltranza" per entrambi i settori"

Come evidenziato dalla Figura 47, esiste un'importante differenza tra uomini e donne: tra gli uomini la frequenza di set conclusi ad oltranza è maggiore anche rispetto a quella del punteggio di 6-4, che in precedenza era il punteggio di moda.

Sostanzialmente la curva di probabilità cresce (logaritmicamente più che linearmente) man mano che aumenta il punteggio del giocatore sconfitto nel set (equivale a dire che aumenta man mano che i parziali sono più combattuti).

Per quanto riguarda il settore femminile invece, si rimane su una distribuzione (simil-) normale con il picco citato in precedenza sul valore di 6-3.

Anche questi risultati sono facilmente spiegabili e perciò si rimanda a quanto affermato nel capitolo 3: le dinamiche di punteggio sono infatti influenzate dall'importanza del servizio e dall'impatto che questo colpo ha sul gioco. Un set di tennis si svolge infatti tramite un'alternanza continua tra turno di servizio e turno di risposta dei due giocatori in campo. Come si è dimostrato nel capitolo 3, gli uomini ottengono molti più punti dal loro servizio e quindi è molto frequente che il punteggio di un set rispecchi l'alternarsi dei turni di servizio e si possa concludere ad oltranza tramite il game di spareggio (tie-break).

Per le donne invece, poiché questo colpo ha meno rilevanza, l'alternanza appena citata di rado viene verificata dai fatti. Hanno cioè preponderanza altri "fattori critici" che determinano scarti di punteggio teoricamente maggiori. Non a caso la curva di probabilità decresce a partire dal punteggio di 6-3, che rimane il valore di picco anche con questa interpretazione statistica. Questo potrebbe spiegare anche il perché, nelle analisi riguardanti il punteggio del match, i dati riguardanti il settore femminile presentassero costantemente scarti più netti: parrebbe insomma essere anche "merito" dell'impatto che il colpo di servizio, e in questo caso i turni di servizio, hanno sul gioco e di conseguenza sul punteggio.

Viene infine presentata un'analisi riguardante eventuali differenze nella frequenza dei punteggi dei set, all'allungarsi della partita. In figura 49 e 50 vengono espresse infatti le frequenze con cui si verificano le sette possibili combinazioni di punteggio nei cinque set possibili.

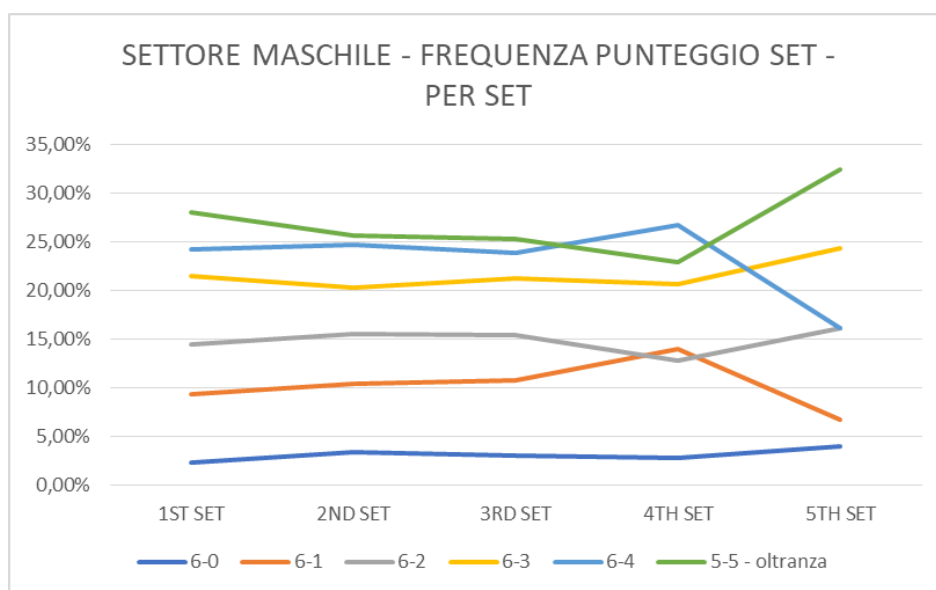


Fig.48: "Frequenze con cui si verificano i vari punteggi dei set al succedersi dei set per il settore maschile"

Per quanto concerne il settore maschile (Figura 48), se nei primi tre set, le frequenze rimangono sostanzialmente costanti, a partire dal quarto parziale sembrano esserci piccoli rimescolamenti: nel quarto set il punteggio di 6-1 sorpassa il 6-2, il 6-4 sorpassa i punteggi

derivati dal 5-5, mentre nel quinto set 6-1 e 6-4 calano vistosamente, mentre riprendono quota i punteggi ad oltranza, il 6-2 e il 6-3.

Per quanto mi riguarda credo che il campione dei dati derivanti dalle partite al terminate quinto set sia troppo limitato (75 partite sulle 1900 maschili nel database) per fare affidamento sui risultati ottenuti. Considerando invece i dati del quarto set, mi sento di tenere valide le informazioni elaborate e dedurre che potrebbe non essere campata per aria l'ipotesi di maggiore probabilità di punteggi "larghi", man mano che si allunga la partita, poiché a quel punto entrano in gioco e diventano influenti fattori critici di varia natura (accennati nel paragrafo 3.1) come le differenze tenuta fisica e psicologica tra i vari giocatori, che si palesano alla distanza e contribuiscono ad allargare la forbice tecnica.

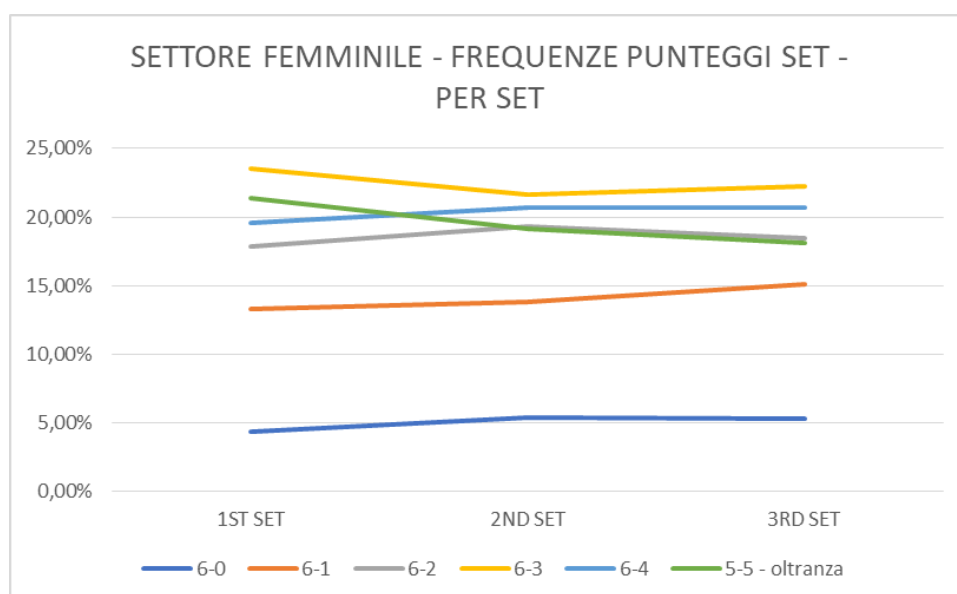


Fig:49: "Frequenze con cui si verificano i vari punteggi dei set al succedersi dei set per il settore femminile"

Nel settore femminile (Figura 49), complice l'inesistenza della formula "3 su 5" e quindi di quarto e quinto set, l'analisi si basa solo sui 3 set giocabili.

Anche qualora esistessero, è più difficile individuare con certezza eventuali tendenze, anche se la progressiva diminuzione di frequenza di set conclusi ad oltranza, fa propendere per validare la considerazione appena fatta per il settore maschile, ovvero la minor probabilità di set combattuti nel punteggio che diminuirebbe teoricamente al succedersi dei set.

4.3 CONSIDERAZIONI FINALI SUL PUNTEGGIO

Ci sono varie considerazioni da affrontare riguardo le analisi statistiche effettuate sui punteggi dei match di Tennis in questo capitolo.

In primo luogo, effettuando una distinzione per turni, ci si rende conto che pare esistere una differenza di valori in campo maggiore nelle fasi finali dei turni maschili rispetto ai tornei femminili, e viceversa per quanto riguarda i turni iniziali. A partire da questa informazione si può dunque ipotizzare che nel settore maschile ci sia una ristretta cerchia di atleti marcatamente superiori alla concorrenza, mentre un livellamento qualitativo a scendere nel ranking. Nel circuito WTA invece sembrerebbe esserci un livellamento ai vertici e poi una costante diminuzione costante del livello delle atlete.

Si deve poi parlare delle differenze emerse tra settore maschile e settore femminile: nel settore femminile i punteggi dei match di tennis sono in genere più netti rispetto a quanto accada negli uomini.

I dati riguardanti le differenze tra le superfici di gioco mostrano che più la superficie è rapida, più le partite sono più equilibrate (nel punteggio), sia per gli uomini sia per le donne.

Infine, le analisi riguardanti i punteggi dei set, mostrano come anche seguendo questo filone, gli equilibri siano più sottili nel settore maschile.

Queste ultime tre considerazioni, se poste in relazione tra loro, parrebbero sottolineare ancora una volta l'importanza del servizio: tutto quanto appena detto è conseguenza infatti dell'impatto che il colpo di servizio ha sul gioco, poiché una maggiore efficacia del colpo implica che il punteggio con più probabilità possa riflettere l'alternanza dei turni di servizio e conseguentemente portare a partite più equilibrate.

Questo "fondamentale" è infatti più determinante nel settore maschile in quanto gli uomini riescono ad effettuarlo con maggiore velocità e potenza, ed è più efficace sulle superfici rapide rispetto a quelle più lente. Una maggiore qualità complessiva del servizio si traduce quindi in punteggi più equilibrati sia per quanto riguarda la prospettiva dei set, che quella dei games.

5.CONCLUSIONI

Nello svolgimento di questo elaborato è stata studiata ed approfondita l'importanza del ruolo che può ricoprire un singolo fattore critico, il servizio, all'interno delle dinamiche di uno sport, il tennis. L'efficacia di questo colpo si palesa attraverso evidenti influenze sul gioco, sul punteggio e su gran parte degli indicatori statistici adottati nella Tennis Data Analytics.

Nell'ambito della Tennis Match Analysis vengono per l'appunto utilizzati alcuni indicatori di performance con lo scopo di esprimere con precisione il rendimento delle specifiche aree critiche di cruciale importanza. In merito agli indicatori utilizzati per il monitoraggio del colpo di servizio si è però rilevata in questa relazione una criticità non trascurabile. Sono infatti impiegati abitualmente indicatori concettualmente poco rappresentativi della realtà, poiché influenzati da diverse variabili e non solamente dal rendimento dell'oggetto di analisi.

L'importanza di utilizzare metriche davvero rilevanti per la misurazione della performance è un argomento di primaria importanza non solo nello sport, ma in tutti quegli ambiti in cui bisogna affrontare ed emergere in un contesto competitivo.

Per questo motivo ci deve dunque essere un continuo studio sottostante che deve mirare alla progettazione e all'utilizzo dei più corretti indicatori di performance per ogni area critica. Oltre ad esprimere la tematica dell'importanza delle tecniche di misurazione della performance, vi è poi un altro obiettivo di questo elaborato, ovvero quello di sottolineare l'importanza dell'utilizzo di un approccio data-driven per effettuare le decisioni strategiche nella gestione di quelle organizzazioni che hanno come scopo il conseguire risultati: economici nel caso delle imprese, o sportivi, nel caso dei team di atleti professionisti.

La scelta di utilizzare lo sport come mezzo per offrire parallelismi affrontando questi argomenti non è dunque casuale.

Lo sport, oltre a rappresentare un settore economico di una certa rilevanza in cui in Italia trovano uno sbocco lavorativo centinaia di migliaia di lavoratori, rappresenta anche una fetta importante dell'industria dell'intrattenimento, mercato che negli ultimi anni, con l'era della tecnologia e dei canali di informazione a portata di tutti, ha avuto un'espansione esponenziale. Gli introiti sempre maggiori derivanti dalla vendita dei diritti televisivi ha come implicazioni il fatto di offrire maggiori possibilità di guadagno agli atleti professionisti (non soltanto a livelli

elitari) e di determinare una maggiore attenzione alla qualità del prodotto televisivo offerto. Tutto questo concorre al venirsi a creare un contesto estremamente competitivo, dove per emergere è richiesto un altissimo livello di attenzione al particolare.

Il mondo della competizione sportiva in senso stretto si può considerare come una metafora rispetto varie tematiche della vita non solo di ogni individuo, ma anche delle imprese. L'ambiente competitivo industriale e quello sportivo richiedono dunque i medesimi sforzi e vi sono delle evidenti somiglianze per quanto riguarda le varie fasi della pianificazione strategica delle attività, che secondo l'approccio moderno deve essere pensata e ragionata in modo data-driven, ovvero facendosi letteralmente guidare dai dati per in modo da effettuare "evidence-based decisions". Ogni decisione deve essere quindi presa razionalmente mediante l'utilizzo di un metodo scientifico che comporta dapprima la raccolta dei dati e delle informazioni potenzialmente rilevanti, effettuata con le più moderne tecnologie di rilevazione e secondo i dettami di figure professionali specializzate. Questi professionisti provvedono poi all'elaborazione dei dati raccolti con l'obiettivo di stabilire correlazioni e connessioni tra le diverse variabili in modo di cercare di spiegare, prevedere e comprendere in modo scientifico le realtà oggetto delle ricerche.

A partire dai risultati di questi processi, vengono quindi effettuate decisioni strategiche che devono riflettere le proprie capacità, caratteristiche e attitudini (si è parlato della statura degli atleti, come si potrebbe parlare per esempio del grado di istruzione dei dipendenti), poichè le peculiarità di un'impresa o di un atleta definiscono "priorità competitive" che a loro volta implicano ben determinate e differenti strategie competitive.

La capacità di monitorare nel tempo le prestazioni degli attori coinvolti, utilizzando indicatori di performance che siano in grado di esprimere il reale rendimento di un determinato fattore critico, riveste una grande rilevanza nel successo di tutto il processo. Gli indicatori di performance, per essere attendibili, debbano essere concepiti in modo tale da riflettere perfettamente il rendimento del fattore critico cui si riferiscono così da garantire oggettività e precisione nel monitoraggio. Tutto questo presuppone che ci debba essere una profonda conoscenza sia della materia affrontata sia delle metodologie scientifiche di analisi statistica.

Si è infatti parlato di Data Analysts e Data Scientists, figure professionali dalle competenze interdisciplinari coinvolte nei processi di Data Analytics, che ricoprono sempre più frequentemente una presenza fissa nei teams degli atleti d'élite.

L'integrazione di queste figure negli staff dei professionisti può essere considerato come lo specchio di quanto accade a livello di gestione delle imprese, dove gli approcci moderni vogliono che vengano intrapresi processi di concurrent engineering (vi è una forte similitudine con lo sport, dove le collaborazioni tra coach e Data Analysts per la progettazione dei piani di allenamento più adatti) e legami con i fornitori sempre più orientati alle collaborazioni di lungo termine (cooperazioni win-win), talvolta integrati addirittura fisicamente nelle fabbriche per gestire meglio il flusso delle informazioni e delle decisioni riguardanti il processo produttivo.

Si è quindi compreso l'importanza del ruolo svolto dalla Data Analytics nei processi di pianificazione strategica in tutti i settori, nello sport come nelle imprese, e l'impatto che giocano le moderne tecnologie nel poter consentire di disporre dati accurati e informazioni rilevanti.

6.RIFERIMENTI

RIFERIMENTI BIBLIOGRAFICI

Comitato Leonardo, P. C. (2018). *Sport e Imprese: play together* .

Gil Fried, C. M. (2017). *Sport Analytics: A data-driven approach to sport business and management*. Routledge.

Hughes, M., Lipoma, M., & Sibilio, M. (2015). *PERFORMANCE ANALYSIS: ELEMENTI DI BASE E ASPETTI APPLICATIVI IN CAMPO EDUCATIVO E INTEGRATIVO*. FRANCO ANGELI.

Krajewski, Ritzman, Mlhotra, Grando, & Secchi. (2015). *Supply Chain Managment*. Pearson Editore.

Mecheri, S., Rioult, F., Mantel, B., Kauffman, F., & Benguigui, N. (2016). The serve impact in tennis: First large-scale study of big Hawk-Eye data: The Serve Impact in Tennis.

RIFERIMENTI SITOGRAFICI

1,2 Il libro bianco dello sport Italiano, CONI, 2020, coni.it

3 “Industria dell’entertainment: solo lo sport fattura 130 miliardi” (M.Frisone, www.ilsole24ore.com, 2019)

4 “La pratica sportiva in Italia”, ISTAT, 2015, istat.it

5 Tavolo Nazionale per la Governance nello Sport. Piano Nazionale per la Promozione dell’Attività Sportiva. 1° edizione. Roma, 2012, archivi.istruzione.it

6 Capgemini Research Institute, www.capgemini.com

7,8 LA CRESCENTE IMPORTANZA DELLA FOOTBALL ANALYTICS, Soccerment Research, soccerment.com

9 “Migliorare le prestazioni sportive, fotogramma per fotogramma”, Technogym.it

- 10 The Role of Data Science in Sports, mastersintdatascience.com
- 11 “Data Scientist vs Data Engineer vs Data Analyst – What really differentiates them?”, data-flair.traninig
- 12 Definizione CONI, scuoladelloSPORT.coni.it
- 13 “Cos’è la performance analysis?”, bskilled.it
- 14 “Key Performance Indicators” valentinomea.it
- 15 “La match analysis”, scienzaemovimento.it
- 16 ITF Global Tennis Report, 2019, itftennis.com
- 17 “Wta-Sap, la tecnologia al servizio del tennis femminile”, 2016, F.Pallone, repubblica.it
- 18 “ATP and Infosys Extend Digital Innovation Partnership”, Infosys.com
- 19 “Tennis e dati: ok, ma quali sono quelli disponibili? Dai più grezzi ai dati Hawkeye”, F. Bertelli, 2020, ubitennis.com
- 20 “Hawk-Eye”, wikipedia.it
- 21 “Hawk-Eye Innovations Ltd has been acquired by Sony Professional Solutions”, oaklinks.com
- 22 “Tennis e dati: come si raccolgono i dati sul campo (e quanti soldi servono)”, Ubitennis.com
- 23 Estratto di intervista a Shane Liyange, Tennis Data Analyst, ubitennis.com
- 24 “Che cos’è un data lake?”, neodatagroup.com
- 25 “Match Charting Project”, settesei.it
- 26 “Le superfici non sono diventate tutte uguali, lo dicono i dati”, V. Vignoli, Ubitennis.com
- 27 ATP Leaderboards, Serve Stats, ATP.com
- 28 “Measuring the Impact of the Serve in Men’s Tennis”, 2017, J. Sackmann, Tennisabstract.com

29 P. O'Donoghue and A. Ballantyne, The impact of speed of service in Grand Slam singles tennis, In *Science and Racket Sports III*, A. Lees, J. F. Kahn, and I. Maynard, eds. London and New York, Routledge, 2004, 179–184

30 How Much Does Height Matter in Men's Tennis?, 2017, J.Sackmann, Tennisabstract.com