

II CORSO FOOTBALL DATA INTELLIGENCE

**VALUTAZIONE E ANALISI MODELLO
EXPECTED ASSIST SU SVILUPPI DI
CALCIO D'ANGOLO**



soccerment

RELATORI

Michele Crestani

Aldo Comi

Antonio Gagliardi

Roberto Angioni

Francesco Bordin

CORSISTA

Salvatore Nizza

26 Settembre - 28 Ottobre 2022

INDICE

PREMESSA	3
1. INTRODUZIONE	3
2. DATI STATISTICI	6
3. DATASET UTILIZZATI	8
4. SVOLGIMENTO	8
4.1 FASE 1 - PREPARAZIONE DATI	8
4.2 FASE 2 - CLASSIFICAZIONE	9
4.2.1 GRUPPO A	9
4.2.2 GRUPPO B	9
4.2.3 GRUPPO C	10
4.2.4 GRUPPO D	10
4.3 FASE 3 - IMPLEMENTAZIONE MODELLO	12
5. ANALISI DEI DATI	13
5.1 VALUTAZIONE MODELLO	13
5.2 RISULTATI	19
6. CONFRONTI	20
7. CONSIDERAZIONI FINALI	24

PREMESSA

Lo scopo di questo studio è quello di valutare e analizzare un modello expected assist ottenuti da un modello di regressione logistica prendendo in esame la stagione 2021-2022 della Serie A per comprendere in quale situazione da calcio d'angolo si ha un beneficio offensivo.

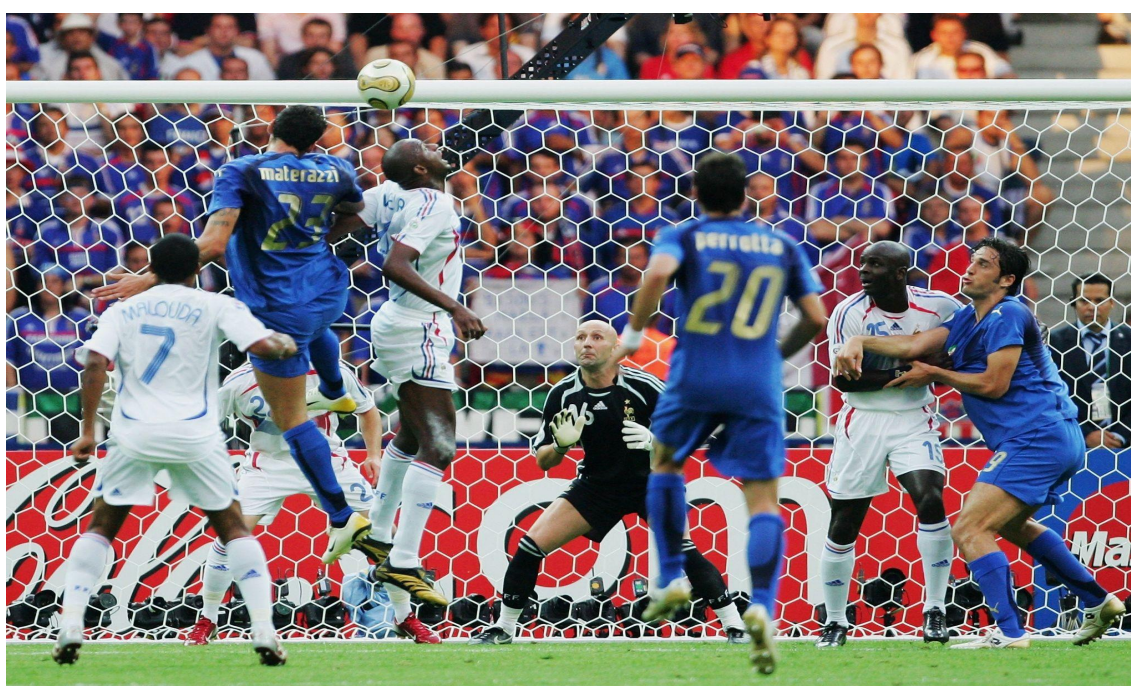
1. INTRODUZIONE

Come recita Il Regolamento del Giuoco del Calcio corredato delle Decisioni Ufficiali FIGC e della Guida Pratica AIA, *“un calcio d'angolo viene assegnato quando il pallone, toccato per ultimo da un calciatore della squadra difendente, ha interamente superato la linea di porta, sia a terra sia in aria, senza che una rete sia stata segnata”*.

Il calcio d'angolo è una situazione chiave di gioco all'interno di una partita di calcio che può permettere di realizzare una rete da palla inattiva e influire in modo importante sul risultato dell'incontro.

Solo per citare alcuni esempi:

- Marco Materazzi, Finale della Coppa del Mondo 2006, Italia – Francia, gol del pareggio su assist di Andrea Pirlo per portare la squadra ai calci di rigore e alla conquista del quarto titolo mondiale



- Sergio Ramos, Finale della Champions League 2013-2014, Real Madrid – Atletico Madrid, gol del pareggio su assist di Luka Modric per portare la squadra ai tempi supplementari e alla conquista della *‘decima’*



- Kostas Manolas, Ritorno Quarti di finale della Champions League 2017-2018, Roma – Barcellona, gol del 3 a 0 su assist di Cengiz Under e passaggio del turno



- Matias Vecino, 38° giornata di ritorno della Serie A 2017-2018, Lazio – Inter, gol vittoria su assist di Marcelo Brozovic e ritorno dell’Inter in Champions League dopo 6 anni

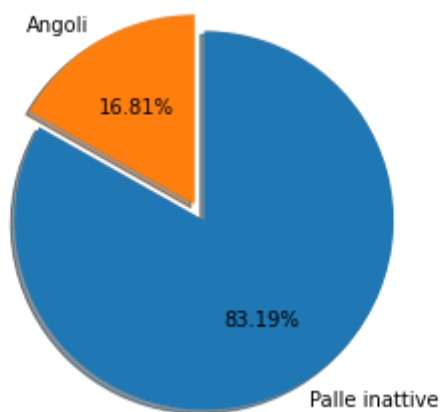


- Divock Origi, Ritorno Semifinali della Champions League 2018-2019, Liverpool – Barcellona, gol del 4 – 0 su assist di Trent Alexander-Arnold e passaggio del turno

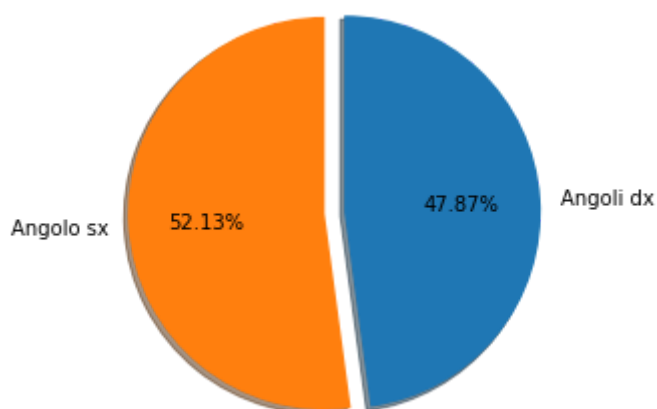


2. DATI STATISTICI

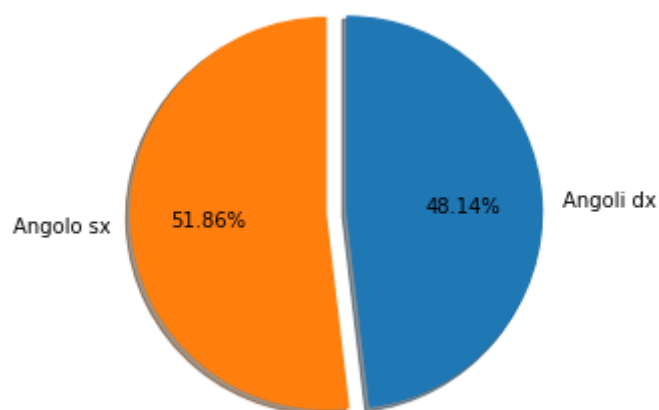
Durante la stagione 2021-22 della Serie A gli eventi da situazione di calcio d'angolo hanno rappresentato il 16,81 % (3.654) delle palle inattive a favore (18.079) che si sono verificati nell'arco del campionato.



Il 52,13 % (1.905) degli angoli è avvenuto dalla parte sinistra rispetto alla direzione di attacco mentre il 47,87 % (1.749) è avvenuto dalla parte destra rispetto alla direzione di attacco.



Considerando invece solamente i passaggi finiti all'interno dell'area di rigore avversaria si ottiene il 51,86% (1.547) di eventi da calcio d'angolo di sinistra e il 48,14% (1.436) di eventi da calcio d'angolo di destra.



Compresa la frequenza delle azioni da calcio d'angolo si è proseguito con la stesura dell'elaborato.

3. DATASET UTILIZZATI

Per sviluppare la tematica sono stati utilizzati i dati derivanti dai seguenti dataset:

- **SICS** – dati posizionali da calcio d’angolo (*SICS-Serie A 2021-22_eventi*)
- **Transfermarkt** – dati su piede preferito (*database_foot_serieA_2021-22*)
- database numeri identificativi giocatori (*player_id_serieA_2021-22*)

4. SVOLGIMENTO

Lo studio è stato suddiviso nelle seguenti fasi:

- FASE 1: Preparazione dati
- FASE 2: Classificazione dati
- FASE 3: Implementazione modello

4.1 FASE 1 – PREPARAZIONE DATI

Prima della classificazione dei calci d’angolo in base alla zona di campo da cui vengono effettuati (lato destro o sinistro) e in base al piede del battitore (destro o sinistro) che sarà presa in esame successivamente nella Fase 2, si è provveduto alla preparazione dei dati, in particolare all’unione del dataset riguardante i dati posizionali di SICS *SICS-Serie A 2021-22_eventi* e del dataset riguardante le informazioni sul piede preferito di ogni giocatore *database_foot_serieA_2021-22* ottenuto da un’operazione di scraping dei dati derivanti da Transfermarkt.

Dato che i due dataset non hanno nessuna chiave in comune tramite la quale effettuare l’unione, si è utilizzato un terzo dataset riguardante i diversi numeri identificativi di ogni giocatore (SICS, Soccerment, Skillcorner) *player_id_serieA_2021-22*.

Si nota come i dataset dei numeri identificativi e di Transfermarkt abbiano in comune il nome dei giocatori da utilizzare come chiave di unione.

Eseguita l'unione di questi ultimi due dataset si è passati a un'ulteriore unione con il dataset **'SICS-Serie A 2021-22_eventi'** utilizzando come chiave di unione il numero identificativo di SICS.

Ultimata questa operazione e non considerando i giocatori ambidestri, per i quali non è stato possibile associare il piede con cui sono stati calciati i calci d'angolo, si è ottenuto il dataset finale con il quale effettuare l'analisi dei dati.

4.2 FASE 2 – CLASSIFICAZIONE

Ottenuto il dataset utile allo studio descritto si è passato alla classificazione dei calci d'angolo in base alla zona di campo da cui vengono effettuati e in base al piede del battitore.

Sono stati classificati quattro gruppi:

- **GRUPPO A:** angolo destro, piede destro battitore (cross con effetto a uscire)
- **GRUPPO B:** angolo destro, piede sinistro battitore (cross con effetto a rientrare)
- **GRUPPO C:** angolo sinistro, piede sinistro battitore (cross con effetto a uscire)
- **GRUPPO D:** angolo sinistro, piede destro battitore (cross con effetto a rientrare)

4.2.1 GRUPPO A (cross con effetto a uscire da destra)

I dati sono stati filtrati secondo:

1. le coordinate di partenza ($P0 \text{ norm } x > 103$, $P0 \text{ norm } y < 2$) del passaggio da calcio d'angolo destro
2. le coordinate di arrivo del passaggio dentro l'area di rigore ($88,5 \leq P1 \text{ norm } x \leq 105$, $14 \leq P1 \text{ norm } y \leq 54$) da calcio d'angolo destro
3. i battitori di piede destro

4.2.2 GRUPPO B (cross con effetto a rientrare da destra)

I dati sono stati filtrati secondo:

1. le coordinate di partenza ($P0 \text{ norm } x > 103$, $P0 \text{ norm } y < 2$) del passaggio da calcio d'angolo destro
2. le coordinate di arrivo del passaggio dentro l'area di rigore ($88,5 \leq P1 \text{ norm } x \leq 105$, $14 \leq P1 \text{ norm } y \leq 54$) da calcio d'angolo destro
3. i battitori di piede sinistro

4.2.3 GRUPPO C (cross con effetto a uscire da sinistra)

I dati sono stati filtrati secondo:

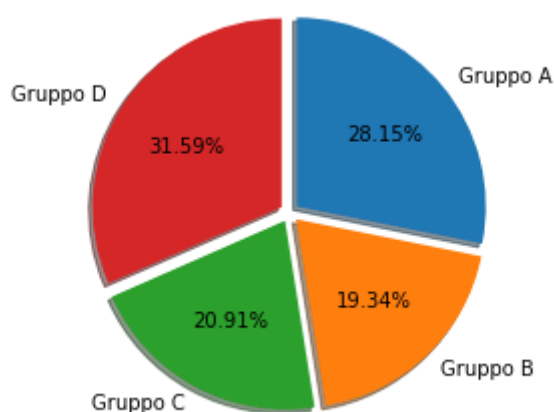
1. le coordinate di partenza ($P0 \text{ norm } x > 103$, $P0 \text{ norm } y > 66$) del passaggio da calcio d'angolo sinistro
2. le coordinate di arrivo del passaggio dentro l'area di rigore ($88,5 \leq P1 \text{ norm } x \leq 105$, $14 \leq P1 \text{ norm } y \leq 54$) da calcio d'angolo sinistro
3. i battitori di piede destro

4.2.4 GRUPPO D (cross con effetto a rientrare da sinistra)

I dati sono stati filtrati secondo:

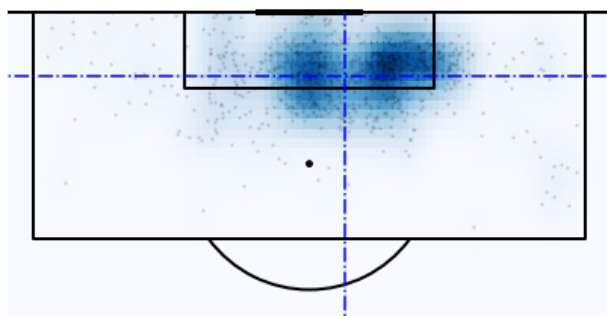
1. le coordinate di partenza ($P0 \text{ norm } x > 103$, $P0 \text{ norm } y > 66$) del passaggio da calcio d'angolo sinistro
2. le coordinate di arrivo del passaggio dentro l'area di rigore ($88,5 \leq P1 \text{ norm } x \leq 105$, $14 \leq P1 \text{ norm } y \leq 54$) da calcio d'angolo sinistro
3. i battitori di piede sinistro

Come riportato dal grafico sottostante il Gruppo A rappresenta il 28,15% (770) degli eventi da calcio d'angolo, il Gruppo B il 19,34% (529), il Gruppo C il 20,91% (572) e il Gruppo D il 31,59% (864).

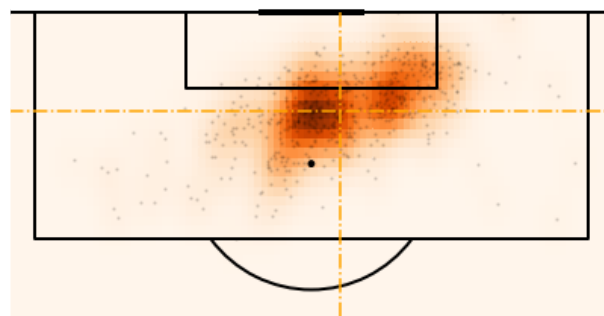


Successivamente sono state riportate su uno scatterplot le coordinate di arrivo dei passaggi effettuati per ognuno dei 4 gruppi con lo scopo di visualizzare attraverso un heatmap su quale porzione dell'area di rigore si ha una maggiore densità di palloni.

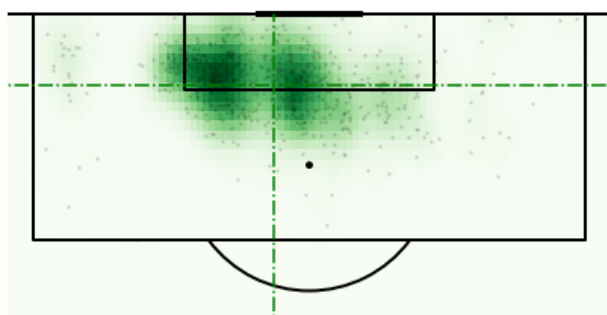
GRUPPO A
angolo dx, piede dx
(effetto a uscire)



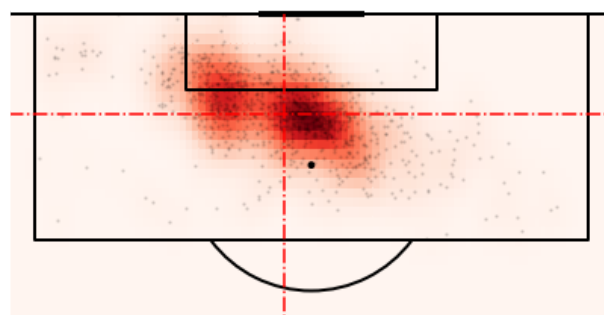
'GRUPPO B
angolo dx, piede sx
(effetto a rientrare)'



GRUPPO C
angolo sx, piede sx
(effetto a uscire)



GRUPPO D
angolo sx, piede dx
(effetto a rientrare)



Per quanto riguarda il gruppo A e il gruppo C (cross con effetto a uscire) la maggior parte dei passaggi cade in corrispondenza del vertice dell'area di porta più vicino al punto di battuta del calcio d'angolo e in corrispondenza della linea longitudinale dell'area di porta in posizione centrale. Per quanto riguarda il gruppo B e il gruppo D (cross con effetto a rientrare) la maggior parte dei passaggi cade in corrispondenza della zona centrale tra la linea longitudinale dell'area di porta e il dischetto di rigore

Le linee tratteggiate orizzontali e verticali rappresentano rispettivamente l'altezza media delle coordinate di tutti i passaggi all'interno dell'area di rigore rispetto al lato lungo del campo e l'altezza media delle coordinate di tutti i passaggi all'interno dell'area di rigore rispetto al lato corto del campo.

Da un punto di vista tattico si può pensare che tali soluzioni vengono adottate per evitare che il portiere avversario intercetti la palla prima che questa raggiunga la zona di potenziale battuta a rete.

4.3 FASE 3 - IMPLEMENTAZIONE MODELLO

Dopo aver descritto le diverse soluzioni da calcio d'angolo attraverso la classificazione in quattro gruppi si è passati all'implementazione di un modello di regressione logistica semplificato di expected assist x_A con l'obiettivo di stimare quale delle diverse soluzioni adottabili porti più beneficio in termini di assist.

Per la creazione del modello si è fatta l'ipotesi che la metrica sia di tipo *shot-centric*, ossia che l'assist dipenda esclusivamente dal fatto che il colpitore abbia ricevuto effettivamente il passaggio. Alla luce della considerazione appena descritta, si può assumere che il modello di expected assist sia paragonabile a quello di expected goal. Quindi, in fase di creazione del modello di expected assist si adotteranno delle ipotesi che valgono per un modello di expected goal (non verranno considerate all'interno del modello le coordinate di partenza dei passaggi dato che tutti gli eventi presi in considerazione derivano da calcio d'angolo, punto in comune per tutti i passaggi).

Per l'implementazione del modello sono state individuate dapprima una variabile target e delle variabili predittive.

Come variabile target è stata creata una nuova colonna per gli assist alla quale, per ogni situazione, è stato associato il valore 1 in caso di passaggio che ha raggiunto il colpitore oppure il valore 0 in caso contrario (se non si è registrato un tiro).

Invece, come variabili predittive sono stati considerati la distanza dal centro della porta, ovvero l'ipotenusa del triangolo formato dalle coordinate di un punto, e l'angolo di porta, ovvero l'angolo sottinteso dall'ampiezza della porta per ogni punto.

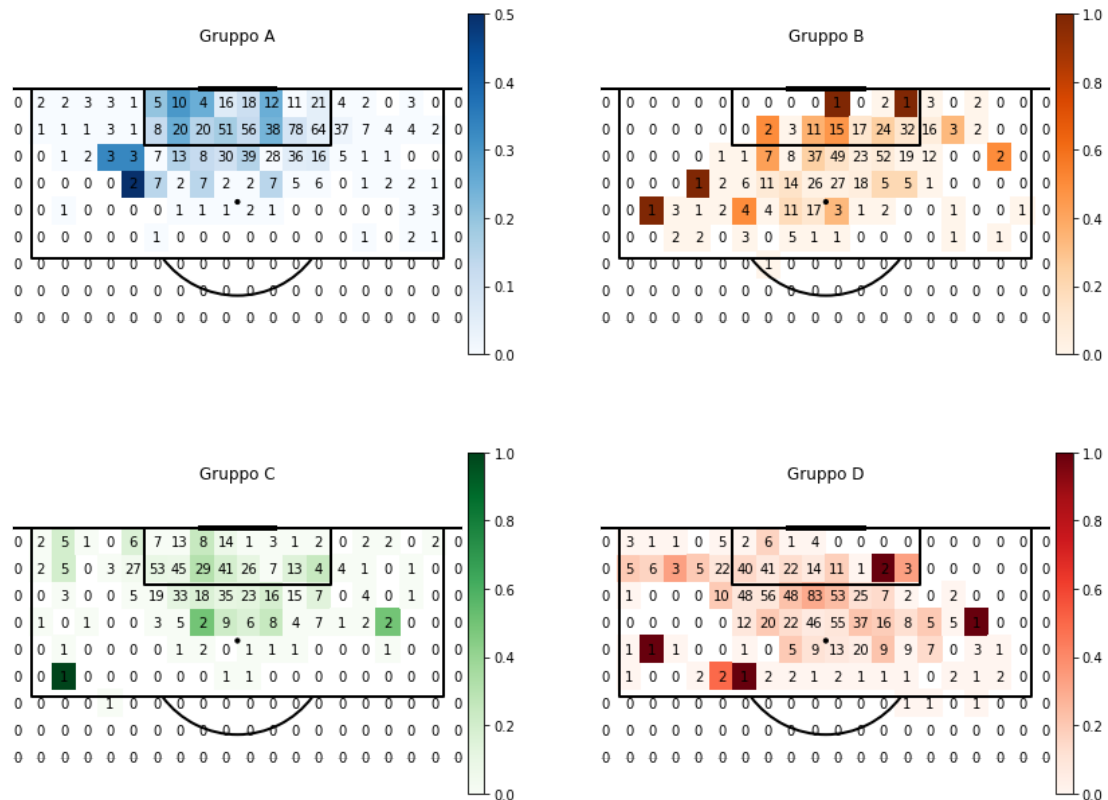
Le due variabili sono legate alla probabilità che un tiro diventi un gol e quindi contemporaneamente alla probabilità che un passaggio diventi un assist. Minore sarà la distanza dalla porta (a meno della porzione di area presso la linea di porta presidiato dal portiere) e maggiore l'angolo di porta, maggiore sarà la probabilità che si verifichi un assist, e viceversa.

Definite la variabile target e le variabili predittive, è stato implementato il modello statistico per gli expected assist attraverso una regressione logistica e infine è stato allenato con i dati a disposizione.

5. ANALISI DEI DATI

5.1 VALUTAZIONE MODELLO

Osservando il conversion rate della variabile target si evince che esistono dei valori estremi dati dalla poca numerosità degli assist in quelle zone dell'area di rigore.



Quindi risolvere questa situazione è stato adottato un modello statistico di regressione logistica per trovare una correlazione tra la variabile target (xA) e le variabili predittive (distanza e angolo di porta).

Come detto in precedenza ci si dovrebbe aspettare quanto segue:

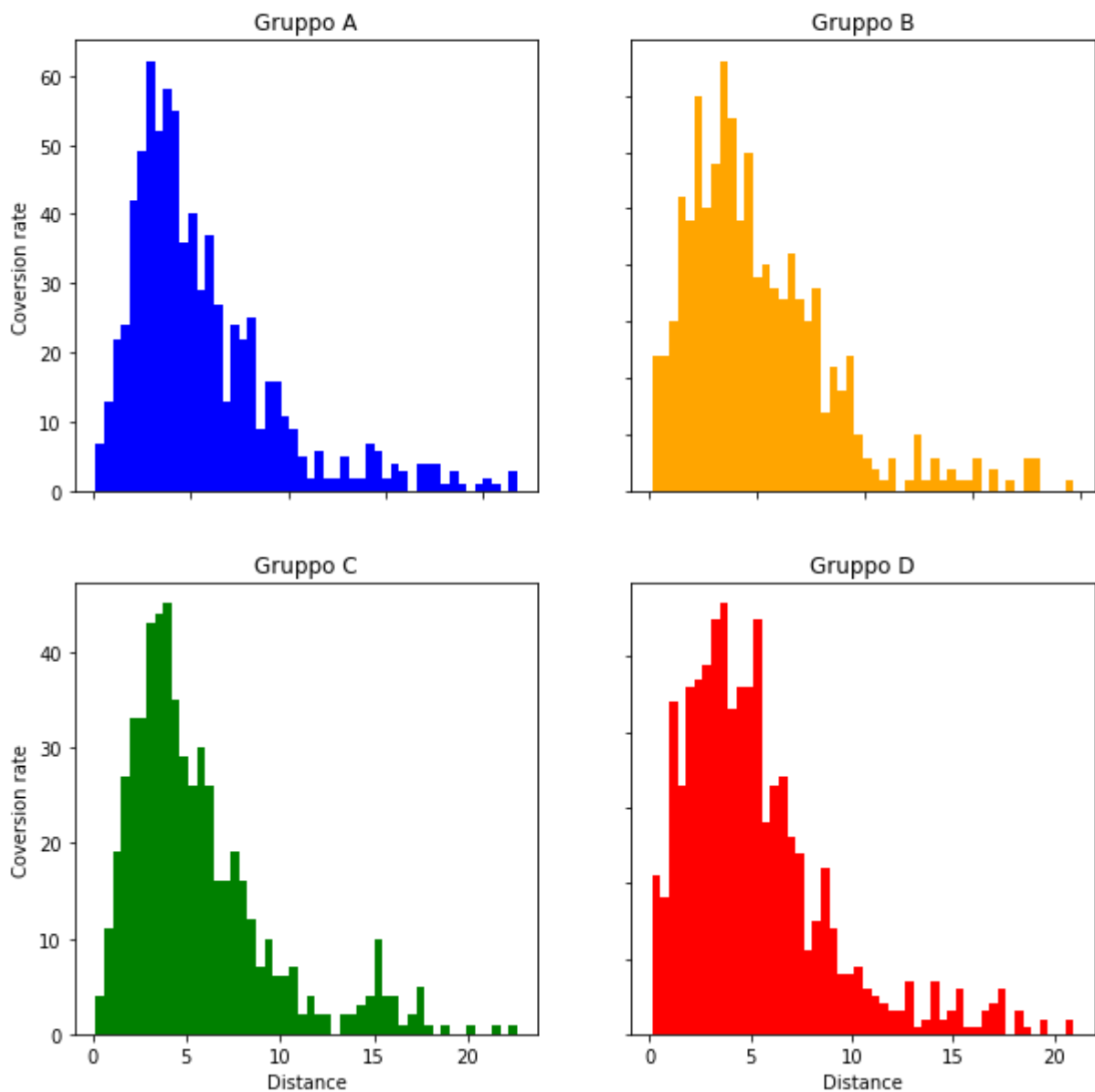
1. minore è la distanza dalla porta e maggiore è la probabilità che si verifichi un assist (a meno della porzione di area a ridosso della linea di porta presidiato dal portiere)
2. maggiore è l'angolo di porta, maggiore è la probabilità che si verifichi un assist

e viceversa

3. maggiore è la distanza dalla porta e minore è la probabilità che si verifichi un assist
4. minore è l'angolo di porta, minore è la probabilità che si verifichi un assist

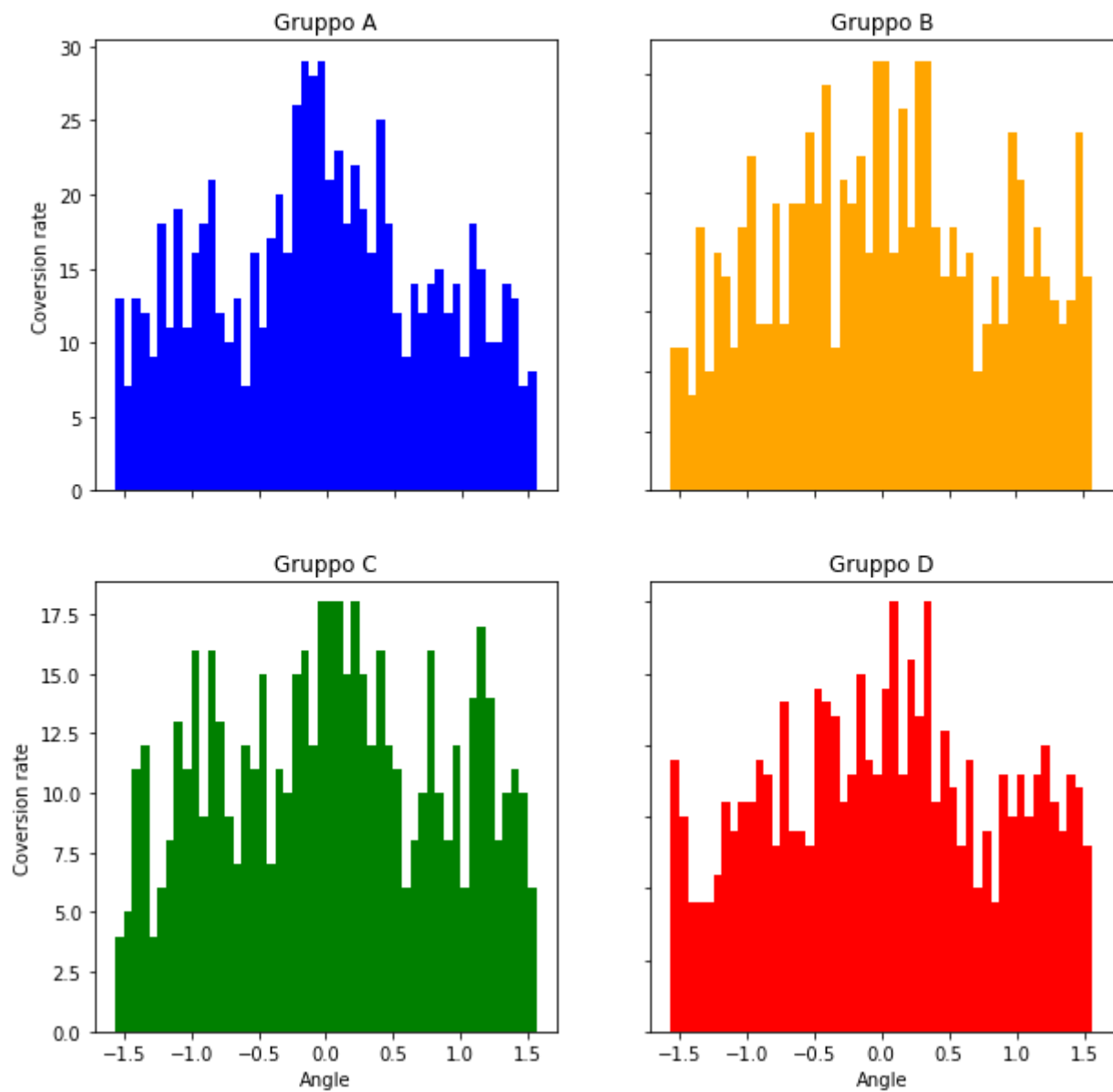
Mettendo in relazione il conversion rate e la distanza dalla porta si osserva che nei quattro casi presi in esame il conversion rate diminuisce all'aumentare della distanza dalla porta (a meno della porzione di area a ridosso della linea di porta) così come ipotizzato in fase di creazione del modello.

Conversion rate depending on distance

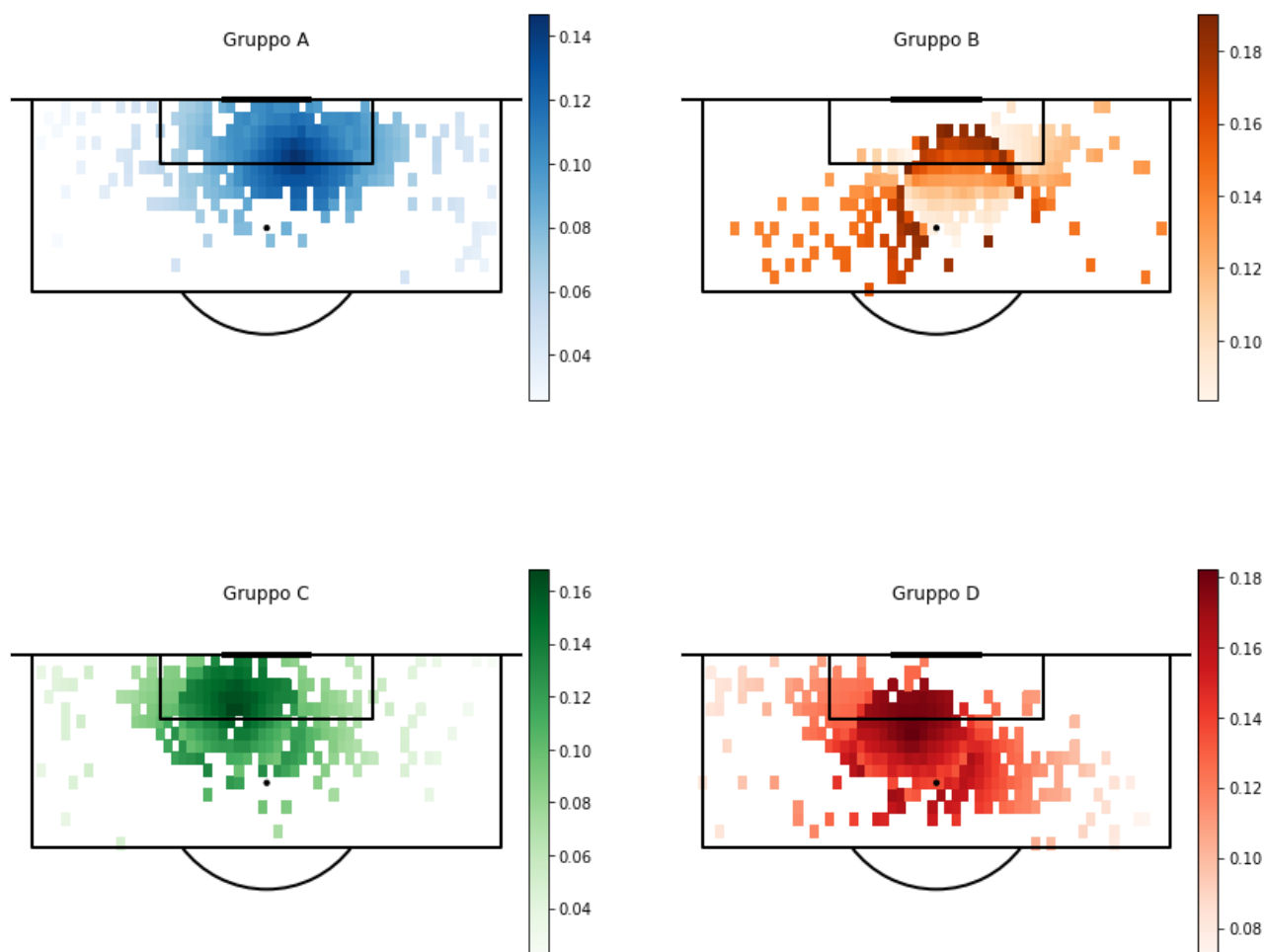


Mettendo, invece, in relazione il conversion rate e l'angolo di porta si osserva che nei quattro casi presi in esame il conversion rate diminuisce al diminuire dell'angolo di porta, anche se non così chiaramente rispetto a quanto ipotizzato durante la creazione del modello.

Conversion rate depending on angle



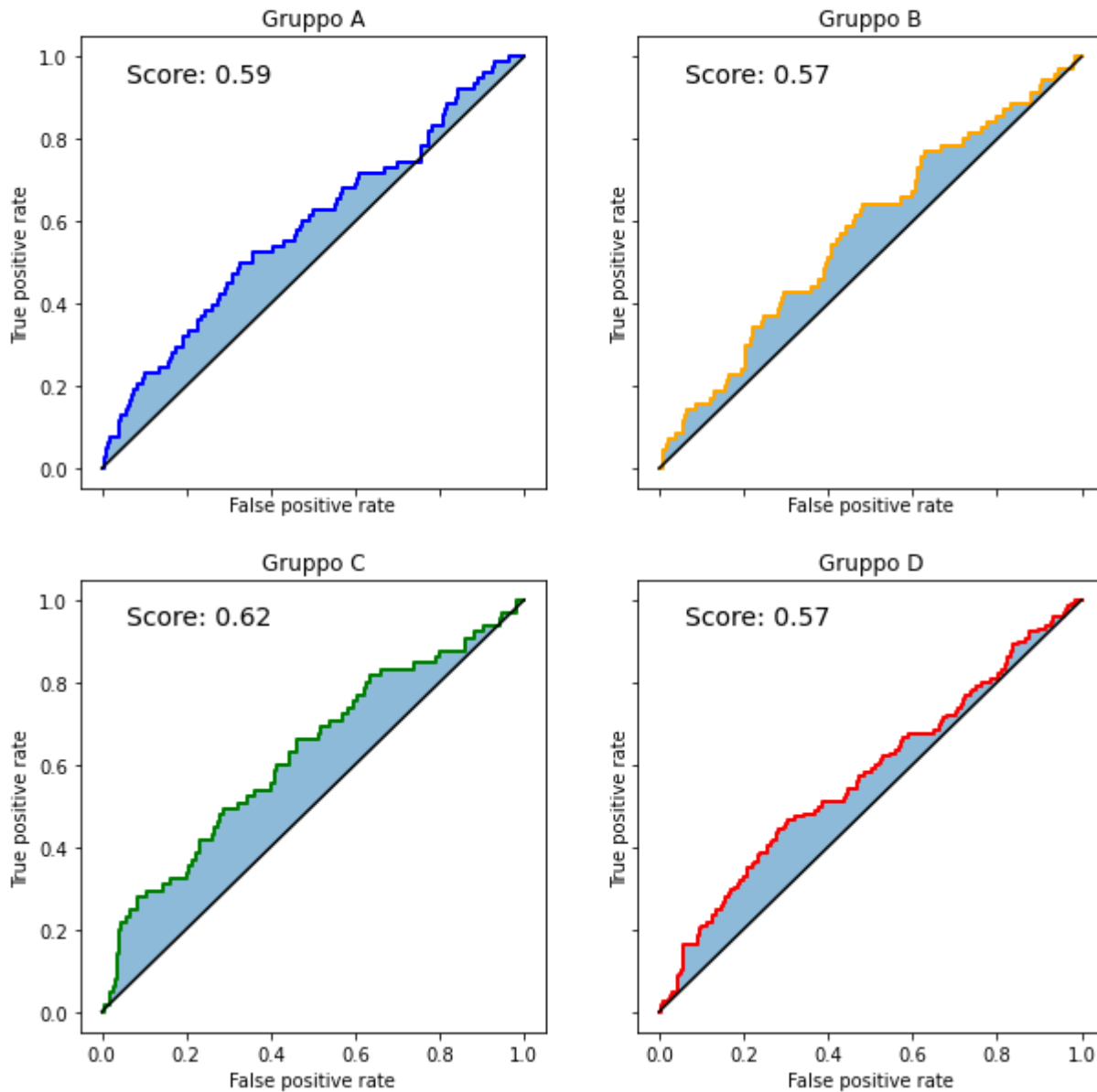
Successivamente, il modello statistico è stato allenato e i valori di expected assist sono stati riportati nei grafici sottostanti:



Sembrerebbe che in tre casi su quattro il modello abbia restituito dei valori in linea con le ipotesi adottate.

Per valutare la bontà del modello da un punto di vista quantitativo si è utilizzato il ROC AUC score che mette in relazione la frazione dei veri positivi e la frazione di falsi positivi.

ROC AUC score

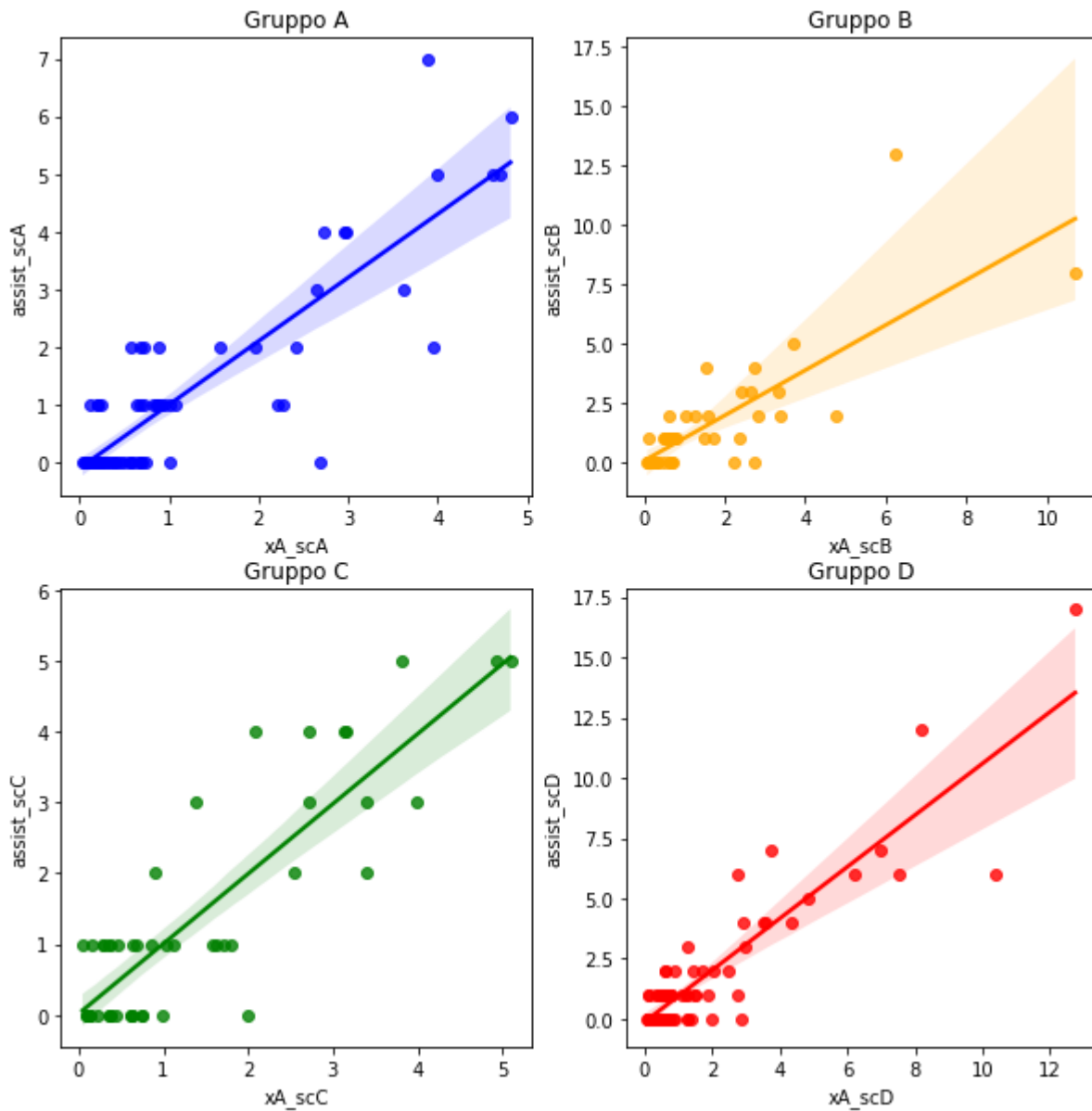


Come è possibile osservare dai grafici il valore di ROC AUC per i quattro casi presi in esame il modello è abbastanza affidabile.

Per valutare invece la bontà del modello dal punto di vista della precisione probabilistica dei singoli valori si è utilizzata la LogLoss.

Dai valori ottenuti (0.32 Gruppo A, 0.39 Gruppo B, 0.35 Gruppo C, 0.42 Gruppo D) si riscontra che il modello è abbastanza accurato.

Infine, sono stati messi in relazione gli expected assist derivanti dal modello statistico e gli assist registrati raggruppati per giocatore derivanti dai dati presenti nel dataset preso in esame per verificare se esiste una correlazione lineare tra le metriche.



Si può constatare che per valori di xA bassi (inferiori a 3) esiste una certa correlazione lineare, mentre per valori di xA più alti non esiste una buona correlazione.

In conclusione, alla luce delle considerazioni effettuate, il modello, seppure con alcuni limiti, può essere utilizzato per descrivere in maniera sommaria gli expected assist.

Per poter migliorare il modello potrebbe essere necessario:

1. costruire il modello su una base molto più ampia di dati derivanti dalle stagioni precedenti e/o da altri campionati;
2. considerare altri parametri come per esempio la posizione dei giocatori della squadra difendente e il tipo di difesa adottata (a uomo o a zona) oppure le *spizzate di testa*.

5.2 RISULTATI

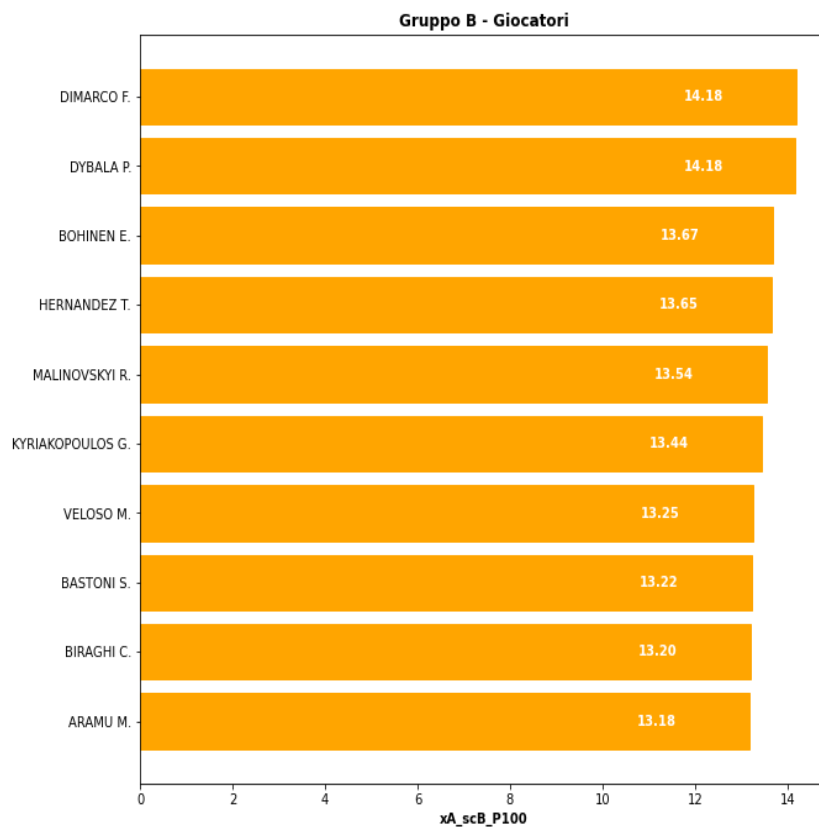
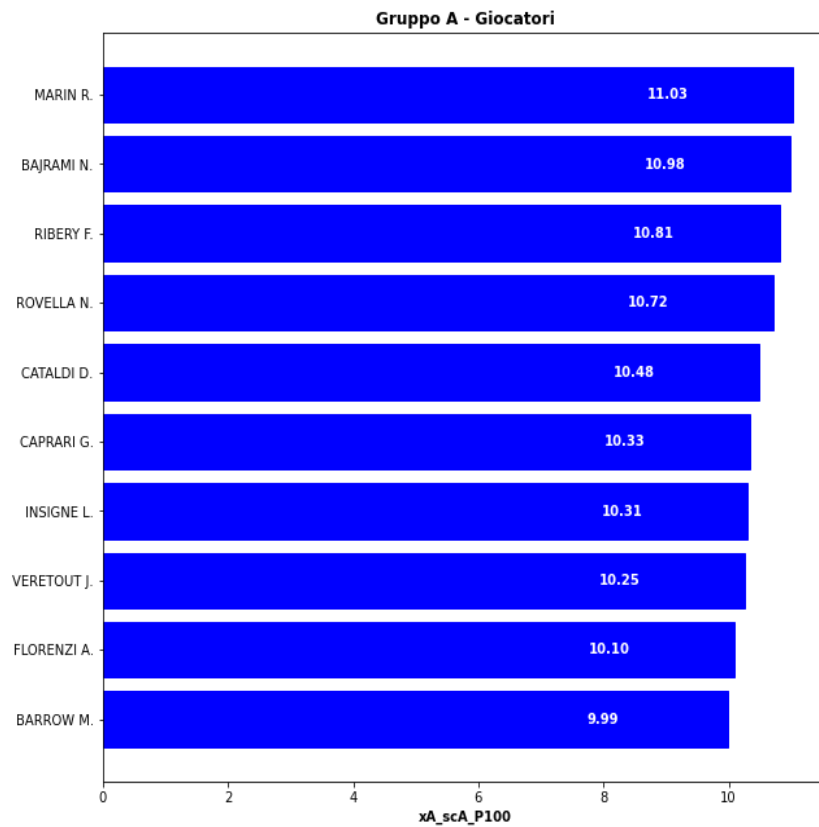
Detto che il modello può essere utilizzato per descrivere gli expected assist, sono stati messi a confronto gli expected assist totali di ogni gruppo per valutare in quale situazione si ottiene un beneficio offensivo.

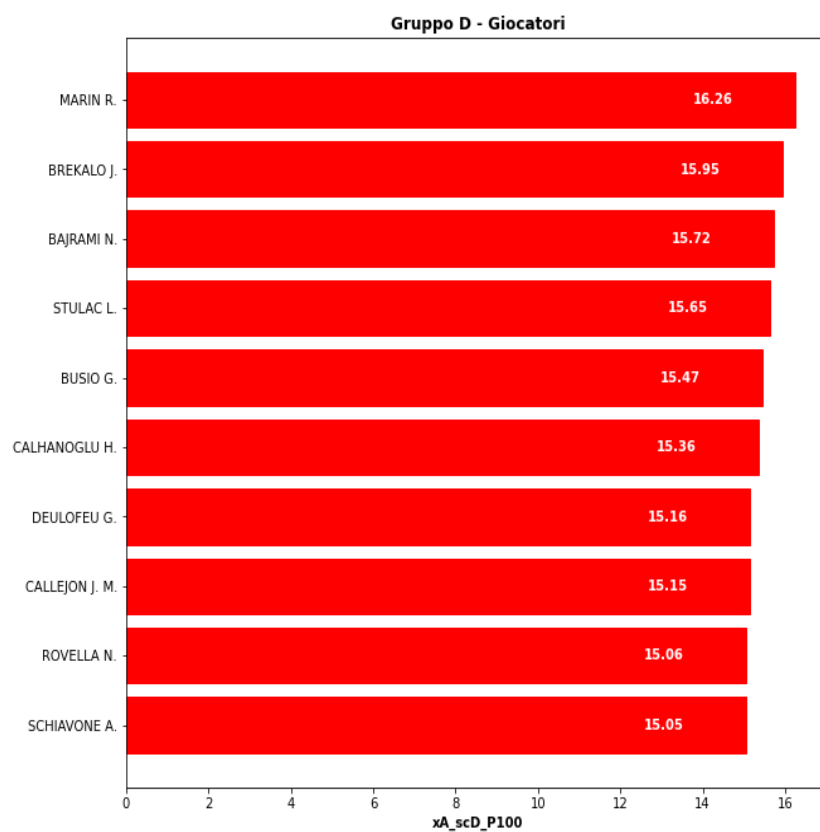
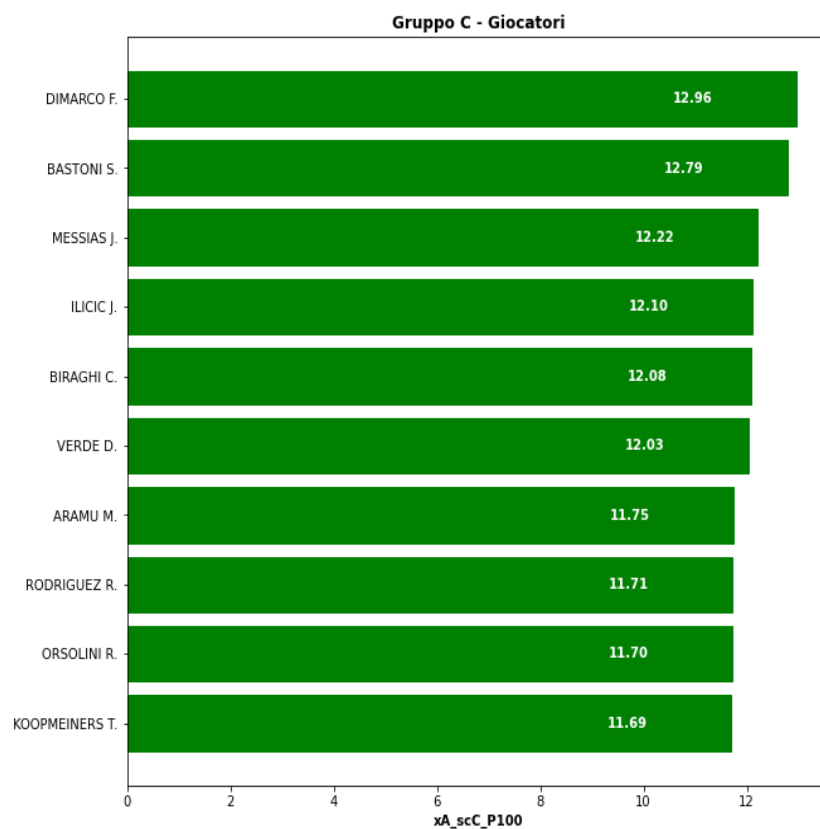
I dati sono stati normalizzati per 100 calci d'angolo per avere una base comune su cui fare il confronto data la diversa quantità di eventi per ogni situazione.

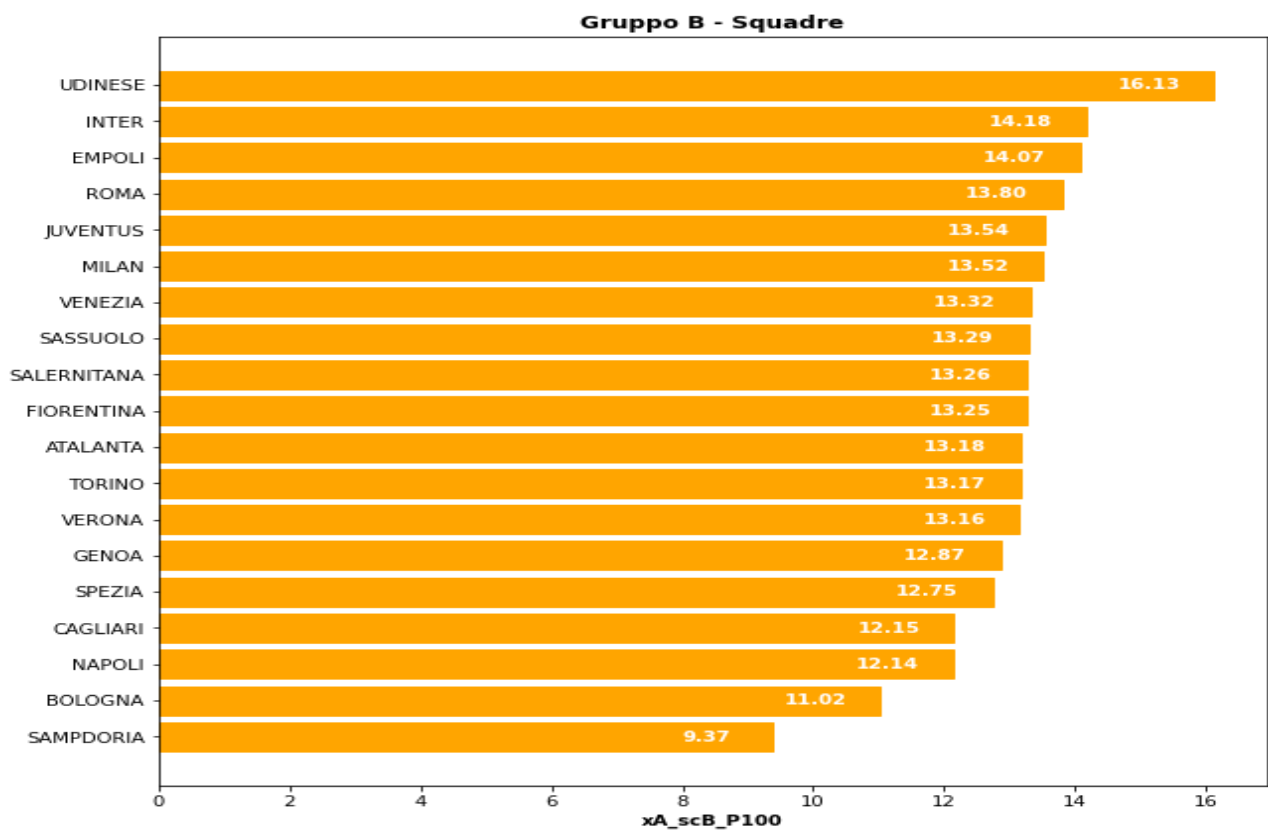
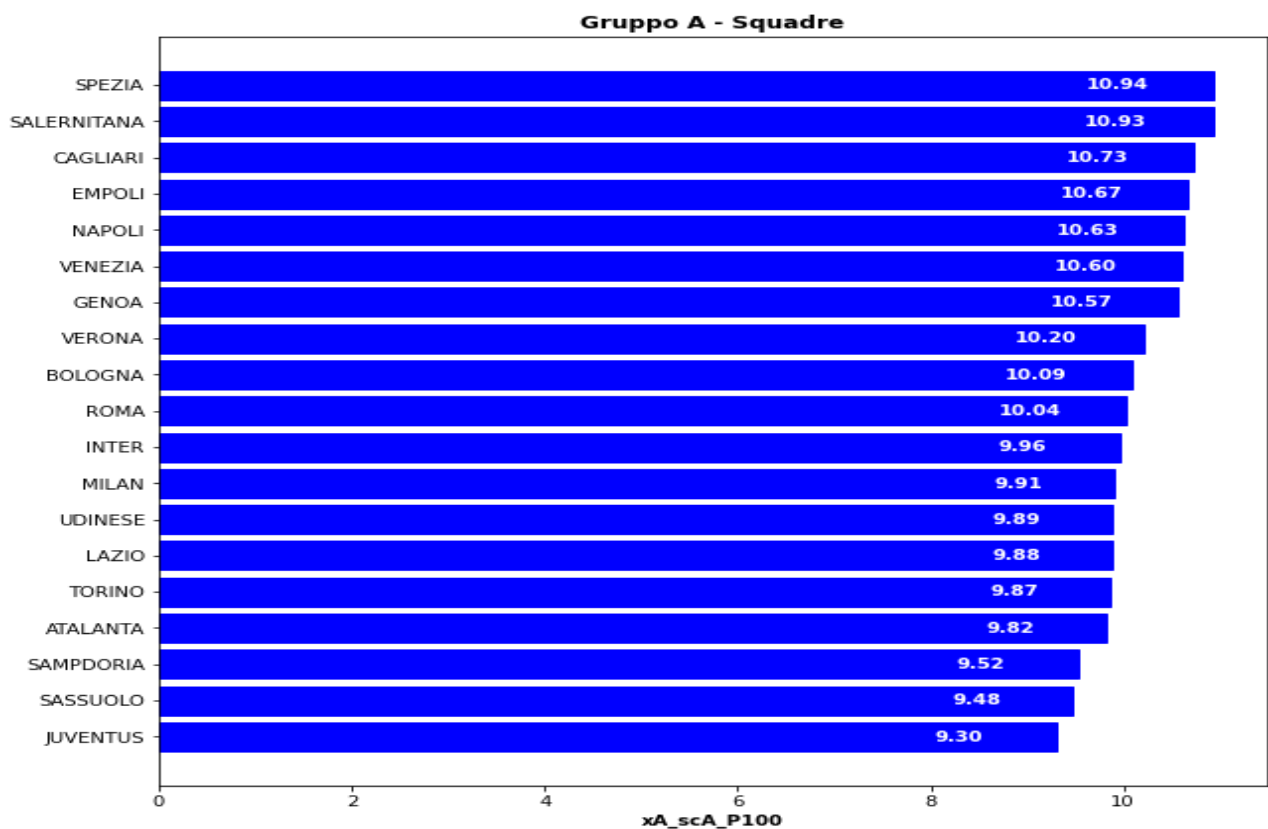
Dai valori si ottiene che gruppi B [13,23 xA] e D [14,93 xA] (cross da calcio d'angolo con effetto a rientrare) portano un beneficio offensivo maggiore in termini di assist rispetto ai gruppi A [10,13 xA] e C [11,36 xA] (cross da calcio d'angolo con effetto a uscire)

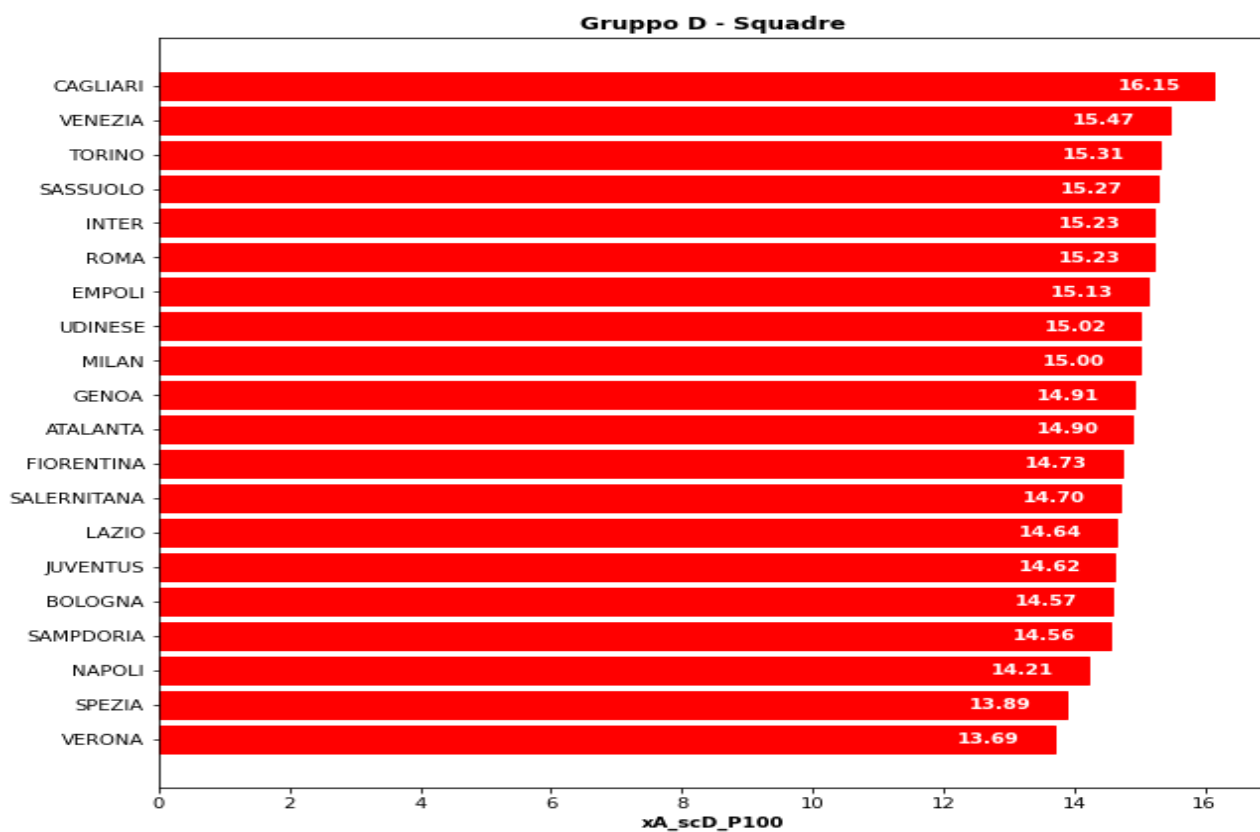
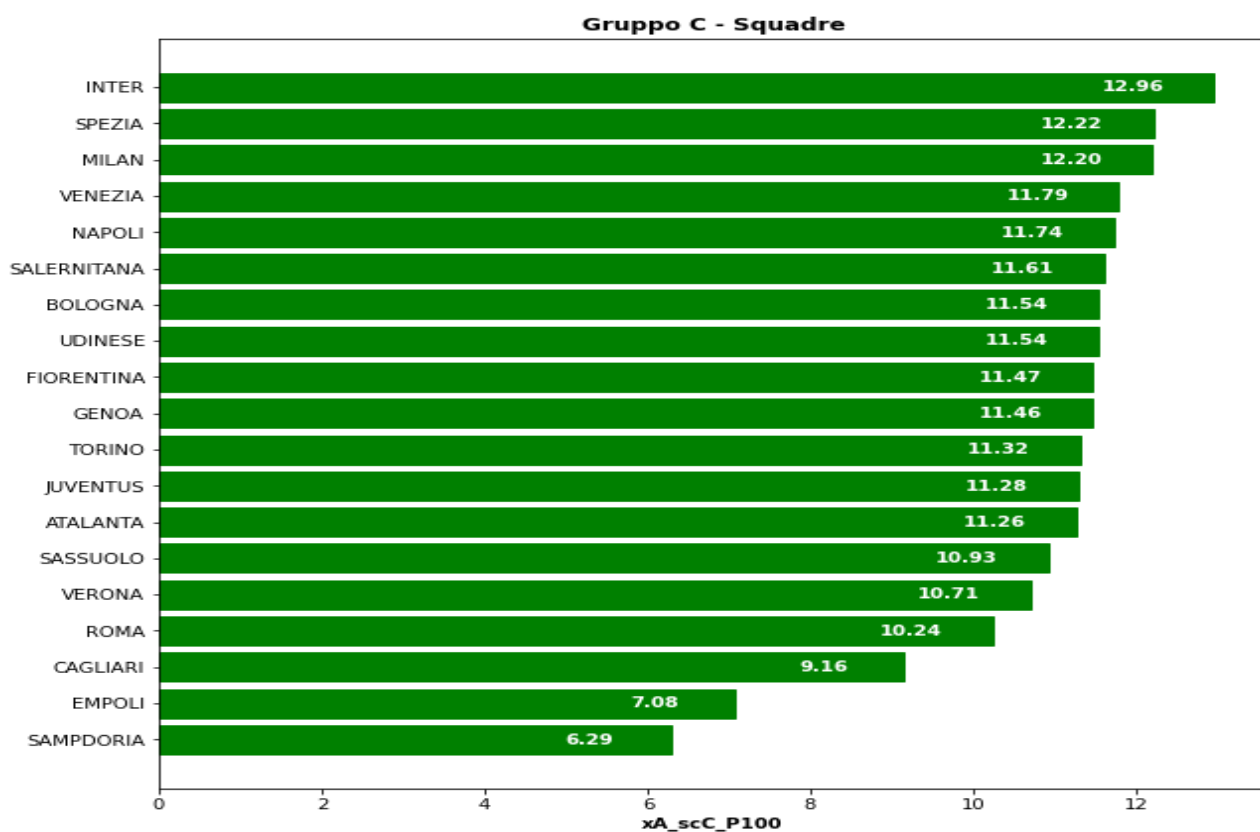
6. CONFRONTI

In conclusione vengono riportati i dati sui battitori (con almeno 10 calci d'angolo battuti) e sulle squadre che generano più expected assist nelle diverse situazioni









7. CONSIDERAZIONI FINALI

Ricapitolando:

- Il calcio d'angolo all'interno di una partita può rappresentare un'arma offensiva per consentire a una squadra di essere pericolosa e poter sbloccare il proprio tabellino marcatori;
- 4 i tipi di situazioni da calcio d'angolo: angolo destro, piede destro battitore (cross con effetto a uscire), angolo destro, piede sinistro battitore (cross con effetto a rientrare), angolo sinistro, piede sinistro battitore (cross con effetto a uscire), angolo sinistro, piede destro battitore (cross con effetto a rientrare);
- il modello statistico di regressione logistica per gli expected assist, allenato sui dati a disposizione, è quantitativamente abbastanza affidabile;
- il modello statistico di regressione logistica per gli expected assist, allenato sui dati a disposizione, è a livello di precisione probabilistica abbastanza accurato;
- esiste una certa correlazione lineare tra i valori del modello di expected assist e gli assist derivanti dagli eventi registrati;
- necessità di migliorare il modello statistico allenandolo su una base più ampia di dati derivanti dalle stagioni precedenti e/o da altri campionati e considerando ulteriori parametri di dipendenza;
- calci d'angolo con cross con effetto a rientrare portano un beneficio offensivo maggiore in termini di assist rispetto a calci d'angolo con cross con effetto a uscire.