
ANALISI DELLE PARTITE DI CALCIO NEI 5 MAGGIORI CAMPIONATI EUROPEI NELLE STAGIONI 2014-2020

Corso di Laurea Magistrale Ingegneria Informatica
e dell'Automazione



Autori

Enrico Maria Sardellini

Simone Di Battista

Marco Barbarella

Anno accademico 2023-2024

Indice

1	Introduzione al dataset	5
1.1	Descrizione	5
1.2	Struttura delle tabelle	5
2	Tableau	9
2.1	Introduzione	9
2.2	Caricamento dati	9
2.3	Data Analysis	10
2.3.1	Dashboard 1: "The best" - Analisi delle prestazioni di due calciatori a confronto negli anni	10
2.3.2	Dashboard 2: "Championship" - Analisi delle prestazioni delle squadre nel corso della stagione	13
2.3.3	Dashboard 3: "Shooters of the match" - Analisi di una singola partita relativa alle occasioni create e alla loro modalità	16
3	Qlik	21
3.1	Introduzione	21
3.2	Caricamento dati	21
3.3	Data Analysis	23
3.3.1	Dashboard 1: Serie A team analysis	23
3.3.2	Dashboard 2: Premier League team analysis	26
3.3.3	Dashboard 3: Shots Analysis	30
4	Power BI	34
4.1	Introduzione	34
4.2	Caricamento dati	34
4.3	Data Analysis	35
4.3.1	Dashboard 1: Analysis of goals, assists, shots and cards of players	35
4.3.2	Dashboard 2: Analysis of goals, xGoals, shot and shot on target of teams per season	38
4.3.3	Dashboard 3: Analysis of bookmakers bets for a match	40

Elenco delle figure

1	Logo di Tableau	9
2	Origine dati Tableau 1	10
3	Dashboard 1: "The best" - Analisi delle prestazioni di due calciatori a confronto negli anni	11
4	Dashboard 1 filtrata per giocatori e arco temporale	12
5	Origine dati Tableau 2	13
6	Dashboard 2: "Championship" - Analisi delle prestazioni delle squadre nel corso della stagione	14
7	Dashboard 2 filtrata per campionato e arco temporale	16
8	Origine dati Tableau 3	17
9	Dashboard 3: "Shooters of the match" - Analisi di una singola partita relativa alle occasioni create e alla loro modalità	18
10	Dashboard 3 filtrata per squadre e arco temporale	19
11	Logo di Qlik	21
12	Origine dati Qlik	22
13	Dashboard 1: Serie A team analysis	24
14	Dashboard 1 filtrata per squadra e stagione	25
15	Dashboard 2: Premier League team analysis	28
16	Dashboard 2 filtrata per le squadre	29
17	Dashboard 3: Shots Analysis	31
18	Dashboard 3 filtrata per calciatore	32
19	Logo di Power BI	34
20	Origine dati Power BI	35
21	Dashboard 1: Analysis of goals, assists, shots and cards of players	36
22	Dashboard 1 filtrata per giocatori e stagione	37
23	Dashboard 2: Analysis of goals, xGoals, shot and shot on target of teams per season	39
24	Dashboard 2 filtrata per lega, squadre e stagione	40
25	Dashboard 3: Analysis of bookmakers bets for a match	41
26	Dashboard 3 filtrata per squadre e stagioni	42

Elenco delle tabelle

1	Campi di <code>Appearances.csv</code>	6
2	Campi di <code>Games.csv</code>	7
3	Campi di <code>Leagues.csv</code>	7
4	Campi di <code>Players.csv</code>	7
5	Campi di <code>Shots.csv</code>	8
6	Campi di <code>Teams.csv</code>	8
7	Campi di <code>Teamstats.csv</code>	8

1 Introduzione al dataset

Questa tesina fornisce un'analisi completa delle partite dei principali campionati europei, concentrandosi sulle informazioni chiave riguardanti squadre, giocatori, tiri in porta, risultati delle partite e altre statistiche dettagliate. Il dataset analizzato comprende le partite giocate nel periodo 2014-2020.

1.1 Descrizione

Il dataset preso in considerazione per questo progetto fornisce dettagliati dati sulle partite e sui calciatori, coprendo il periodo dal 2014 al 2020.

La raccolta comprende informazioni riguardanti risultati, gol, gol attesi, minutaggio, posizione dei calciatori, cartellini e caratteristiche delle azioni. È una risorsa completa e interessante per ottenere informazioni approfondite sulle dinamiche delle partite e sulle performance dei singoli giocatori.

È possibile accedere a questo dataset tramite il seguente link:

<https://www.kaggle.com/datasets/technika148/football-database>.

1.2 Struttura delle tabelle

Il dataset è composto da sette file `.csv` :

- **Appearances.csv** : ogni riga rappresenta un calciatore che ha giocato in una specifica partita. La chiave principale è data dalle colonne `gameID` e `playerID` , con le altre colonne che mostrano le prestazioni del giocatore in tale partita.
- **Games.csv** : ogni riga rappresenta la singola partita identificata dalla colonna `gameID` . Da questo file si può ottenere il risultato della partita, le squadre che si sono affrontate, la data in cui si è svolta e le aspettative che c'erano.
- **League.csv** : ogni riga rappresenta uno dei 5 campionati in cui si sono giocate le partite prese in analisi.
- **Players.csv** : ogni riga rappresenta uno dei calciatori che hanno giocato nelle partite prese in analisi.
- **Shots.csv** : ogni riga rappresenta un singolo tiro che è avvenuto in una specifica partita. Le colonne mostrano chi ha tirato il pallone, da chi gli è provenutto, come lo ha calciato e in quale situazione.
- **Teams.csv** : ogni riga rappresenta una delle squadre che hanno giocato nelle partite prese in analisi
- **Teamstats.csv** : ogni riga rappresenta l'andamento che ha avuto una squadra in quella specifica partita, mostrando i gol fatti, il risultato, il numero dei falli ed i cartellini.

Di seguito sono forniti per esteso gli elenchi completi dei campi delle due tabelle.

Tabella 1: Campi di `Appearances.csv`

Campo	Descrizione
<code>gameID</code>	ID della partita
<code>playerID</code>	ID del giocatore
<code>goals</code>	Gol
<code>ownGoals</code>	Autogol
<code>shots</code>	Tiri effettuati
<code>xGoals</code>	Gol attesi
<code>xGoalsChain</code>	Gol attesi di ogni possesso in cui il giocatore è coinvolto
<code>xGoalsBuildup</code>	Gol attesi senza passaggi chiave e tiri
<code>assists</code>	Assist
<code>keyPasses</code>	Passaggi chiave
<code>xAssists</code>	Assist attesi
<code>position</code>	Posizione in campo (testuale)
<code>positionOrder</code>	Posizione in campo (numerica)
<code>yellowCard</code>	Ammonizioni
<code>redCard</code>	Espulsioni
<code>time</code>	Tempo giocato
<code>substituteIn</code>	Giocatore subentrato
<code>substituteOut</code>	Giocatore sostituito
<code>leagueID</code>	ID della lega

Tabella 2: Campi di `Games.csv`

Campo	Descrizione
gameID	ID univoco della partita
leagueID	ID della lega
season	Stagione
date	Data e ora della partita
homeTeamID	ID della squadra di casa
awayTeamID	ID della squadra ospite
homeGoals	Gol segnati dalla squadra di casa
awayGoals	Gol segnati dalla squadra ospite
homeProbability	Probabilità di vittoria della squadra di casa per il sito Understat
drawProbability	Probabilità di pareggio per il sito Understat
awayProbability	Probabilità di vittoria della squadra ospite per il sito Understat
homeGoalsHalfTime	Gol segnati dalla squadra di casa nel primo tempo
awayGoalsHalfTime	Gol segnati dalla squadra ospite nel primo tempo
B365H	Quota Bet365 per la vittoria della squadra di casa
B365D	Quota Bet365 per il pareggio
B365A	Quota Bet365 per la vittoria della squadra ospite
BWH	Quota Betway per la vittoria della squadra di casa
BWD	Quota Betway per il pareggio
BWA	Quota Betway per la vittoria della squadra ospite
IWH	Quota Interwetten per la vittoria della squadra di casa
IWD	Quota Interwetten per il pareggio
IWA	Quota Interwetten per la vittoria della squadra ospite
PSH	Quota Pinnacle Sports per la vittoria della squadra di casa
PSD	Quota Pinnacle Sports per il pareggio
PSA	Quota Pinnacle Sports per la vittoria della squadra ospite
WHH	Quota William Hill per la vittoria della squadra di casa
WHD	Quota William Hill per il pareggio
WHA	Quota William Hill per la vittoria della squadra ospite
VCH	Quota VC Bet per la vittoria della squadra di casa
VCD	Quota VC Bet per il pareggio
VCA	Quota VC Bet per la vittoria della squadra ospite
PSCH	Quota Pinnacle Sports chiusa per la vittoria della squadra di casa
PSCD	Quota Pinnacle Sports chiusa per il pareggio
PSCA	Quota Pinnacle Sports chiusa per la vittoria della squadra ospite

Tabella 3: Campi di `Leagues.csv`

Campo	Descrizione
leagueID	ID univoco della lega
name	Nome della lega
understatNotation	Notazione della lega usata su Understat

Tabella 4: Campi di `Players.csv`

Campo	Descrizione
playerID	ID univoco del giocatore
name	Nome del giocatore

Tabella 5: Campi di `Shots.csv`

Campo	Descrizione
<code>gameID</code>	ID della partita
<code>shooterID</code>	ID del giocatore che ha effettuato il tiro
<code>assisterID</code>	ID del giocatore che ha fatto l'assist
<code>minute</code>	Minuto in cui è stato effettuato il tiro
<code>situation</code>	Situazione di gioco in cui è stato effettuato il tiro
<code>lastAction</code>	Ultima azione prima che il tiro fosse effettuato
<code>shotType</code>	Tipo di tiro
<code>shotResult</code>	Risultato del tiro
<code>xGoal</code>	Valore degli Expected Goals (xG) del tiro
<code>positionX</code>	Posizione (asse lungo del campo) da cui è stato effettuato il tiro
<code>positionY</code>	Posizione (asse corto del campo) da cui è stato effettuato il tiro

Tabella 6: Campi di `Teams.csv`

Campo	Descrizione
<code>teamID</code>	ID univoco della squadra
<code>name</code>	Nome della squadra

Tabella 7: Campi di `Teamstats.csv`

Campo	Descrizione
<code>gameID</code>	ID della partita
<code>teamID</code>	ID della squadra
<code>season</code>	Stagione
<code>date</code>	Data e orario della partita
<code>location</code>	Casa (H) o trasferta (A)
<code>goals</code>	Gol segnati
<code>xGoals</code>	Gol attesi
<code>shots</code>	Tiri
<code>shotsOnTarget</code>	Tiri in porta
<code>deep</code>	Passaggi completati entro 18 metri dalla porta (esclusi i cross)
<code>ppda</code>	Passaggi concessi per azione difensiva nella metà campo avversaria
<code>fouls</code>	Falli
<code>corners</code>	Calci d'angolo
<code>yellowCards</code>	Cartellini gialli
<code>redCards</code>	Cartellini rossi
<code>result</code>	Vittoria (W), pareggio (D), sconfitta (L)

2 Tableau

Passiamo ora alle analisi effettuate con il primo software, Tableau. Inizieremo con una breve descrizione del software e del processo di caricamento dei dati. Successivamente, forniremo un elenco delle analisi realizzate, insieme alle visualizzazioni presenti nella dashboard, accompagnate da spiegazioni dettagliate per ciascuna..

2.1 Introduzione

Tableau è un software di data visualization che fornisce una piattaforma flessibile per esplorare, analizzare e comunicare informazioni tramite grafici interattivi, dashboard e report. Creato con l'obiettivo di facilitare la comprensione dei dati, Tableau permette agli utenti di trasformare facilmente informazioni complesse in rappresentazioni visive chiare e intuitive.



Figura 1: Logo di Tableau

Uno dei principali vantaggi di Tableau è la sua facilità d'uso, che consente anche a chi non ha esperienza nell'analisi dei dati di creare visualizzazioni accattivanti. La piattaforma si distingue per la sua flessibilità, supportando un'ampia varietà di fonti di dati e offrendo integrazioni con numerosi strumenti e piattaforme.

Tableau Desktop è la versione dell'applicazione pensata per un utilizzo individuale, fornendo un ambiente locale per creare, analizzare e condividere visualizzazioni dei dati. Viene apprezzato per la sua interfaccia intuitiva *drag-and-drop*, che permette di trascinare facilmente dimensioni e misure per generare grafici e dashboard interattivi. Con Tableau Desktop, è possibile collegarsi a diverse fonti di dati, incluse basi di dati relazionali, fogli di calcolo, servizi cloud e altre tipologie.

Un aspetto di rilievo è la capacità di interagire in tempo reale con i dati, consentendo agli utenti di esplorare dinamicamente le informazioni e ottenere immediatamente nuovi spunti. In alternativa, è possibile eseguire un'estrazione dei dati utilizzando l'*Hyper Data Engine* di Tableau per operare su di essi, migliorando la velocità di esecuzione delle singole query.

2.2 Caricamento dati

La prima operazione da effettuare, come negli altri software, è il caricamento dei dati dalla sorgente. In Tableau, ciò è notevolmente semplificato grazie ad un meccanismo *drag and drop* con il quale si trascinano i file `.csv` nella schermata centrale dove poi saranno effettuati i *join* e rappresentate graficamente le relazioni tra le varie tabelle.

In questo caso, dunque, sono stati effettuati molteplici e vari *join* data la quantità di file csv a disposizione. Per effettuare un *join* basta semplicemente selezionare i campi e l'operatore "=" dall'interfaccia fornita.

Una volta fatto ciò, l'interprete di Tableau si occupa di verificare che non ci siano errori di lettura dai file originali, così da poter procedere con la costruzione dei fogli e delle dashboard.

2.3 Data Analysis

Di seguito sono presentate le dashboard realizzate e saranno discusse le principali considerazioni effettuate su di esse. Nello specifico, le dashboard in questione sono tre, ognuna che ruota attorno ad una tematica specifica:

- Dashboard 1: "The best" - Analisi delle prestazioni di due calciatori a confronto negli anni.
- Dashboard 2: "Championship" - Analisi delle prestazioni delle squadre nel corso della stagione.
- Dashboard 3: "Shooters of the match" - Analisi di una singola partita relativa alle occasioni create e alla loro modalità.

In tutte le dashboard sono contenuti: i grafici realizzati separatamente nei rispettivi fogli di lavoro, le legende relative ai colori utilizzati per differenziare le rappresentazioni e le finestre con i filtri interattivi selezionabili dall'utente nella parte destra.

2.3.1 Dashboard 1: "The best" - Analisi delle prestazioni di due calciatori a confronto negli anni

Questa dashboard, come riportato nella Figura 2, presenta i seguenti *join*:

- $\text{gameID } \{appearances.csv\} = \text{gameID } \{games.csv\};$
- $\text{playerID } \{appearances.csv\} = \text{playerID } \{players.csv\};$
- $\text{playerID } \{players.csv\} = \text{shooterID } \{shots.csv\};$
- $\text{leagueID } \{games.csv\} = \text{leagueID } \{leagues.csv\};$
- $\text{homeTeamID } \{games.csv\} = \text{teamID } \{teams.csv\};$
- $\text{teamID } \{teams.csv\} = \text{teamID } \{teamstats.csv\};$

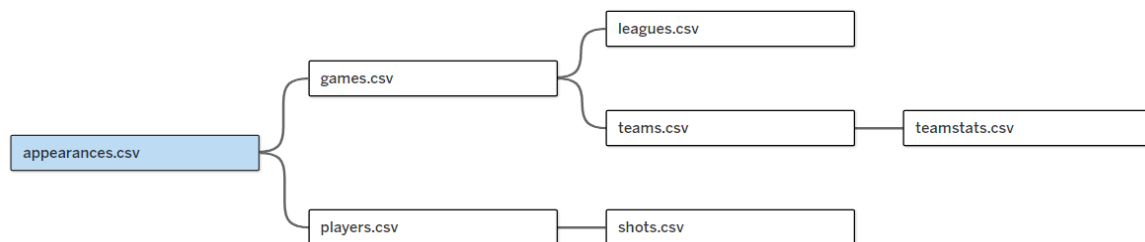


Figura 2: Origine dati Tableau 1

La prima analisi diagnostica effettuata con Tableau è di tipo diagnostico mette a confronto l'andamento sportivo di due calciatori in un determinato lasso di tempo.

La dashboard, come riportato nella Figura 3, è divisa in tre tipologie di fogli duplicati per mostrare il confronto:

- *a sinistra* si trova un grafico a linee che mostra l'andamento stagionale dei calciatori considerando il numero di gol e assist realizzati.

- *in alto al centro* è presente una mappa di calore rappresentante in ordine decrescente le posizioni che ha ricoperto il singolo calciatore nell'arco temporale considerato.
- *in basso al centro* si trova un grafico a torta che mostra la percentuale di partite in cui il calciatore ha preso o non ha preso un cartellino giallo o rosso o entrambi.

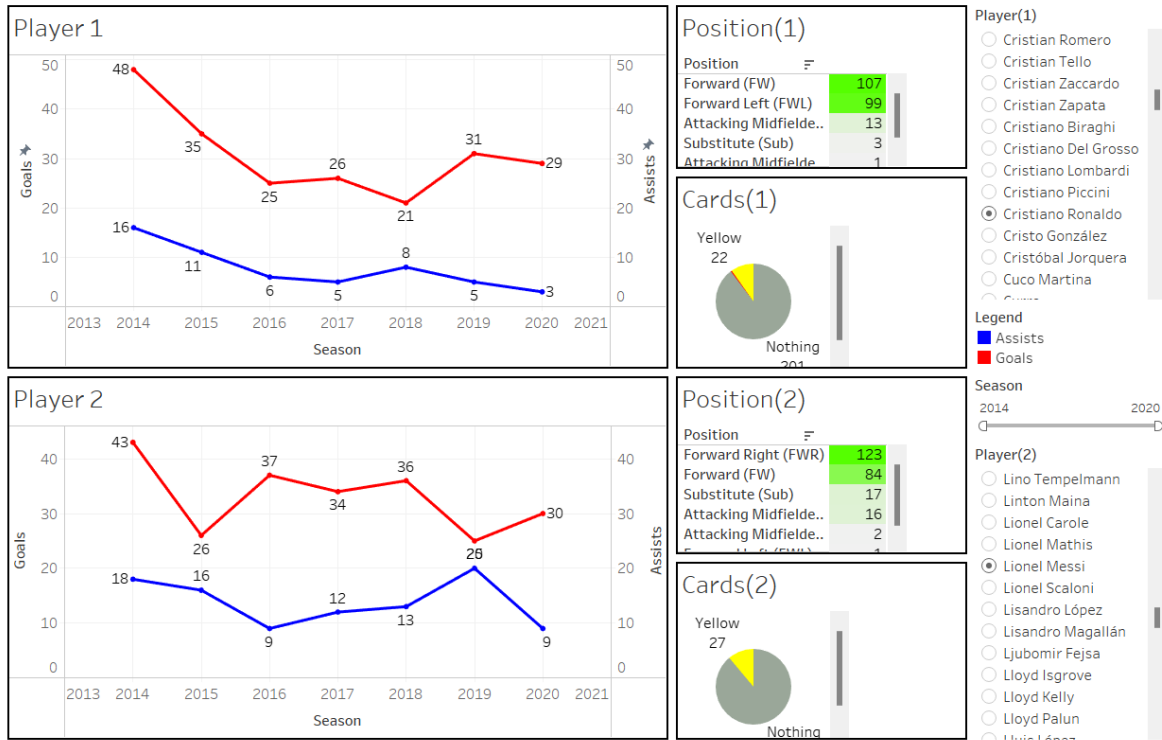


Figura 3: Dashboard 1: "The best" - Analisi delle prestazioni di due calciatori a confronto negli anni

Anche senza applicare alcun filtro, è già possibile effettuare alcune considerazioni osservando i grafici.

Per esempio, possiamo notare che, dal grafico a linee, entrambi i giocatori hanno avuto un picco di prestazioni in termini di gol segnati in determinate stagioni, ma con una fluttuazione delle loro prestazioni negli anni successivi. Questa variabilità potrebbe essere legata a fattori come il cambio di squadra o infortuni o variazioni nella posizione. Agendo sul filtro *Season* si può andare a ridurre l'arco temporale, limitandolo ad una specifica stagione, per verificare come il numero di partite e il cambio di posizione abbia agito sulle prestazioni del singolo.

La mappa di calore, quindi, mette in evidenza le diverse posizioni occupate dai giocatori in campo. Si può dedurre la versatilità di un giocatore in base al numero di posizioni diverse in cui ha giocato. Si può altresì notare l'importanza di un determinato giocatore dal numero di valori "Substitution" presenti. Per esempio, un giocatore che ha giocato in più ruoli potrebbe essere considerato più flessibile e utile per l'allenatore rispetto a un giocatore che ha occupato prevalentemente una sola posizione. Allo stesso tempo un numero basso di gare da subentrato dalla panchina insieme ad un elevato numero di gol e assist ci mostrano un giocatore centrale nel progetto della squadra.

Il grafico a torta, invece, offre una chiara rappresentazione del comportamento disciplinare dei giocatori in campo, mostrando la frequenza con cui ricevono cartellini gialli o rossi. Questo grafico mi

determina l'affidabilità di un giocatore. Infatti un calciatore che riceve un gran numero di cartellini è adatto ad un determinato tipo di squadra, con uno stile di gioco più difensivo, piuttosto che ad un altro che a parità di uomini è tecnicamente superiore all'altra. Anche le posizioni occupate mi incidono sul numero di cartellini. Un difensore è più propenso a prenderne un numero maggiore rispetto ad un attaccante.

Oltre a ciò, utilizzare i filtri ci consente di confrontare tutti i calciatori presenti nel database. Nello specifico, i filtri presenti in tale dashboard sono:

- filtro sul giocatore 1;
- filtro sul giocatore 2;
- filtro sull'arco temporale (2014-2020).

Nella Figura 4 abbiamo, infatti, applicato tutti i filtri alla dashboard: variando entrambi i calciatori e l'arco temporale considerato.



Figura 4: Dashboard 1 filtrata per giocatori e arco temporale

La dashboard offre la possibilità di fare un confronto diretto tra due giocatori, mettendo in mostra principalmente le capacità realizzative ma anche la versatilità in campo ed il comportamento disciplinare.

In particolare si può notare come non si hanno dati relativi al *Player 2: Marco Borriello* nel 2018 e 2019. Questo può essere dovuto o ad un lungo infortunio che lo ha tenuto bloccato per tutta la stagione, oppure risultava in quel periodo svincolato o in una squadra non appartenente ai campionati del database. Per quanto riguarda *Player 1: Manolo Gabbiadini* si può dire che abbia un andamento più costante rispetto all'altro calciatore ma soprattutto che, nonostante non sia un titolare fisso essendo

partito dalla panchina per quasi la metà delle gare, abbia inciso sempre con alcuni picchi in determinati anni.

Analizzando quindi le tendenze passate, si può tentare di prevedere le performance future dei giocatori risultando utile non solo alle squadre o allenatori che necessitano di acquistare qualcuno da inserire nella loro rosa, ma anche ad appassionati di fantacalcio che devono valutare su quali difensori, centrocampisti o attaccanti puntare e su quali no.

2.3.2 Dashboard 2: "Championship" - Analisi delle prestazioni delle squadre nel corso della stagione

Questa dashboard, come riportato nella Figura 5, presenta i seguenti *join*:

- $\text{gameID } \{\text{appearances.csv}\} = \text{gameID } \{\text{games.csv}\};$
- $\text{playerID } \{\text{appearances.csv}\} = \text{playerID } \{\text{players.csv}\};$
- $\text{playerID } \{\text{players.csv}\} = \text{shooterID } \{\text{shots.csv}\};$
- $\text{leagueID } \{\text{games.csv}\} = \text{leagueID } \{\text{leagues.csv}\};$
- $\text{homeTeamID } \{\text{games.csv}\} = \text{teamID } \{\text{teamstats.csv}\};$
- $\text{teamID } \{\text{teamstats.csv}\} = \text{teamID } \{\text{teams.csv}\};$

La differenza con i join della dashboard precedente sta nello scambio tra *teamstats.csv* e *teams.csv*, col fine di risolvere un problema che si è creato tra *leagues.csv* e *teamstats.csv*

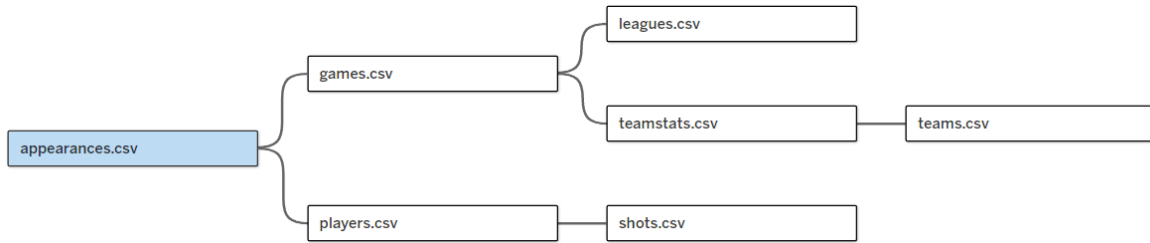


Figura 5: Origine dati Tableau 2

La seconda dashboard realizzata con Tableau è anch'essa di tipo diagnostico e il suo scopo è quello di analizzare l'andamento di tutte le squadre di un determinato campionato e in una determinata stagione.

Tale dashboard, come riportato nella Figura 6, è divisa in quattro fogli:

- *in alto a sinistra* si trova una grafico a barre orizzontali che calcola i punti ottenuti da ogni squadra considerando i risultati di ogni partita giocata fino a quel momento;
- *in alto a destra* è presente una tabella che riporta non solo il numero di tiri compiuti da ogni squadra in quell'arco temporale, ma anche il numero di quei tiri che era diretto verso la porta avversaria e, di questi, il numero che si è poi trasformato in una rete;
- *in basso a sinistra* si trovano due grafici a barre che mostrano, per ogni squadra, il numero di cartellini gialli e rossi presi nel corso della stagione.

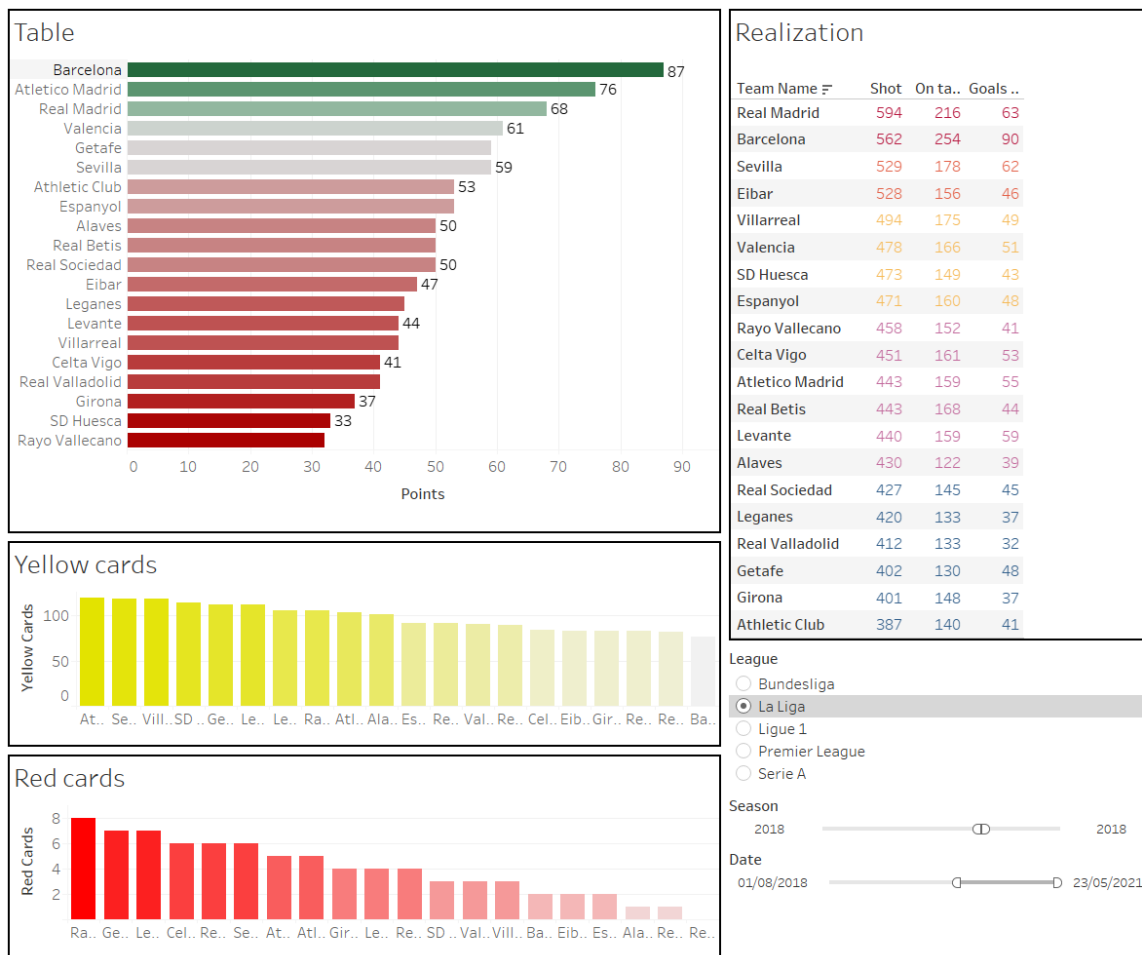


Figura 6: Dashboard 2: "Championship" - Analisi delle prestazioni delle squadre nel corso della stagione

Anche senza applicare alcun filtro, è possibile già trarre alcune considerazioni interessanti dai grafici e dalle visualizzazioni proposte.

Osservando il grafico a barre orizzontali *Table*, notiamo come il *Barcellona* domini la stagione con un punteggio di 87, seguito dall'*Atletico Madrid* con 76 punti e dal *Real Madrid* con 68. Le squadre come *Rayo Vallecano* e *SD Huesca*, che occupano gli ultimi posti, mostrano un distacco notevole sia in termini di risultati che di competitività.

La tabella *Realization* riporta dettagli importanti sui tiri totali, tiri in porta e gol segnati da ciascuna squadra. Ad esempio, il *Real Madrid* risulta avere il maggior numero di tiri totali (594), ma non si traduce nel maggior numero di gol segnati, suggerendo una potenziale inefficienza nella finalizzazione. Al contrario, squadre come il *Barcellona*, con un numero di tiri inferiore (562), mostrano una maggiore precisione, segnando 90 gol. Questo confronto suggerisce che le differenze non dipendono solo dal volume di tiri, ma anche dalla qualità del gioco. Infine osservando l'*Atletico Madrid*, che occupa la seconda posizione in classifica, sprofonda nella tabella occupando la undicesima posizione per tiri, ma anche per tiri in porta risultando nettamente distaccata dalle prime due.

Il grafico *Yellow cards* offre una chiara panoramica del comportamento disciplinare delle squadre. Notiamo come l'*Atletico Madrid* sia la squadra con il numero più alto di cartellini gialli, il che potrebbe

riflettere uno stile di gioco aggressivo o molto difensivo. Al contrario, squadre come il *Barcellona* e il *Real Madrid*, che si trovano tra le prime posizioni della classifica, mostrano un numero inferiore di ammonizioni, suggerendo uno stile di gioco più tecnico e controllato.

Analogamente, il grafico *Red cards* evidenzia squadre come il *Rayo Vallecano* e il *Girona* con il maggior numero di espulsioni. Questo dato potrebbe indicare difficoltà difensive che spingono i giocatori a commettere falli gravi, incidendo negativamente sulle loro prestazioni complessive. Squadre come il *Barcellona* e l'*Atletico Madrid*, invece, si distinguono per un numero limitato di cartellini rossi, il che contribuisce a mantenere una rosa competitiva in campo durante le partite.

Come per la precedente dashboard, anche in questa è possibile applicare alcuni filtri per andare più nel dettaglio con le analisi. I filtri che si trovano in tale dashboard riguardano:

- filtro sulla lega;
- filtro sulle stagioni (anche se è consigliato prendere in considerazione una stagione alla volta);
- filtro sulla data (ogni stagione va dal 1/8 di quell'anno al 31/7 dell'anno successivo).

A scopo dimostrativo, nella Figura 7 viene cambiata sia la lega, che l'anno, andando sempre a considerarne uno solo, ed effettuato un ulteriore filtraggio sulla data. Considerando sia il filtraggio dalla data che del tiro si è preso in riferimento il periodo che va dal 1/8/2020 al 3/2/2021.

La nuova classifica parziale evidenzia una lotta per il primo posto serrata con pochi punti di distacco tra la prima squadra, ovvero il *Manchester City*, e le inseguatrici. Nelle posizioni di fondo classifica troviamo squadre come il *West Bromwich Albion* (12 punti) e lo *Sheffield United* (11 punti), evidenziando la loro difficoltà a competere in questo inizio di stagione.

La tabella *Realization* ci mostra che il *Liverpool* guida la classifica per numero di tiri totali (345), con 43 gol segnati, dimostrando una buona capacità di finalizzazione. Così come le altre squadre che si trovano nelle posizioni più alte in classifica. Squadre di bassa classifica come il *West Bromwich Albion* e lo *Sheffield United*, evidenziano una scarsa capacità offensiva. Questo suggerisce che i problemi di queste squadre non siano solo difensivi ma anche nella creazione e finalizzazione delle occasioni.

Questo viene anche confermato dal numero di cartellini gialli presi a differenza delle prime classificate che registrano un numero inferiore di cartellini, confermando uno stile di gioco più tecnico e disciplinato.

Dal grafico *Red cards*, notiamo che l'*Arsenal*, una delle pretendenti al titolo inglese, è la squadra con il maggior numero di espulsioni (5), un dato che potrebbe aver inciso negativamente sui loro risultati. Mentre squadre della parte alta della classifica non registrano espulsioni, confermando un maggiore controllo in campo.

Analizzando questa dashboard, risulta evidente come possa essere un utile strumento per comprendere le performance delle squadre e il loro comportamento in campo. Può supportare allenatori e staff tecnici nell'identificare punti di forza e debolezze, come l'efficienza offensiva o il comportamento disciplinare, oltre a fornire informazioni preziose per decisioni strategiche.

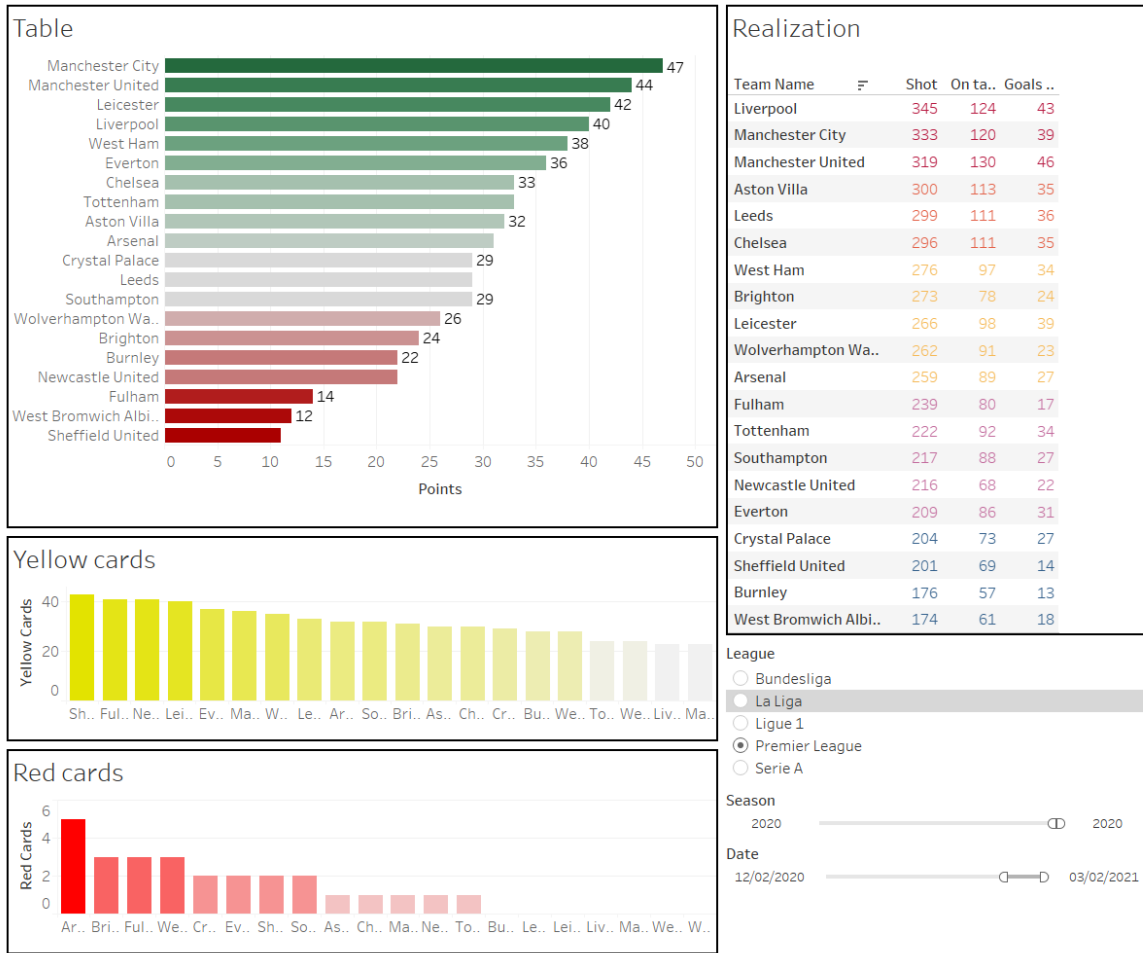


Figura 7: Dashboard 2 filtrata per campionato e arco temporale

2.3.3 Dashboard 3: "Shooters of the match" - Analisi di una singola partita relativa alle occasioni create e alla loro modalità

Questa dashboard, come riportato nella Figura 8, presenta i seguenti *join*:

- $\text{gameID } \{appearances.csv\} = \text{gameID } \{games.csv\};$
- $\text{leagueID } \{games.csv\} = \text{leagueID } \{leagues.csv\};$
- $\text{homeTeamID } \{games.csv\} = \text{teamID } \{teams.csv\};$
- $\text{teamID } \{teams.csv\} = \text{teamID } \{teamstats.csv\};$
- $\text{gameID } \{games.csv\} = \text{gameID } \{shots.csv\};$
- $\text{shooterID } \{shots.csv\} = \text{playerID } \{assister.csv\};$
- $\text{assisterID } \{shots.csv\} = \text{playerID } \{shooter.csv\};$

Le differenze con i join precedenti in questa origine sono dovute al fatto che questa dashboard si concentra principalmente su *shots.csv* e serviva un join diretto con *games.csv*. Inoltre Tableau non

permette il doppio join tra due file, quindi è stato necessario inserire un altro *players.csv* e rinominarli *assister.csv* e *shooter.csv* al fine di distinguerli meglio.

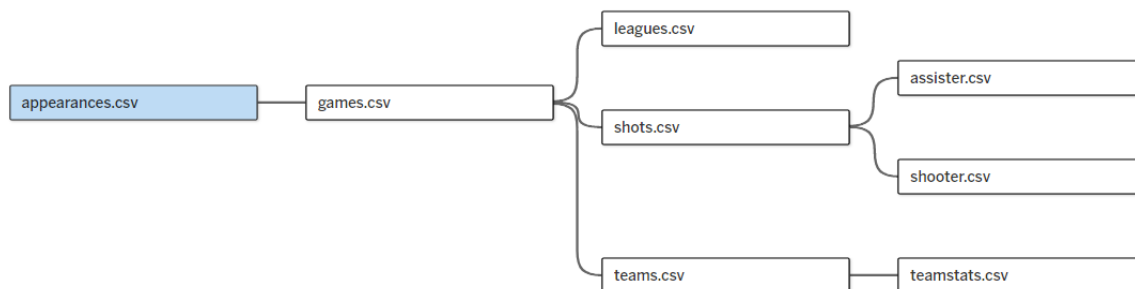


Figura 8: Origine dati Tableau 3

La terza e ultima dashboard realizzata mostra le azioni create nel corso di una specifica partita, nell'arco di una o più stagioni, ed il risultato finale mettendo in risalto le caratteristiche delle azioni. Viene mostrata inoltre una partita in maniera simile a quella riportata nei vari siti di telecronaca sportiva.

La dashboard in questione, come riportato nella Figura 9, è divisa in quattro fogli:

- *in alto* troviamo una tabella che mostra il calciatore che ha calciato in uno specifico minuto e il risultato della propria conclusione;
- *in basso a sinistra* è presente un grafico a bolle che mostra, in base alla dimensione delle bolle, come maggiormente sono stati realizzati i tiri.
- *al centro* è presente una semplice tabella con le due squadre che si sono affrontate ed il risultato finale della partita.
- *in basso a destra* è presente un grafico ad albero mostrante le principali situazioni da cui sono emerse le varie possibilità di tiro.

Anche senza applicare alcun filtro, è già possibile effettuare alcune considerazioni osservando i grafici.

La tabella *Shooter* elenca i giocatori che hanno tirato in porta, con l'indicazione del minuto e del risultato del tiro, evidenziando i diversi esiti con colori distintivi dove il rosso viene usato per il colore del gol, che si distingue in mezzo agli altri colori scuri. Il grafico purtroppo non distingue i giocatori di una squadra piuttosto che l'altra. Questo perchè avrebbe richiesto un eccessivo lavoro manuale per distinguere tutte le eccezioni. Vengono di conseguenza visualizzati i protagonisti individuali che hanno effettuato un tiro in porta. In caso di *goal*, viene anche visualizzato chi ha effettuato l'assist. Il grafico può essere utile nell'individuare determinati periodi di gioco in cui questo è stato particolarmente concitato oppure giocatori che hanno influenzato l'andamento della gara, in positivo segnando un goal o in negativo se questo ha tirato un gran numero di volte nel corso della gara senza però concludere a rete.

Il grafico a bolle punta a mostrare in che modo sono stati effettuati i tiri. Mostra sia la parte del corpo coinvolta la quale determina pure il colore delle bolle (piede destro=rosso; sinistro=arancione; testa=blu), sia l'azione da cui si è generato il tiro (open play, penalty, set piece...). Le considerazioni

che possono essere effettuate è che il piede destro è ampiamente utilizzato così come le *open play* che sono di gran lunga superiori a qualsiasi altra situazione di gioco. Ciò ci fa intendere che la partita è stata giocata ad alti ritmi con poche interruzioni di gioco.

Quest'ultimo punto risulta particolarmente evidente se analizziamo il grafico *Action*. Mentre in *Situation* l'origine del tiro si mantiene molto vaga con *open play*, qui viene aggiunto un livello di dettaglio maggiore mostrando indicando in che modo le occasioni da goal vengono create, offrendo indicazioni su quali siano le modalità di attacco più utilizzate. In questo caso il *Pass* è la modalità di attacco più utilizzata il che ci permette di capire che le due squadre hanno preferito una costruzione dal basso e centrale, sfruttando poco le fasce. Questo viene ulteriormente confermato dal numero di colpi di testa del grafico precedente che sono stati veramente pochi.

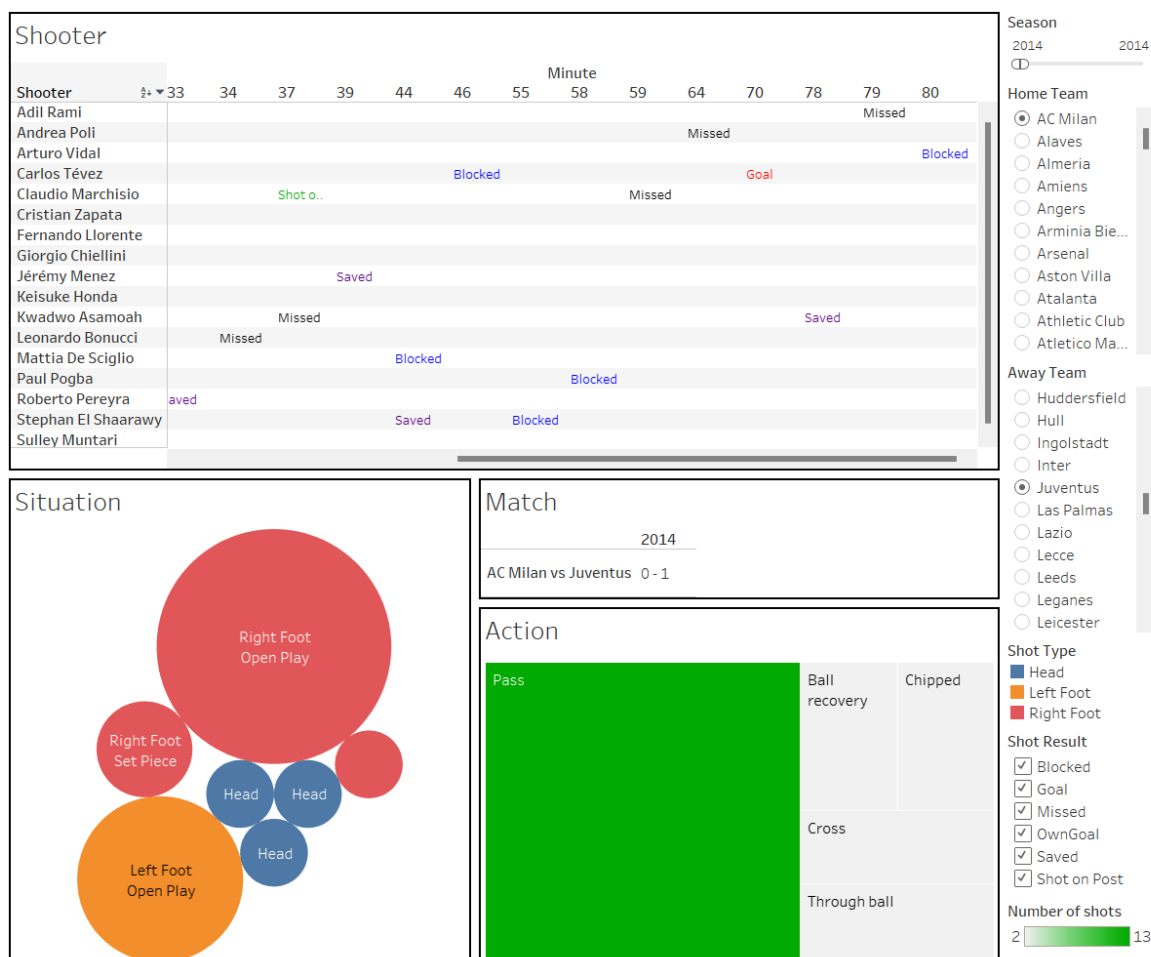


Figura 9: Dashboard 3: "Shooters of the match" - Analisi di una singola partita relativa alle occasioni create e alla loro modalità

Anche in questa dashboard è possibile applicare alcuni filtri per andare più nel dettaglio con le analisi. I filtri che si trovano in tale dashboard riguardano:

- filtro sulla squadra di casa;
- filtro sulla squadra ospite;
- filtro sul risultato del tiro;
- filtro sulle stagioni.

A scopo dimostrativo, nella Figura 10 viene effettuato un filtraggio sul risultato del tiro, selezionando solo *goal*, e per anno andando non solo a considerare una singola partita, ma tutte le volte che quella partita si è giocata.



Figura 10: Dashboard 3 filtrata per squadre e arco temporale

I filtri applicati a tale dashboard ne hanno generato una nuova con più dettagli utili e variegati da analizzare.

Il grafico *Match* assume adesso una rilevanza non mostrando un singolo risultato, ma tutti i risultati di quella gara nei diversi anni. Si può notare che il pareggio in questa gara non viene contemplato da nessuna delle due squadre.

Il grafico *Shooter* adesso mostra 2 ulteriori informazioni. Ovvero quale giocatore impatta maggiormente in quella gara il che è utile saperlo se quel giocatore in una gara futura si trova ancora in quella squadra. L'altra considerazione che si può ricavare è il periodo di gioco in cui si sono concentrati i gol. Viene mostrato che nel periodo tra il 61° e 66° minuto di gioco sono stati realizzati la maggior parte dei goal, ben 7. Mentre i tiri che si concludono a rete sono veramente scarsi nei minuti finali di gioco.

Il grafico a bolle non presenta particolari differenze con la dashboard prima dei filtri, ma possiamo notare che il grafico *Action* è cambiato leggermente. Il passaggio rimane la forma preferita dalle squadre per segnare una rete, ma emergono *cross* e *through ball* il cui significato è che lo stile di gioco delle squadre non è fisso, ma vengono create occasioni da gol anche mediante l'uso delle ali e dei lanci dei difensori.

Analizzando quindi le tendenze passate, si può tentare di prevedere le performance future delle squadre e dei singoli calciatori risultando utile agli appassionati di fantacalcio che devono valutare se schierare o meno un attaccante e agli scommettitori che possono scommettere non solo sul risultato, ma anche sui marcatori, sul minuto di gioco in cui una rete viene segnata e sulla modalità di segnatura di una rete.

3 Qlik

Nella sezione che segue, approfondiremo le analisi condotte utilizzando il secondo dei tre software. Inizieremo con una breve descrizione del processo di caricamento dei dati, evidenziando le difficoltà riscontrate e le soluzioni implementate per superarle. Successivamente, presenteremo un elenco delle analisi svolte e delle relative visualizzazioni, ciascuna accompagnata dalle motivazioni e dalle conclusioni raggiunte.

3.1 Introduzione

Nel panorama delle soluzioni di Business Intelligence, Qlik si afferma come uno dei leader nel settore, offrendo una suite di strumenti software che facilitano la visualizzazione, l'analisi e l'interpretazione dei dati. Fondata nel 1993 in Svezia, Qlik ha rivoluzionato il modo in cui le organizzazioni interagiscono con i dati, spostando l'attenzione dall'elaborazione di report statici alla scoperta dinamica e interattiva delle informazioni.



Figura 11: Logo di Qlik

Qlik propone due prodotti principali, ciascuno rivolto a specifiche esigenze nell'ambito della Data Analytics: *QlikView* e *Qlik Sense*. Entrambi i software si caratterizzano per un approccio distintivo nella gestione e visualizzazione delle informazioni, ma si differenziano notevolmente per il tipo di utenza a cui si rivolgono e per le funzionalità disponibili.

QlikView è progettato per supportare la *"Guided Analytics"*, offrendo alle organizzazioni la possibilità di sviluppare applicazioni analitiche su misura che rispondano alle esigenze specifiche del business. In questo scenario, l'utente finale assume un ruolo di consultazione, senza la possibilità di modificare o ampliare l'applicazione con nuovi elementi analitici. Al contrario, Qlik Sense è concepito come uno strumento di *"Self-service Analytics"*, mirato a democratizzare l'accesso e l'analisi dei dati. Con Qlik Sense, l'utente finale ha la libertà di esplorare autonomamente i dati disponibili, creando report personalizzati e dashboard interattive.

A differenza delle tradizionali piattaforme di Business Intelligence, che si basano su database relazionali e meccanismi di interrogazione tramite query, Qlik Sense si distingue per l'uso di un *Associative Engine*. Questo motore consente a Qlik Sense di superare le limitazioni del linguaggio SQL, originariamente progettato per l'elaborazione di dati statici. Grazie al suo motore associativo, Qlik Sense è in grado di eseguire calcoli e aggregazioni in tempo reale, aggiornando automaticamente le analisi e rivelando le connessioni tra i dati in modo altamente efficiente. Questo approccio non solo facilita una comprensione più profonda delle informazioni, ma rende l'analisi dei dati accessibile e intuitiva per gli utenti finali, permettendo loro di esplorare liberamente le relazioni nascoste nei dati senza dover formulare complesse query SQL.

3.2 Caricamento dati

Nella fase di caricamento dei dati, sono state utilizzate tre tabelle principali: `appearances`, `updated_teamStats` e `updated_games`. Queste tabelle sono state associate tramite la colonna `gameID`, come

consigliato da Qlik. Tuttavia, va notato che l'associazione non è stata ampiamente sfruttata, limitandosi principalmente all'utilizzo del campo season. Le tabelle updated sono state successivamente modificate manualmente, poiché l'interfaccia di Qlik non si è rivelata immediatamente intuitiva. Questo processo di modifica ha richiesto un po' di tempo e attenzione per garantire che i dati fossero correttamente formattati e pronti per l'analisi. Inoltre, la tabella updateds_shots è stata caricata separatamente dalle altre tabelle, mantenendo una propria struttura per facilitare analisi specifiche sui tiri, senza influenzare le associazioni create tra le altre tabelle. Per descrivere al meglio in che modo sono state modificate le tabelle:

- La tabella "appearances" non ha subito modifiche rispetto alla tabella originale del dataset.
- La tabella "updated_games" è una versione modificata della tabella originale "games", in cui sono state aggiunte due nuove colonne: "homeTeamName" e "awayTeamName".
- La tabella "updated_teamStats" è una versione modificata della tabella originale "teamStats", in cui è stata aggiunta una nuova colonna: "teamName".
- La tabella "updates_shots" è una versione modificata della tabella originale "shots", in cui è stata aggiunta una nuova colonna: "shooterName".

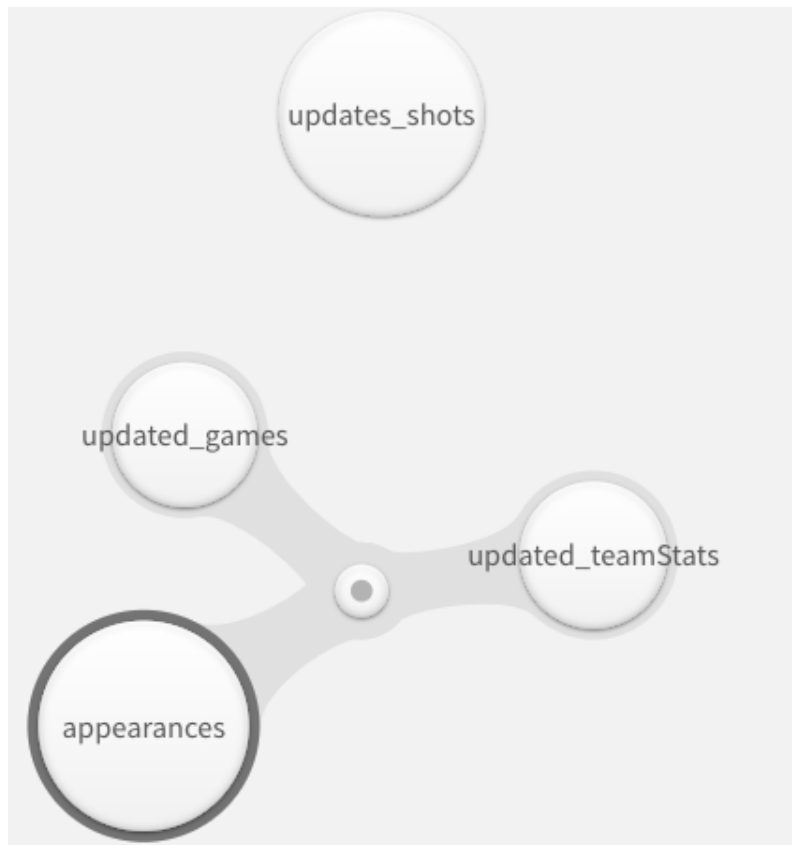


Figura 12: Origine dati Qlik

3.3 Data Analysis

I dati, ora disponibili per l'analisi, ci consentono di procedere alla realizzazione di fogli interattivi mediante l'utilizzo del software Qlik. Nelle sezioni successive, si procederà con l'analisi dettagliata dei vari fogli, con l'obiettivo di esaminare i grafici presenti in ciascuno di essi. Nel dattaglio, sono stati realizzati tre fogli:

- Dashboard 1: Serie A team analysis
- Dashboard 2: Premier League team analysis
- Dashboard 3: Shots analysis

3.3.1 Dashboard 1: Serie A team analysis

Il primo foglio presentato è il risultato della prima analisi condotta utilizzando il software Qlik Sense, con l'obiettivo di familiarizzare con gli strumenti offerti dal tool e ottenere una panoramica generale delle squadre partecipanti al campionato di Serie A. Le visualizzazioni realizzate, come riportato nella Figura 13, mostrano la distribuzione delle prestazioni delle squadre selezionate, confrontate con la Juventus, evidenziando parametri chiave come i punti stagionali, i risultati delle partite, i gol realizzati e attesi (xGoals), nonché il confronto su falli e cartellini. Questo approccio fornisce una visione completa dell'andamento delle squadre nelle varie stagioni. Il confronto con la Juventus è stato scelto perché gli anni analizzati rappresentano un periodo di forte successo per il club torinese, rendendolo un punto di riferimento nel campionato di Serie A.

Procediamo ora con una descrizione dettagliata degli elementi costitutivi del foglio:

- Il grafico a linee, *in alto a sinistra*, mostra l'andamento dei punti stagionali di un team selezionato in confronto alla Juventus.
- Il grafico a torta, *in alto al centro*, rappresenta la distribuzione percentuale dei risultati delle partite, suddivisi in vittorie (W), pareggi (D) e sconfitte (L).
- Il grafico a barre raggruppato, *in basso a sinistra*, confronta i gol effettivi (goals) e i gol attesi (xGoals) tra un team selezionato e Juventus.
- Il grafico a barre in pila, *in basso a destra*, confronta il numero di falli, cartellini gialli e cartellini rossi ricevuti da un team selezionato e dalla Juventus.

I filtri presenti nella dashboard 1 consentono di personalizzare e restringere i dati visualizzati, offrendo una panoramica più mirata sulle squadre e le stagioni selezionate.

Ecco una spiegazione dei filtri visibili nell'immagine:

- Filtro "teamName": Questo filtro permette di selezionare una squadra da analizzare.
- Filtro "season": Consente di selezionare una o più stagioni da analizzare. Questo filtro è utile per osservare l'evoluzione delle prestazioni di una squadra o per confrontare più stagioni tra loro.

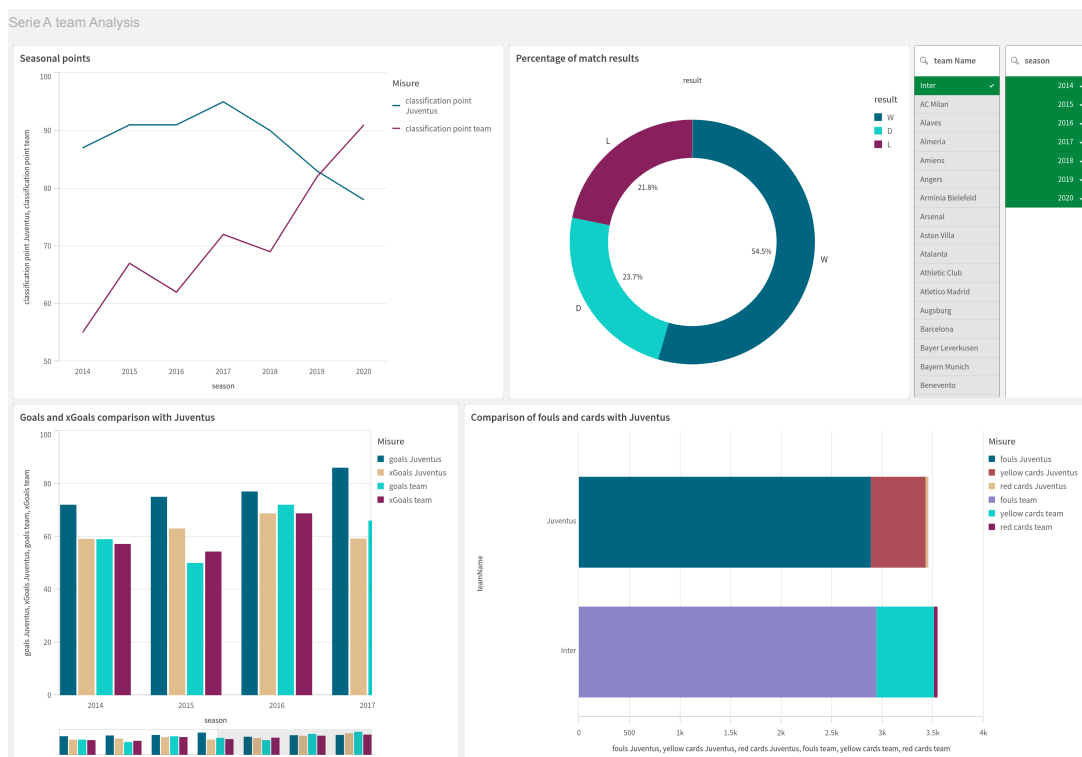


Figura 13: Dashboard 1: Serie A team analysis

Diamo una descrizione dettagliata dei grafici presenti nel foglio di analisi dedicato al confronto tra Inter e Juventus. Nel grafico *Seasonal points*, viene mostrato l'andamento dei punti stagionali ottenuti dall'Inter e dalla Juventus nelle stagioni dal 2014 al 2020. La linea azzurra rappresenta i punti della Juventus, che ha mantenuto un livello elevato nel corso degli anni. Nel 2014, la Juventus ha totalizzato 87 punti, raggiungendo 91 punti nel 2015 e mantenendo questo punteggio anche nel 2016. Nel 2017, ha ottenuto 95 punti, confermando la sua superiorità, mentre nel 2018 ha chiuso con 90 punti. Tuttavia, negli anni successivi, si è registrata una flessione, con 83 punti nel 2019 e 78 punti nel 2020. La linea viola rappresenta l'Inter, che ha mostrato un progresso costante. Nel 2014, l'Inter ha ottenuto 55 punti, ma ha migliorato le sue prestazioni negli anni successivi, totalizzando 67 punti nel 2015, 62 punti nel 2016, e 72 punti nel 2017. Nel 2018, l'Inter ha raggiunto 69 punti, e ha ottenuto 82 punti nel 2019 e 91 punti nel 2020, evidenziando una crescita significativa.

Nel grafico *Percentage of match results*, viene mostrata la distribuzione percentuale dei risultati delle partite dell'Inter dal 2014 al 2020. Si osserva che l'Inter ha vinto il 54.5% delle partite, ha pareggiato il 23.7% e ha perso il 21.8%. Questo grafico fornisce un'ottima visione d'insieme sulla consistenza dei risultati dell'Inter, evidenziando una prevalenza di vittorie rispetto ad altri esiti.

Nel grafico *Goals and xGoals comparison with Juventus*, viene rappresentato il confronto tra i gol effettivi "goals" e i gol attesi "xGoals" per Juventus e Inter dal 2014 al 2020. Nel 2014, la Juventus ha segnato 72 gol con 59.08 xGoals, dimostrando una buona efficienza offensiva. Negli anni successivi, la Juventus ha continuato a migliorare, segnando 75 gol nel 2015 e 77 gol nel

2016, con un picco di 86 gol nel 2017. Nel 2018, ha realizzato 70 gol e ha chiuso con 76 gol nel 2019 e 77 gol nel 2020, mostrando una buona capacità di finalizzazione. L'Inter, nel 2014, ha totalizzato 59 gol con 57.17 xGoals e ha avuto una flessione nel 2015 con 50 gol. Tuttavia, dal 2016 in poi, ha mostrato miglioramenti significativi, segnando 72 gol nel 2016, 66 gol nel 2017, e raggiungendo un massimo di 89 gol nel 2020, a fronte di 78.35 xGoals.

Nel grafico *Comparison of fouls and cards with Juventus*, viene descritto il confronto tra il numero di falli, cartellini gialli e cartellini rossi tra Inter e Juventus dal 2014 al 2020. La Juventus ha commesso circa 2.89k falli, con 541 cartellini gialli e 25 cartellini rossi. L'Inter, invece, ha registrato un numero leggermente superiore di falli, circa 2.94k, con 568 cartellini gialli e 35 cartellini rossi. Questi dati indicano che, sebbene entrambe le squadre abbiano mantenuto una certa disciplina, l'Inter ha mostrato una maggiore aggressività in campo, evidenziata dal numero più alto di falli e cartellini.

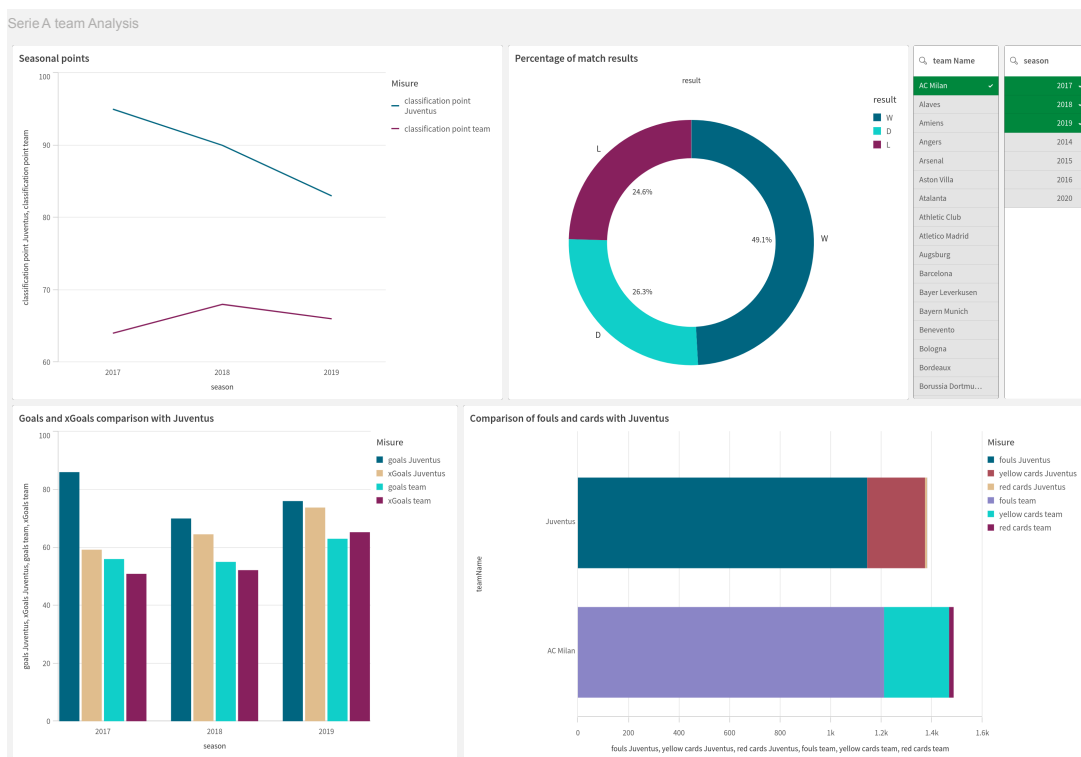


Figura 14: Dashboard 1 filtrata per squadra e stagione

Diamo una descrizione dettagliata dei grafici presenti nel foglio di analisi dedicato al confronto tra Milan e Juventus, come mostrato nella Figura 14. Nel grafico *Seasonal points*, in particolare, viene mostrato l'andamento dei punti stagionali ottenuti dal Milan e dalla Juventus nelle stagioni 2017, 2018 e 2019. La linea azzurra rappresenta i punti della Juventus, che mostra una leggera flessione nel corso degli anni, ma si mantiene costantemente sopra gli 80 punti. La linea viola rappresenta il Milan, che ha un rendimento inferiore, con punti tra i 60 e i 70 in tutte le stagioni considerate. Questo grafico

evidenzia come la Juventus abbia dominato in termini di punti, con una differenza evidente rispetto al Milan.

Nel grafico *Percentage of match results* viene mostrata la distribuzione percentuale dei risultati delle partite del Milan in termini di vittorie, pareggi e sconfitte nelle stagioni 2017, 2018 e 2019. È possibile notare che il Milan ha vinto il 49.1% delle partite, ha pareggiato il 26.3% e ha perso il 24.6%. Questo grafico fornisce una rapida visione d'insieme sulla consistenza dei risultati del Milan in termini di percentuali, evidenziando una leggera prevalenza di vittorie rispetto agli altri esiti.

Nel grafico *Goals and xGoals comparison with Juventus* viene rappresentato il confronto tra i gol effettivi ("goals") e i gol attesi ("xGoals") delle squadre Milan e Juventus per le stagioni 2017, 2018 e 2019. La Juventus ha segnato e ha avuto un numero di gol attesi più alto rispetto al Milan, dimostrando una maggiore capacità offensiva. La Juventus ha mantenuto un valore alto sia nei gol effettivi che nei gol attesi, mentre il Milan ha avuto prestazioni inferiori su entrambi i fronti. Questo confronto suggerisce che la Juventus sia stata più efficiente sia in termini di realizzazione che nelle occasioni create. Si osserva un dato interessante per il Milan nella stagione 2019, infatti, in questa stagione il Milan ha avuto un numero di xGoals superiore ai gol effettivamente realizzati, il che indica che, secondo le aspettative basate sulle occasioni create, il Milan avrebbe dovuto segnare più gol di quelli effettivamente realizzati. In contrasto, la Juventus nella stagione 2017 ha segnato molti più gol rispetto agli xGoals previsti, indicando un'efficacia offensiva superiore alle aspettative. Questo significa che, nonostante il numero di opportunità create (xGoals) fosse in linea con le previsioni, la Juventus è riuscita a concretizzare molto meglio di quanto indicato dalle statistiche.

Nel grafico *Comparison of fouls and cards with Juventus*, viene descritto il confronto tra il numero di falli, cartellini gialli e cartellini rossi tra Milan e Juventus nelle stagioni 2017, 2018 e 2019.

Il Milan ha commesso un numero maggiore di falli rispetto alla Juventus, 1210 contro 1140 ma entrambe le squadre hanno un numero simile di cartellini gialli (257 - 229). Per quanto riguarda i cartellini rossi, entrambe le squadre hanno ricevuto un numero relativamente basso, 8 per la Juventus e 18 per il Milan, indicando che sono state abbastanza disciplinate nelle stagioni analizzate.

Tra le due dashboard possiamo individuare alcune differenze:

- Andamento dei punti: L'Inter ha mostrato un miglioramento più evidente nel confronto con la Juventus, mentre il Milan ha avuto punteggi più stabili, ma inferiori.
- Disciplina: L'Inter ha mostrato maggiore aggressività rispetto alla Juventus, mentre il Milan ha registrato un numero maggiore di falli e cartellini rossi, evidenziando problemi di disciplina.

Questa dashboard fornisce una visione chiara delle performance delle squadre, mettendo in evidenza parametri come punti stagionali, gol, e disciplina. È uno strumento utile per monitorare l'andamento nel tempo, individuare aree di miglioramento e confrontare squadre in modo rapido e visivo. Ideale per tecnici e dirigenti, permette di prendere decisioni strategiche basate su dati concreti.

3.3.2 Dashboard 2: Premier League team analysis

Il foglio presentato è il risultato di un'analisi condotta con Qlik Sense per esplorare le prestazioni di una squadra della Premier League, mettendo a confronto i gol fatti in casa e in trasferta e le medie delle quote di scommesse Bet365 con quelle del Leicester. Le visualizzazioni mostrate, come indicato nella Figura 15, evidenziano il confronto tra la squadra selezionata e il Leicester in termini di gol segnati in casa e in trasferta, oltre alle quote di scommesse per le partite giocate in entrambe le situazioni. Inoltre, il foglio include l'analisi dei gol concessi in casa e in trasferta, concentrandosi solo sulla squadra

selezionata della Premier League. Questo approccio permette di ottenere una visione completa delle prestazioni offensive e difensive della squadra nel contesto della Premier League, integrata con una valutazione delle aspettative di scommesse, offrendo così una panoramica chiara su fattori chiave del rendimento in diversi contesti di gara. La scelta di confrontare la squadra selezionata con il Leicester nasce dal fatto che il Leicester ha vinto la Premier League nella stagione 2015-2016, un risultato straordinario per una squadra considerata "piccola" rispetto ai club più blasonati. Questo confronto consente di valutare le performance della squadra selezionata rispetto a una realtà che ha saputo raggiungere il successo contro avversari più forti.

Procediamo ora con una descrizione dettagliata degli elementi costitutivi del foglio:

- Il grafico a barre raggruppato, *in alto a sinistra*, confronta i goal segnati in casa e in trasferta tra un team selezionato e Leicester.
- Il grafico a dispersione, *in alto a destra*, rappresenta le prestazioni difensive di una squadra per ogni stagione, con i gol concessi in casa sull'asse X e i gol concessi in trasferta sull'asse Y. Ogni punto nel grafico corrisponde a una stagione specifica.
- Il grafico lineare, *in basso a destra*, mostra il confronto tra le quotazioni medie di probabilità di vittoria in casa e in trasferta del team selezionato e del Leicester.

Nel Foglio 2 sono presenti due filtri e l'obiettivo è di selezionare la stessa squadra in entrambi i filtri affinché i grafici forniscano un'analisi coerente. Per consentire la selezione dello stesso team in entrambi i filtri, sono stati utilizzati due stati alternativi: `stateHome` e `stateAway`. I filtri sono i seguenti:

- Filtro "home team name": Consente di selezionare la squadra che gioca in casa.
- Filtro "away team name" : Consente di selezionare la squadra che gioca in trasferta.

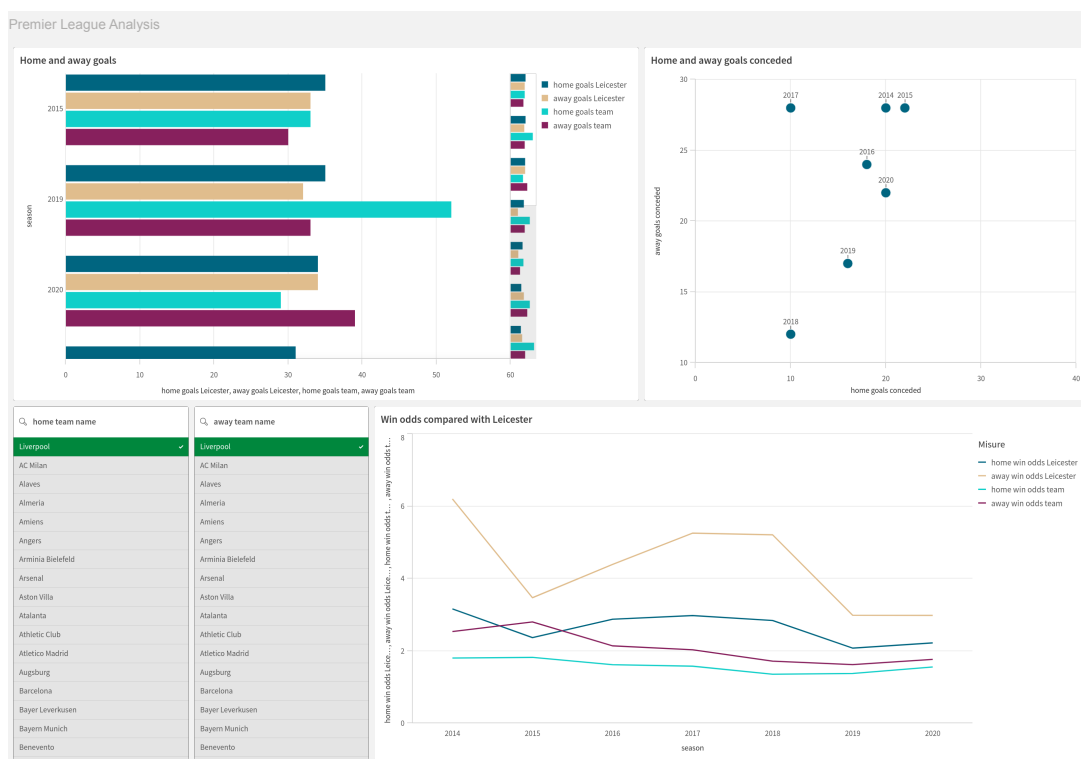


Figura 15: Dashboard 2: Premier League team analysis

Diamo una descrizione dettagliata dei grafici presenti nel foglio di analisi dedicato al confronto tra Liverpool e Leicester:

Nel grafico *Home and Away Goals*, viene mostrato l'andamento dei gol segnati in casa e in trasferta dal Liverpool e dal Leicester nelle stagioni dal 2014 al 2020.

La linea azzurra rappresenta i gol realizzati in casa dal Liverpool, che ha mostrato un trend in crescita, passando da 30 gol nel 2014 a un massimo di 55 gol nel 2018. Tuttavia, nel 2015, il Leicester ha segnato più gol in casa rispetto al Liverpool, totalizzando 35 gol, un chiaro segno del loro rendimento offensivo durante la stagione in cui hanno vinto il campionato.

Anche i gol segnati in trasferta dal Liverpool, rappresentati dalla linea bordeaux, mostrano una progressione, passando da 22 gol nel 2014 a 39 gol nel 2020. I gol realizzati in trasferta dal Leicester, indicati dalla linea beige, sono stati 18 nel 2014 e hanno raggiunto il massimo di 34 nel 2020, evidenziando un andamento in crescita.

Nel grafico *Home and Away Goals Conceded*, vengono analizzati i gol subiti in casa e in trasferta dal Liverpool. I gol concessi in casa dal Liverpool, mostrano un calo significativo nel 2017/2018, con soli 10 gol subiti, e si attestano a 20 gol nel 2020. Per quanto riguarda i gol concessi in trasferta, il Liverpool ha subito un massimo di 28 gol nel 2014/2015, con una leggera flessione nel 2018, dove ha concesso solo 12 gol.

Il grafico *Win Odds Compared with Leicester* mette a confronto le quotazioni medie per la vittoria in casa e in trasferta tra Liverpool e Leicester. Le quotazioni di vittoria in casa per il Leicester, indicate dalla linea blu, mostrano un picco di 3.16 nel 2014 e scendono a 2.08 nel 2019, mentre le quotazioni per la vittoria in trasferta oscillano, partendo da 6.20 nel 2014 e raggiungendo 2.99 nel 2020. Le quotazioni

di vittoria in casa per il Liverpool, rappresentate dalla linea azzurra, rimangono costantemente basse, partendo da 1.80 nel 2014 e scendendo fino a 1.35 nel 2018. Le quote per le vittorie in trasferta, indicate dalla linea bordeaux, mostrano una leggera flessione, partendo da 2.53 nel 2014 e scendendo a 1.62 nel 2019. È interessante notare che con il passare delle stagioni, le quote sono in calo indicavando un miglioramento delle performance delle due squadre: il Liverpool si contendeva frequentemente il titolo, mentre il Leicester, pur essendo una squadra di media-alta classifica, ha dimostrato di poter competere con i top club. Questi grafici offrono una visione chiara delle prestazioni di Liverpool e Leicester, evidenziando i punti di forza e le debolezze di entrambe le squadre nel corso degli anni, con particolare attenzione alla stagione in cui il Leicester ha vinto il campionato nel 2015 e al successo del Liverpool nel 2019.

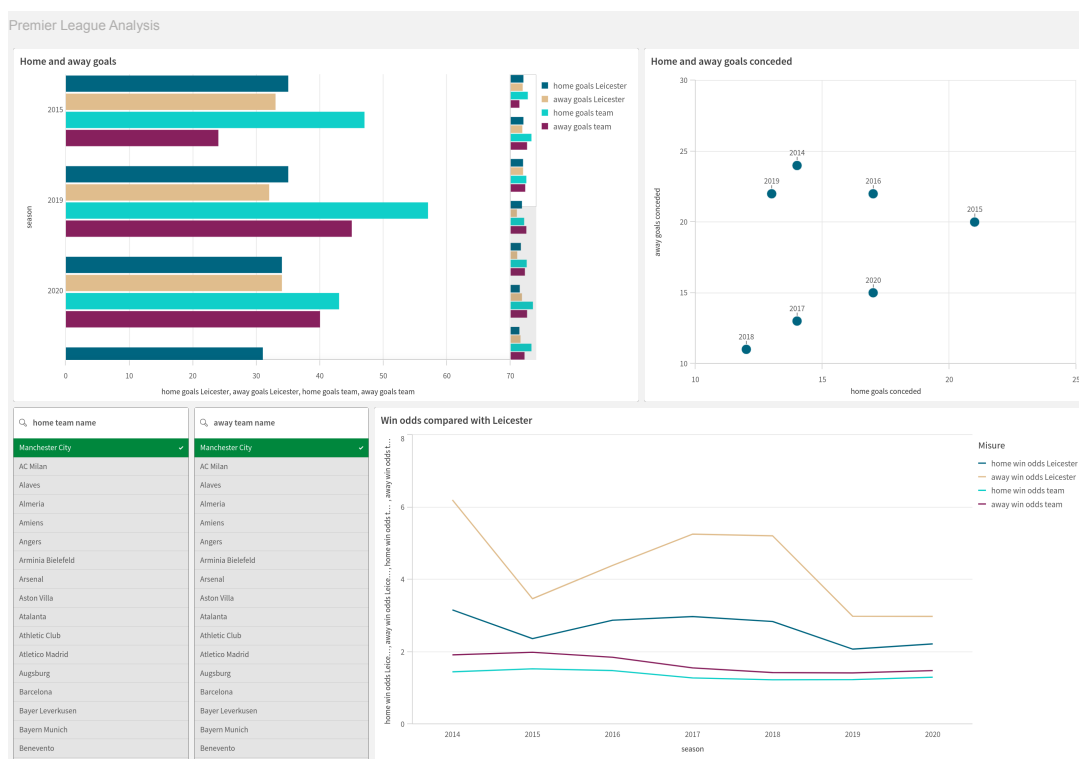


Figura 16: Dashboard 2 filtrata per le squadre

Diamo una descrizione dettagliata dei grafici presenti nel foglio di analisi dedicato al confronto tra Manchester City e Leicester:

Come mostrato nella Figura 16, il grafico *Home and Away Goals* mostra l'andamento dei gol segnati in casa e in trasferta dal Manchester City e dal Leicester nelle stagioni dal 2014 al 2020. La linea azzurra rappresenta i gol realizzati in casa dal Manchester City, che ha mostrato un andamento top, passando da 44 gol nel 2014 a 47 gol nel 2015, fino a un picco di 61 gol nel 2017. Nel 2018, il Manchester City ha totalizzato 57 gol, mantenendo lo stesso punteggio nel 2019, per poi scendere a 43 gol nel 2020. Questo trend crescente evidenzia l'elevata capacità offensiva del Manchester City durante gli anni di massimo rendimento. È importante notare che, nel 2015, nonostante il Manchester City avesse totalizzato un numero complessivo di gol (casa e trasferta) superiore a quello del Leicester, non è riuscito a vincere il

campionato.

Anche i gol segnati in trasferta dal Manchester City, rappresentati dalla linea bordeaux, mostrano una certa variabilità. Il Manchester City ha iniziato con 39 gol nel 2014 e ha registrato un massimo di 45 gol nel 2017, chiudendo a 40 gol nel 2020.

Nel grafico *Home and Away Goals Conceded*, vengono analizzati i gol subiti in casa e in trasferta dal Manchester City. I gol concessi in casa dal Chelsea, rappresentati dalla linea viola, mostrano un calo significativo nel 2018, con soli 12 gol subiti, e si attestano a 17 gol nel 2020. Per quanto riguarda i gol concessi in trasferta, indicati dalla linea gialla, il Manchester City ha subito un massimo di 24 gol nel 2014, con una leggera flessione nel 2018, dove ha concesso solo 11 gol.

Il grafico *Win Odds Compared with Leicester* mette a confronto le quotazioni medie per la vittoria in casa e in trasferta tra Manchester City e Leicester. Le quotazioni di vittoria in casa per il Manchester City, rappresentate dalla linea azzurra, rimangono costantemente basse, partendo da 1.45 nel 2014 e scendendo fino a 1.23 nel 2018. Le quote per le vittorie in trasferta, indicate dalla linea blu scuro, mostrano una leggera flessione, partendo da 1.92 nel 2014 e scendendo a 1.42 nel 2019.

È interessante osservare che le quotazioni di vittoria del Manchester City sono diminuite negli anni 2017 e 2018, durante i quali ha conquistato il campionato. Anche nel 2020, quando il Manchester City ha vinto nuovamente, le quotazioni erano leggermente più alte rispetto agli anni precedenti.

Tra le due dashboard possiamo individuare alcune differenze. In generale, la dashboard del Manchester City evidenzia un andamento offensivo e una capacità difensiva notevole, con un chiaro picco nelle prestazioni nelle stagioni 2017/2018. La dashboard del Liverpool mostra una crescita costante nei gol realizzati, ma una maggiore vulnerabilità in trasferta. Il Leicester, che ha vinto il campionato nel 2015, funge da termine di confronto per entrambe le squadre, evidenziando l'importanza di performance elevate anche da parte delle cosiddette "piccole" realtà nel campionato di Premier League.

Queste dashboard offre un'opportunità per monitorare l'andamento delle prestazioni delle squadre, evidenziando le dinamiche di gioco nel tempo e consentendo confronti significativi tra le squadre. Fornisce dati concreti per prendere decisioni strategiche, analizzare le quotazioni di scommesse e presentare informazioni in modo chiaro a stakeholder e analisti, rendendola versatile e potente per l'analisi delle performance sportive.

3.3.3 Dashboard 3: Shots Analysis

Il foglio presentato si concentra sull'analisi dei tiri effettuati dai giocatori durante il campionato. Come mostrato nella Figura 17, la dashboard rappresenta graficamente vari aspetti legati ai tiri, tra cui il numero e i vari tipi di tiri per minuto di gioco, la tipologia dei tiri (distinti in colpi di testa, tiri con il piede destro, sinistro e con altre parti del corpo) e i risultati dei tiri in base alla situazione di gioco (calcio di rigore, punizione diretta, da corner o azione in movimento).

Procediamo ora con una descrizione dettagliata degli elementi costitutivi del foglio:

- Il grafico lineare, *in alto a sinistra*, mostra l'andamento del numero di tiri suddivisi per tipologia (colpi di testa, tiri con il piede destro, sinistro e altre parti del corpo) durante i 90 minuti di gioco.
- Il grafico a torta, *in alto a destra*, mostra la distribuzione percentuale delle diverse tipologie di tiro, evidenziando la frequenza con cui vengono utilizzati i vari tipi di conclusioni, come tiri con il piede destro, sinistro, colpi di testa e altre parti del corpo.

- Il grafico a barre raggruppato, *in basso a sinistra*, rappresenta i risultati dei tiri in base alla situazione di gioco da cui sono originati.

Nella dashboard 3 è presente un filtro, che è il seguente:

- Il filtro "shooter name": permette di selezionare il giocatore specifico per visualizzare i dettagli relativi ai suoi tiri.

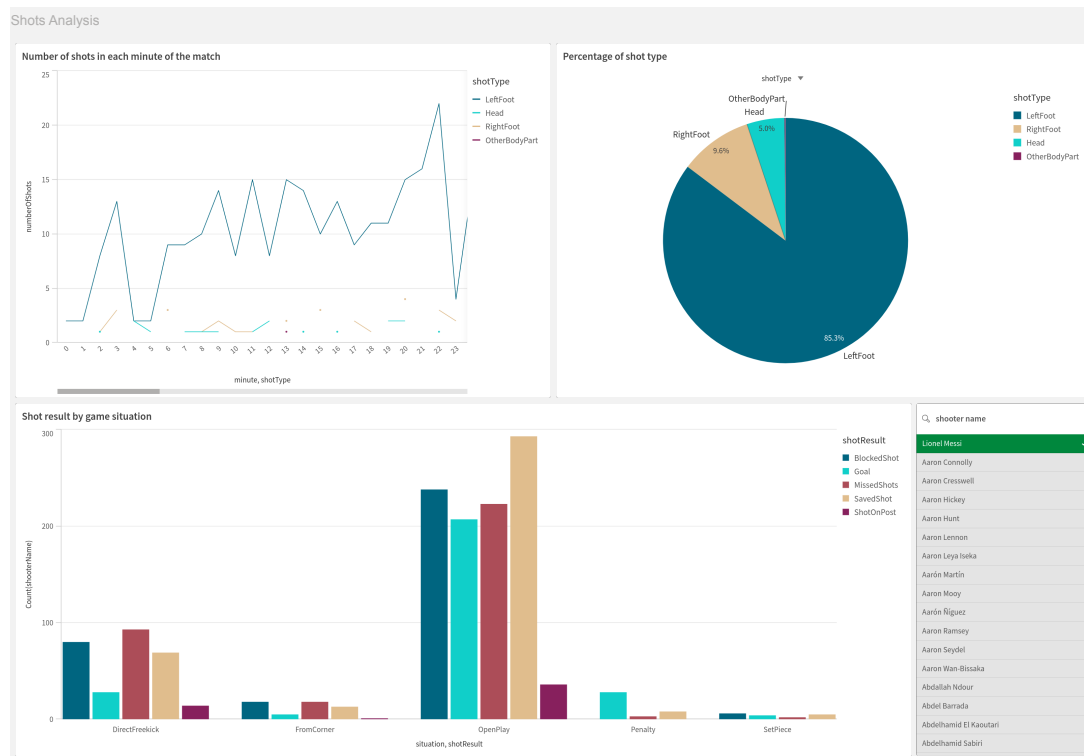


Figura 17: Dashboard 3: Shots Analysis

Diamo una descrizione dettagliata dei grafici presenti nel foglio di analisi dedicato a Lionel Messi:

Nel grafico *Number of shots in each minute of the match* viene mostrato il numero di tiri tentati durante i primi 96 minuti di gioco, suddivisi per tipo di tiro (piede sinistro, piede destro, testa e altre parti del corpo). Come prevedibile per un giocatore come Lionel Messi, che è notoriamente mancino, la maggior parte dei tiri è stata effettuata con il piede sinistro. I picchi più alti si registrano nei minuti 22°, 29°, 36° e 50°, con una media compresa tra i 15 e i 24 tiri in questi intervalli, dimostrando quanto Messi preferisca il piede sinistro durante il gioco. I tiri con il piede destro, pur meno frequenti, mostrano un andamento costante, con alcune crescite isolate, come al 50° minuto. Messi utilizza raramente la testa o altre parti del corpo per i suoi tiri, riflettendo il suo stile di gioco più incentrato sulla tecnica e il dribbling piuttosto che sul gioco aereo.

Il grafico *Percentage of shot type* sottolinea ulteriormente la predominanza del piede sinistro nei tiri di Messi: ben l'85.3% dei suoi tiri provengono da questo piede, evidenziando la sua abilità straordinaria con il sinistro. Solo il 9.6% dei tiri è stato effettuato con il piede destro, dimostrando che, pur essendo

meno frequente, Messi utilizza occasionalmente il destro in situazioni particolari. I colpi di testa rappresentano solo il 5.0%, mentre i tiri con altre parti del corpo sono praticamente inesistenti (0.1%), un aspetto che riflette anche la sua statura e il suo stile di gioco, che non si basa sulla statura fisica ma sulla rapidità e la precisione.

Il grafico *Shot result by game situation* mostra i risultati dei tiri (bloccati, goal, tiri sbagliati, tiri salvati, colpi sul palo) in relazione alla situazione di gioco. Quando si analizzano i tiri in azioni di gioco aperto, Messi dimostra la sua grande efficacia: 207 goal e 238 tiri bloccati, a testimonianza della sua capacità di superare le difese avversarie in spazi stretti. Le punizioni dirette, un altro marchio di fabbrica di Messi, vedono 28 goal segnati su 80 tiri bloccati e 93 tiri sbagliati, dimostrando sia la frequenza che l'efficacia di questa sua abilità. In situazioni di rigore, Messi ha segnato 28 goal, a testimonianza della sua freddezza sotto pressione. I calci d'angolo e le situazioni di calcio piazzato contribuiscono a un numero ridotto di goal, riflettendo un contesto in cui Messi tende più a servire assist piuttosto che cercare la finalizzazione diretta.

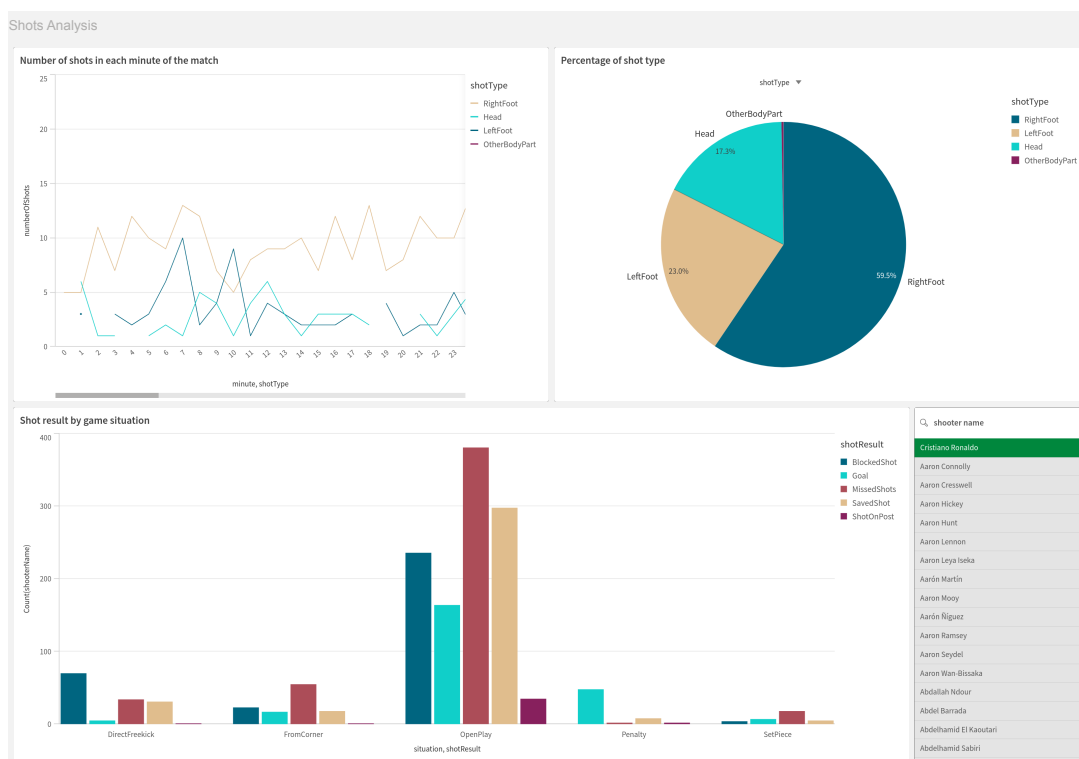


Figura 18: Dashboard 3 filtrata per calciatore

Diamo una descrizione dettagliata dei grafici presenti nel foglio di analisi dedicato a Cristiano Ronaldo:

Come mostrato nella Figura 18, il grafico *Number of shots in each minute of the match* mostra il numero di tiri tentati durante i primi 96 minuti di gioco, suddivisi per tipo di tiro (piede destro, piede sinistro, testa e altre parti del corpo). Come prevedibile per Cristiano Ronaldo, che è noto per la sua potenza e versatilità, la maggior parte dei tiri è stata effettuata con il piede destro. Picchi significativi di tiri con il piede destro si registrano nei minuti 30°, 46° e 50°, con un numero di tentativi

che oscilla tra 10 e 22. Nonostante sia ambidestro, i tiri con il piede sinistro sono meno frequenti, ma comunque utilizzati in modo consistente. Ronaldo, celebre per le sue abilità nel gioco aereo, ha una presenza importante nei colpi di testa, con un numero costante di tiri di testa nei vari minuti di gioco, in particolare nei minuti 40° e 88°.

Il grafico *Percentage of shot type* evidenzia la distribuzione percentuale delle diverse tipologie di tiro. La predominanza del piede destro è evidente, con il 59.5% dei tiri provenienti da questo piede, confermando l'abilità di Ronaldo nel suo piede preferito. Tuttavia, la sua capacità di usare il piede sinistro si riflette nel 23.0% dei tiri effettuati con quest'ultimo. I colpi di testa, una delle caratteristiche distintive di Ronaldo, costituiscono una parte importante del totale con il 17.3%, a dimostrazione della sua eccellenza nel gioco aereo. I tiri con altre parti del corpo sono praticamente inesistenti, rappresentando solo lo 0.3%.

Il grafico *Shot result by game situation* mostra i risultati dei tiri (bloccati, goal, tiri sbagliati, tiri salvati, colpi sul palo) in relazione alla situazione di gioco. Ronaldo dimostra una grande capacità di segnare in varie situazioni, con una predominanza di goal nelle azioni di gioco aperto: 164 goal segnati, accompagnati da 236 tiri bloccati e 381 tiri sbagliati. Nei calci di rigore, Ronaldo ha segnato 48 goal, evidenziando la sua affidabilità in momenti cruciali. Le sue punizioni dirette, altro punto forte del suo repertorio, contano 5 goal su 70 tiri bloccati. Ronaldo è inoltre particolarmente pericoloso sui calci d'angolo, con 17 goal segnati da corner, riflettendo la sua abilità di finalizzare situazioni di gioco aereo.

Tra le due dashboard possiamo individuare alcune differenze:

- Messi è molto più focalizzato sull'uso del piede sinistro, mentre Ronaldo dimostra una maggiore varietà nelle tipologie di tiro, alternando piede destro, sinistro e testa.
- Nel gioco aereo, Ronaldo si distingue nettamente per la sua superiorità, come dimostrato dalla percentuale più alta di colpi di testa, mentre Messi tende a evitare questo tipo di conclusioni, complice anche la sua statura fisica.
- Entrambi eccellono nelle punizioni e nei rigori, ma Messi registra una maggior frequenza di goal su punizione diretta, mentre Ronaldo risulta più prolifico nei rigori.
- Ronaldo si dimostra più versatile nelle situazioni di calcio d'angolo e nel gioco aereo, mentre Messi predilige azioni di gioco aperto e punizioni dirette per sfruttare le sue abilità.

In linea generale, questa dashboard presenta vantaggi significativi, consentendo di visualizzare in modo chiaro le tipologie di tiri, le situazioni di gioco e i risultati associati, facilitando l'identificazione di tendenze. Si rivela utile per confrontare le prestazioni di diversi giocatori e offre agli allenatori e agli analisti l'opportunità di preparare al meglio le partite contro avversari specifici, ottimizzando le strategie in base a dati concreti.

4 Power BI

Di seguito vengono presentate le analisi effettuate utilizzando il terzo e ultimo software considerato. In primo luogo, viene fornita una breve descrizione della fase di caricamento dei dati, mettendo in evidenza le problematiche emerse e le soluzioni adottate per affrontarle. Successivamente, si propone un elenco delle analisi eseguite, accompagnate dalle relative rappresentazioni grafiche nelle dashboard, ciascuna delle quali è corredata di spiegazioni e delle conclusioni raggiunte.

4.1 Introduzione

Power BI è una piattaforma di business intelligence sviluppata da Microsoft, che offre una vasta gamma di strumenti per analizzare e visualizzare i dati. Uno dei suoi punti di forza principali è la perfetta integrazione con altre applicazioni Microsoft, come Excel e Azure, facilitando così l'uso per chi già sfrutta l'ecosistema Microsoft.



Figura 19: Logo di Power BI

Power BI utilizza un modello di dati colonnare altamente compresso, che consente prestazioni elevate nella gestione di grandi volumi di dati. La piattaforma mette a disposizione strumenti intuitivi per la trasformazione dei dati e la creazione di misure personalizzate.

In particolare, per il processo di *ETL*, Power BI si distingue rispetto agli altri software grazie a un ambiente completo e ricco di funzionalità per la pulizia dei dati. Il suo editor di query, *Power Query*, offre un'interfaccia grafica che consente di generare script M per modificare i dati prima di importarli nel modello.

Anche nella creazione di misure, Power BI offre vantaggi rispetto agli altri strumenti grazie al linguaggio *DAX*, che include oltre 250 funzioni per definire formule avanzate.

4.2 Caricamento dati

La prima operazione effettuata, come negli altri software, è il caricamento dei dati dalla sorgente. In Power BI, questo processo è notevolmente semplificato grazie a un'interfaccia intuitiva che consente di importare facilmente i file .csv. Una volta caricati, le tabelle vengono collegate automaticamente da Power BI, rendendo immediato il lavoro con i dati. Tuttavia è stato necessario applicare delle modifiche al dataset originale per poter eseguire delle operazioni di analisi. Queste modifiche consistono sia nell'aggiunta di nuove colonne personalizzate sia la creazione di nuove tabelle.

Ecco una spiegazione dettagliata delle principali tabelle:

- La tabella "appearances" non ha subito modifiche rispetto alla tabella originale del dataset.
- La tabella "players" non ha subito modifiche rispetto alla tabella originale del dataset.
- La tabella "shots" non ha subito modifiche rispetto alla tabella originale del dataset.

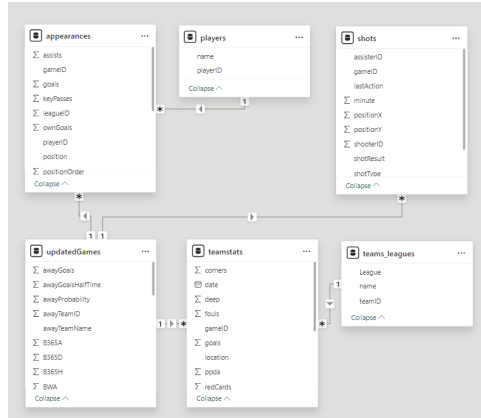


Figura 20: Origine dati Power BI

- La tabella "teamstats" non ha subito modifiche rispetto alla tabella originale del dataset.
- La tabella "updatedGames" è una versione modificata della tabella originale "games", in cui sono state aggiunte nuove colonne: "homeTeamName" e "awayTeamName", "averageBookmakersHome", "averageBookmakersDraw", "averageBookmakersAway", "maxBetHomeWin", "maxBetDraw", "maxBetAwayWin".
- La tabella "teamsleagues" è una nuova tabella creata per unire i nomi delle squadre, la lega a cui appartengono e l'id della squadra. .

4.3 Data Analysis

I dati, ora pronti per l'analisi, permettono di creare report interattivi utilizzando il software Power BI. Nelle prossime sezioni verrà condotta un'analisi approfondita dei vari report, con l'obiettivo di esaminare i grafici presenti in ciascuno di essi. In particolare, sono stati sviluppati tre report distinti:

- Dashboard 1: Analysis of goals, assists, shots and cards of players
- Dashboard 2: Analysis of goals, xGoals, shot and shot on target of teams per season
- Dashboard 3: Analysis of bookmakers bets for a match

4.3.1 Dashboard 1: Analysis of goals, assists, shots and cards of players

Questa dashboard, mostrata nella Figura 21, offre una visione esaustiva delle prestazioni individuali dei giocatori, ponendo l'attenzione su metriche fondamentali come goal, assist, tiri effettuati e comportamenti disciplinari in campo. Grazie a una combinazione di visualizzazioni intuitive e dati aggregati, il foglio consente di analizzare in dettaglio il contributo di ciascun giocatore, evidenziando non solo le capacità realizzative e di supporto, ma anche la precisione e l'efficienza sotto porta. Inoltre, include un'analisi dei comportamenti disciplinari, come il numero di cartellini ricevuti, che offre spunti interessanti per valutare l'impatto dei giocatori dal punto di vista tattico e della correttezza. La presenza di funzionalità interattive, come filtri per stagioni o giocatori specifici, consente di personalizzare l'analisi, rendendo il foglio uno strumento estremamente versatile e dinamico. Questo approccio integrato permette non solo di identificare i migliori marcatori e assistman, ma anche di approfondire aspetti critici come l'aggressività o l'efficienza nei tiri, offrendo un quadro completo delle performance in campo.

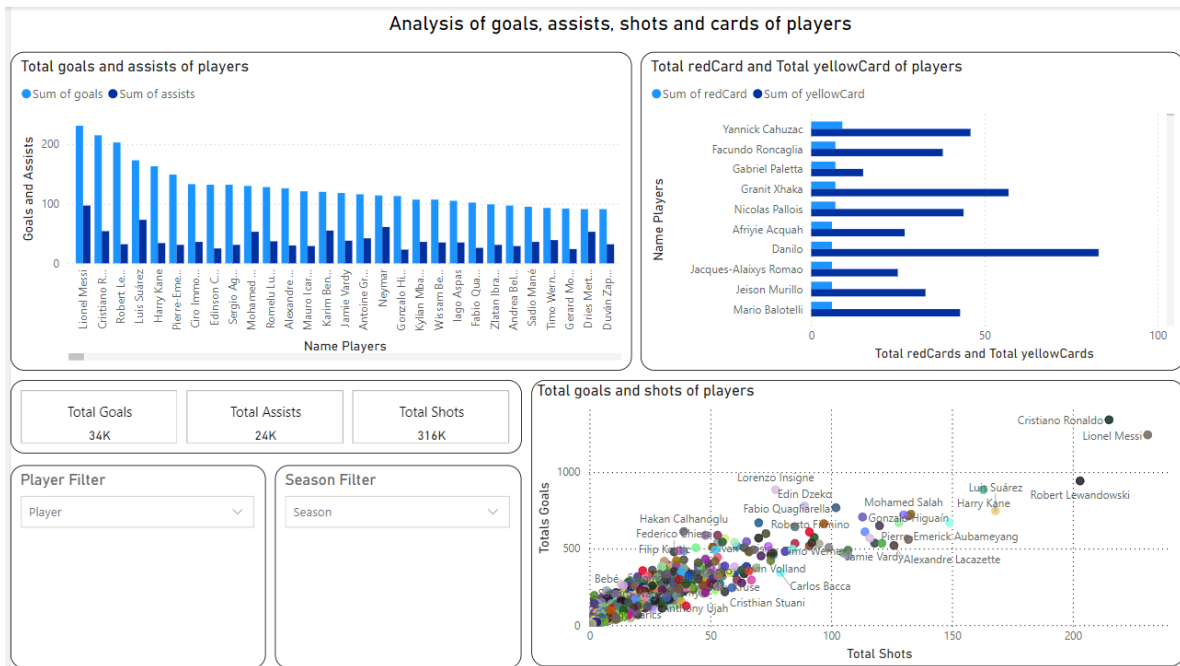


Figura 21: Dashboard 1: Analysis of goals, assists, shots and cards of players

Procediamo ora con una descrizione dettagliata degli elementi costitutivi del foglio:

- La sezione *in alto a sinistra* è dedicata all'analisi dei goal e degli assist realizzati dai giocatori, rappresentati tramite un grafico a colonne.
- La sezione *in alto a destra* della dashboard si concentra sull'analisi dei cartellini gialli e rossi ricevuti dai giocatori, rappresentati attraverso un grafico a righe.
- La sezione *in basso a sinistra* presenta un riepilogo sintetico delle metriche aggregate relative ai goal, assist e tiri totali.
- La sezione *in basso a destra* della dashboard è dedicata all'analisi dei goal e dei tiri totali effettuati dai giocatori, rappresentata attraverso un grafico a dispersione.

La dashboard in alto a sinistra fornisce una panoramica chiara e dettagliata delle performance individuali, mettendo in evidenza i contributi dei giocatori più rilevanti. Ogni barra nel grafico rappresenta un giocatore e include due metriche fondamentali, somma dei goal, rappresentata dalla colonna blu più scura, indica il numero totale di reti realizzate da ciascun giocatore e somma degli assist, rappresentata dalla colonna blu più chiara, evidenzia il numero di passaggi decisivi effettuati. Questa visualizzazione consente di confrontare le prestazioni dei vari giocatori e di identificare rapidamente quelli con un impatto maggiore sia nella finalizzazione che nella costruzione del gioco. L'obiettivo di questa analisi è evidenziare i giocatori che non solo segnano con regolarità, ma contribuiscono anche alla creazione di opportunità per i loro compagni di squadra, fornendo così un quadro completo del loro impatto sul campo. La dashboard in alto a destra permette di identificare i comportamenti disciplinari dei giocatori durante le partite, evidenziando i più frequenti destinatari di sanzioni. Ogni barra del grafico rappresenta un giocatore e include due misure principali, la somma dei cartellini gialli, rappresentata dalla colonna blu più chiara, mostra il totale dei cartellini gialli ricevuti e la somma dei cartellini rossi,

rappresentata dalla colonna blu più scura, indica il numero totale di espulsioni dirette. L'obiettivo di questa sezione è analizzare il comportamento dei giocatori, fornendo informazioni utili per identificare chi potrebbe rappresentare un rischio per la squadra a causa di sanzioni frequenti. Questo tipo di analisi può essere particolarmente rilevante per decisioni tattiche o per la gestione del comportamento dei giocatori. La dashboard in basso a sinistra è costituita da tre indicatori numerici che offrono una panoramica immediata delle performance complessive dei giocatori. Gli indicatori principali sono *Total Goals* il quale indica il numero totale di goal realizzati da tutti i giocatori analizzati, *Total Assists* il quale mostra il numero totale di assist effettuati e *Total Shots* il quale rappresenta il totale dei tiri eseguiti. I valori complessivi consentono di valutare rapidamente l'entità dei dati analizzati e di mettere in relazione le diverse metriche. Il rapporto tra goal e assist evidenzia l'importanza delle collaborazioni tra i giocatori. Il confronto tra goal e tiri totali fornisce un'indicazione generale sull'efficienza realizzativa complessiva. La sezione serve come punto di partenza per approfondimenti successivi, offrendo agli utenti una sintesi chiara e di facile interpretazione. Questi numeri aggregati forniscono una base per identificare trend generali e guidare analisi più dettagliate nelle altre sezioni della dashboard. La dashboard in basso a destra offre una visione dettagliata della relazione tra il numero di tiri e il numero di goal, permettendo di valutare l'efficienza realizzativa dei giocatori. L'asse delle X *Total Shots* rappresenta il numero totale di tiri effettuati da ciascun giocatore mentre l'asse delle Y *Total Goals* indica il numero totale di goal realizzati. Questo grafico fornisce informazioni utili per confrontare i giocatori non solo in base alla loro capacità di segnare, ma anche in relazione alla quantità di tentativi necessari per raggiungere quei risultati. È uno strumento chiave per individuare i giocatori più affidabili in fase realizzativa e quelli che potrebbero necessitare di migliorare la propria precisione.

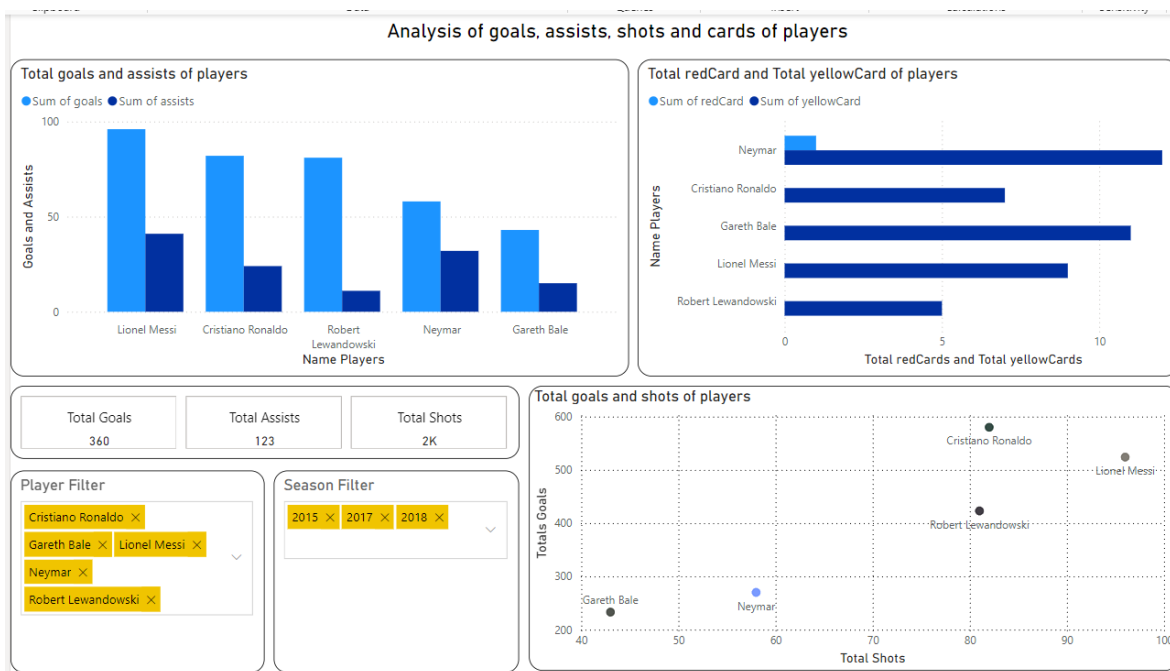


Figura 22: Dashboard 1 filtrata per giocatori e stagione

Oltre a ciò è possibile applicare dei filtri:

- Player filter: scegliere uno o più giocatori.
- Season filter: scegliere una o più stagioni.

La dashboard, come mostrato nella Figura 22, è stata personalizzata applicando due filtri principali: il filtro giocatore e il filtro stagione al fine di concentrare l'analisi su un sottoinsieme specifico di dati. Il filtro giocatore include i seguenti atleti: Cristiano Ronaldo, Lionel Messi, Robert Lewandowski, Neymar e Gareth Bale, mentre il filtro stagione limita l'analisi agli anni 2015, 2017 e 2018. Grazie a questa selezione, il foglio consente di osservare le prestazioni di questi giocatori d'élite in un periodo mirato, escludendo dati irrilevanti o non pertinenti. I filtri applicati hanno un impatto significativo sui risultati visualizzati. Ad esempio, il totale dei goal, assist e tiri riportati nella dashboard riflette solo il contributo dei giocatori selezionati nelle stagioni filtrate: 360 goal, 123 assist e 2.000 tiri totali. Questo approccio consente di effettuare confronti mirati tra i giocatori filtrati, analizzando le loro performance individuali e collettive in un contesto ristretto. Inoltre, il grafico a dispersione dei goal e dei tiri dei giocatori evidenzia come Lionel Messi e Cristiano Ronaldo continuino a dominare sia in termini di tiri totali sia di goal segnati, confermando la loro efficacia realizzativa nelle stagioni considerate. Neymar, Gareth Bale e Robert Lewandowski si posizionano più in basso, indicando un numero inferiore di tiri e goal rispetto ai primi due. Il filtro stagionale permette di identificare specifici trend di performance in questi anni, mentre il filtro per giocatore garantisce un'analisi comparativa più mirata, ideale per evidenziare le differenze tra i protagonisti del calcio mondiale in un determinato periodo.

4.3.2 Dashboard 2: Analysis of goals, xGoals, shot and shot on target of teams per season

La successiva dashboard, mostrata nella Figura 23, offre un'analisi approfondita delle performance delle squadre di calcio, focalizzandosi su metriche chiave come goal segnati, goal attesi detti *xGoals*, tiri totali e tiri in porta, distribuiti lungo più stagioni. Attraverso una visualizzazione chiara e interattiva, consente di identificare trend stagionali, confrontare l'efficienza offensiva delle squadre e valutare la correlazione tra le aspettative di performance e i risultati effettivi. Inoltre, la dashboard fornisce un quadro completo delle capacità delle squadre nel trasformare i tiri in risultati concreti, permettendo analisi comparative tra club di diverse leghe o stagioni. Questo approccio integrato rende il foglio uno strumento essenziale per comprendere le dinamiche del calcio moderno, supportando decisioni tattiche e strategiche.

Procediamo ora con una descrizione dettagliata degli elementi costitutivi del foglio:

- *In alto a sinistra* troviamo il grafico a linee che rappresenta l'andamento dei goal segnati da ciascuna squadra nelle diverse stagioni.
- *In alto a destra* vi è un grafico a dispersione che mostra la relazione tra goal totali e *xGoals* per ogni squadra.
- *In basso a destra* il grafico a barre confronta i tiri totali e i tiri in porta per ogni squadra.

Questa dashboard è stata progettata per fornire un'analisi completa delle performance delle squadre di calcio nel tempo, con un focus su metriche chiave come goal, *xGoals*, tiri totali e tiri in porta. L'obiettivo principale è aiutare gli utenti a comprendere i trend stagionali, la precisione offensiva delle squadre e il loro rendimento in relazione alle aspettative. La dashboard in alto a sinistra utilizza un grafico a linee per mostrare l'andamento dei goal segnati da ciascuna squadra nel corso delle stagioni. Questa visualizzazione è utile per identificare trend di miglioramento o peggioramento, costanza o variabilità nella performance di ogni squadra, consentendo agli analisti di concentrarsi su squadre che mostrano cambiamenti significativi nel rendimento. La dashboard in alto a destra presenta un grafico

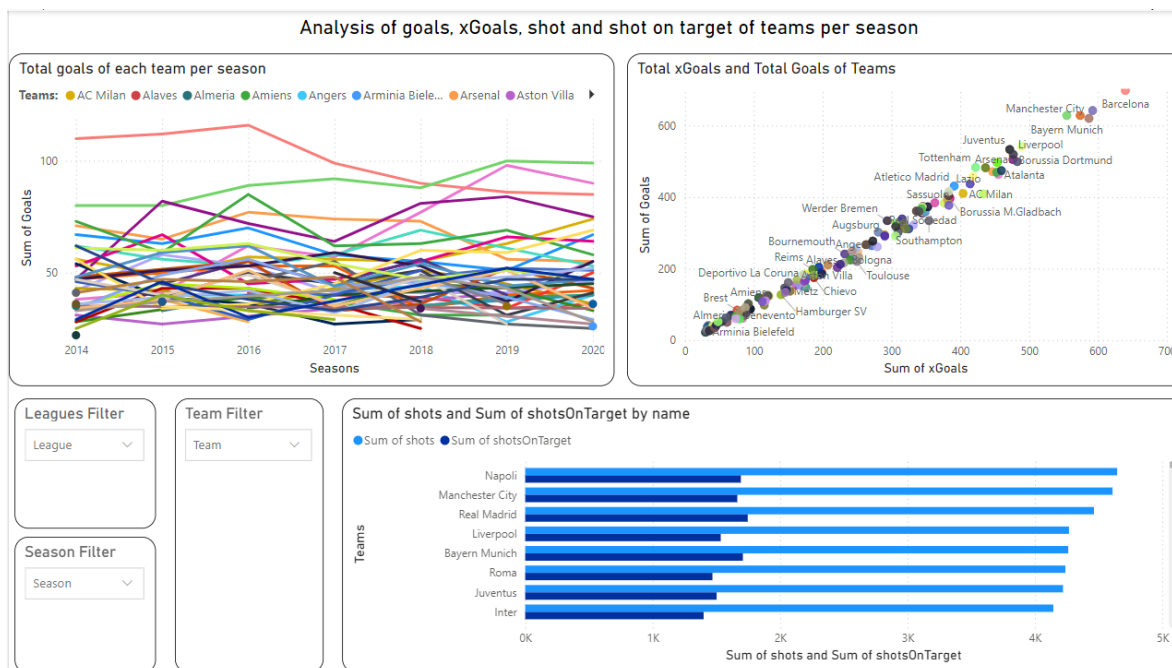


Figura 23: Dashboard 2: Analysis of goals, xGoals, shot and shot on target of teams per season

a dispersione che mette in relazione i goal segnati con gli *xGoals* per ogni squadra. Questo strumento consente di valutare quanto le squadre abbiano rispettato, superato o deluso le aspettative in termini di rendimento offensivo. Squadre che si trovano vicine alla diagonale mostrano una corrispondenza tra goal reali e goal attesi, mentre quelle lontane dalla linea evidenziano una performance oltre o sotto le aspettative. La dashboard in basso a destra utilizza un grafico a barre per confrontare i tiri totali e i tiri in porta delle squadre. Questa analisi aiuta a valutare l'efficienza delle squadre nell'inquadrare la porta, fornendo un'indicazione chiara sulle capacità offensive e sull'efficacia realizzativa delle diverse squadre. Complessivamente, questo foglio offre una panoramica olistica delle capacità offensive delle squadre nel calcio, analizzando trend stagionali, efficienza rispetto agli obiettivi e precisione dei tiri.

Oltre a ciò è possibile applicare dei filtri:

- Leagues filter: scegliere uno o più leghe europee.
- Season filter: scegliere una o più stagioni.
- Team filter: scegliere una o più squadre.

Come mostrato nella Figura 24, sono stati applicati filtri specifici per isolare le analisi sulle squadre che partecipano alla La Liga nelle stagioni 2015, 2016 e 2020. Inoltre, il filtro sulle squadre restringe l'analisi a sei team: Atletico Madrid, Granada, Real Madrid, Real Sociedad, Valencia e Villarreal. Questi filtri consentono di concentrare l'analisi su un insieme mirato di dati, escludendo informazioni non pertinenti e facilitando il confronto diretto tra le squadre selezionate. Grazie ai filtri applicati, il foglio offre un'analisi dettagliata e focalizzata sulle performance di queste squadre in termini di goal segnati, goal attesi, tiri totali e tiri in porta. Nel grafico a linee in alto a sinistra, è possibile osservare l'andamento dei goal segnati da ciascuna squadra nelle stagioni selezionate, evidenziando trend di crescita o calo. Il grafico a dispersione in alto a destra mostra la correlazione tra i goal effettivi e gli *xGoals*, evidenziando squadre come il Real Madrid che sovraperformano rispetto alle

aspettative, mentre altre, come Granada, sono più distanti dalla linea ideale di corrispondenza. Infine, il grafico a barre in basso a destra confronta i tiri totali e i tiri in porta delle squadre filtrate, mostrando chiaramente il dominio offensivo del Real Madrid in termini di volume di tiri rispetto alle altre squadre. Complessivamente, il foglio risulta essere uno strumento preciso e utile per comprendere le dinamiche offensive delle squadre della La Liga nelle stagioni selezionate, fornendo insight chiari e comparativi tra le squadre considerate.

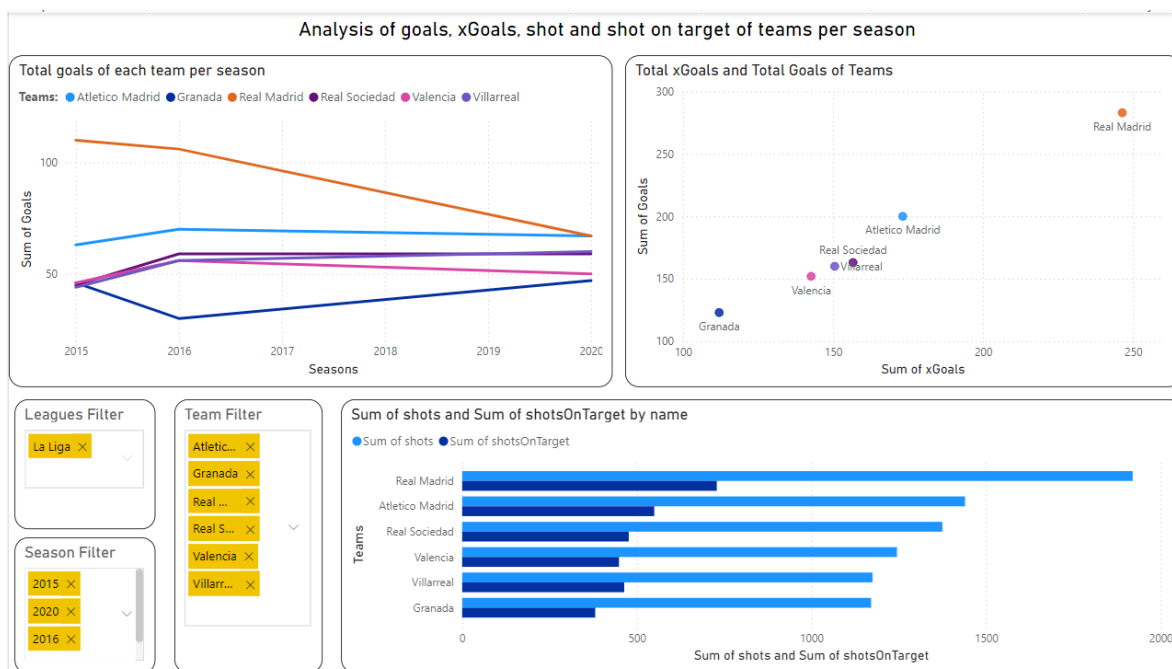


Figura 24: Dashboard 2 filtrata per lega, squadre e stagione

4.3.3 Dashboard 3: Analysis of bookmakers bets for a match

L'ultima dashboard, mostrata nella Figura 25, fornisce un'analisi dettagliata delle quote dei bookmaker e delle statistiche sui goal nelle partite di calcio, evidenziando l'evoluzione delle probabilità e dei risultati nel tempo. L'obiettivo principale è offrire una panoramica sulle tendenze delle quote per le vittorie in casa, i pareggi e le vittorie in trasferta, mettendole in relazione con i goal segnati. Le analisi consentono di valutare l'accuratezza delle previsioni dei bookmaker, confrontando le quote con i risultati reali, e di identificare eventuali differenze significative tra le aspettative e i dati effettivi. La distribuzione delle probabilità, combinata con i risultati complessivi e le quote più alte per ogni partita, offre agli utenti uno strumento utile per comprendere le dinamiche delle scommesse e l'andamento delle partite, supportando decisioni strategiche o approfondimenti analitici.

Procediamo ora con una descrizione dettagliata degli elementi costitutivi del foglio:

- Il grafico *in alto a sinistra* è un grafico a linee che rappresenta l'andamento delle medie delle quote dei bookmaker per vittorie in casa, pareggi e vittorie in trasferta nel corso delle stagioni.
- Il grafico *in alto a destra* è un grafico ad area che rappresenta il totale dei goal segnati in casa e in trasferta per ciascuna stagione.

- *In basso a sinistra* è presente una tabella che riporta le partite con le quote massime offerte dai bookmaker per vittorie in casa, pareggi e vittorie in trasferta, organizzate per stagione, squadra di casa e squadra in trasferta.
- *In basso a sinistra* è presente un grafico a torta che mostra le probabilità medie di vittoria in casa, vittoria in trasferta e pareggio, evidenziando la distribuzione percentuale di queste probabilità calcolate dai bookmaker.

I grafici di questa dashboard sono progettati per analizzare le tendenze delle quote dei bookmaker e metterle in relazione con i risultati effettivi delle partite di calcio. Il grafico a linee in alto a sinistra mostra l'evoluzione delle quote medie per vittorie in casa, pareggi e vittorie in trasferta nel tempo, consentendo di osservare eventuali cambiamenti nelle aspettative dei bookmaker stagione dopo stagione. Il grafico ad area in alto a destra confronta i goal totali segnati in casa e in trasferta, evidenziando l'andamento complessivo e il contributo di ciascuna categoria nel tempo. Il grafico a torta in basso a sinistra rappresenta la distribuzione percentuale delle probabilità di vittoria in casa, pareggio e vittoria in trasferta, offrendo una visione immediata del peso relativo delle tre categorie. Infine, la tabella in basso a sinistra elenca le partite con le quote più alte per vittorie in casa, pareggi e vittorie in trasferta, fornendo una visione dettagliata delle scommesse più significative per ogni stagione. Complessivamente, questi grafici aiutano a valutare l'accuratezza delle previsioni dei bookmaker e le dinamiche delle probabilità rispetto ai risultati reali.

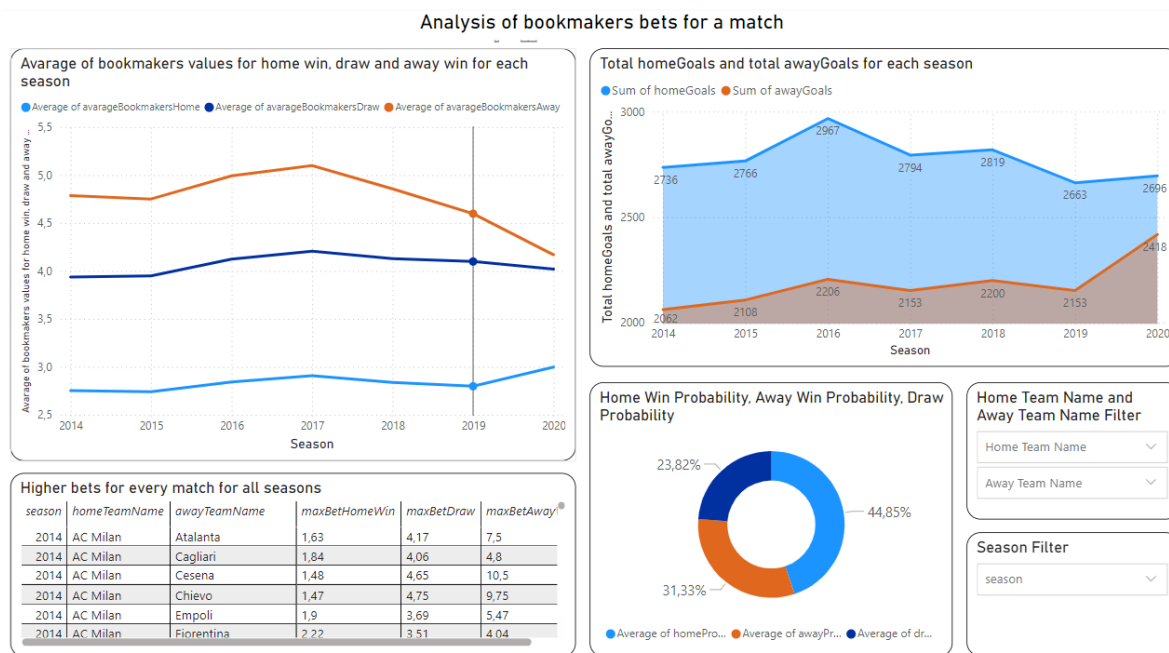


Figura 25: Dashboard 3: Analysis of bookmakers bets for a match

Oltre a ciò è possibile applicare dei filtri:

- Home Team Name and Away Team Name: scegliere la squadra di casa e la squadra ospite.
- Season filter: scegliere una o più stagioni.

Come viene mostrato nella Figura 26, sono stati applicati filtri specifici per analizzare esclusivamente le partite in cui l'AC Milan è stato scelto come squadra di casa e il Bologna come squadra in trasferta,

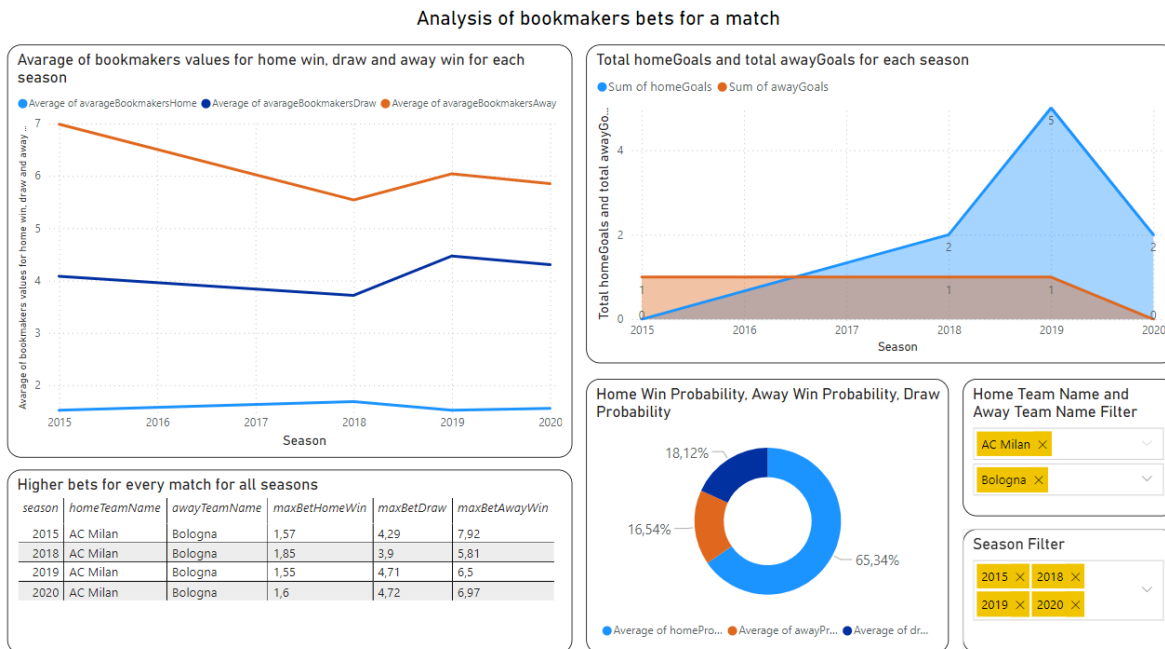


Figura 26: Dashboard 3 filtrata per squadre e stagioni

nelle stagioni 2015, 2018, 2019 e 2020. Questi filtri consentono di focalizzare l'analisi sui dati relativi a questi incontri, permettendo di osservare in modo mirato le dinamiche delle quote dei bookmaker e i risultati effettivi. Grazie a questa selezione, il foglio mostra l'andamento delle quote medie dei bookmaker per vittorie in casa, pareggi e vittorie in trasferta, evidenziando come le aspettative siano cambiate nel tempo per le partite tra queste due squadre. I goal totali segnati in casa e in trasferta per ogni stagione forniscono un confronto diretto delle performance offensive del Milan e del Bologna. Il grafico a torta illustra la distribuzione delle probabilità medie calcolate dai bookmaker, con una netta preferenza verso le vittorie in casa del Milan rispetto ai pareggi e alle vittorie in trasferta del Bologna. Infine, la tabella in basso elenca le partite filtrate, mostrando per ciascuna stagione le quote massime per vittorie in casa, pareggi e vittorie in trasferta, offrendo una visione dettagliata delle scommesse più significative. Complessivamente, il foglio si presenta come uno strumento specifico e dettagliato per comprendere le dinamiche di queste partite, sia dal punto di vista delle aspettative dei bookmaker che dei risultati effettivi.

Siti web consultati

1. Understat: provides modern football figures like 'expected goals' <https://understat.com/>
2. Football data: comprehensive betting data <https://www.football-data.co.uk/>