

# BETS INSIGHTS



## Group 4

Davide Airaghi – *Engineer*  
Luca Gaddoni – *Business Analyst*  
Daniele Raimondi – *Data Scientist*  
Federico Reggiani – *Mathematician*  
Chiara Teruzzi – *Engineer*



# AGENDA

1. Obiettivi

2. Introduzione

3. Raccolta dati e Architettura

4. Insight e Visualizzazione

5. Difficoltà riscontrate e Limiti Soluzione

6. Next Step

# OBIETTIVI



- ☐ Capire le variabili e gli eventi che influenzano l'**andamento delle quote delle partite di calcio**

NEXT STEP

- ☐ Individuare e analizzare le **inefficienze del mercato**

NEXT STEP

- ☐ Stimare la **curva su cui si muovono le quote**

# INTRODUZIONE

Definizioni e concetti chiave



$$\text{Quota} = \frac{1}{\text{Probabilità}}$$



La probabilità reale è da stimare e viene modellata in base alle **aspettative del mercato**.



Ogni **scommettitore ha aspettative differenti** e le valuta soggettivamente e, a livello aggregato, domanda e offerta determinano la probabilità attesa e quindi la quota

*Da cosa dipendono le aspettative?*



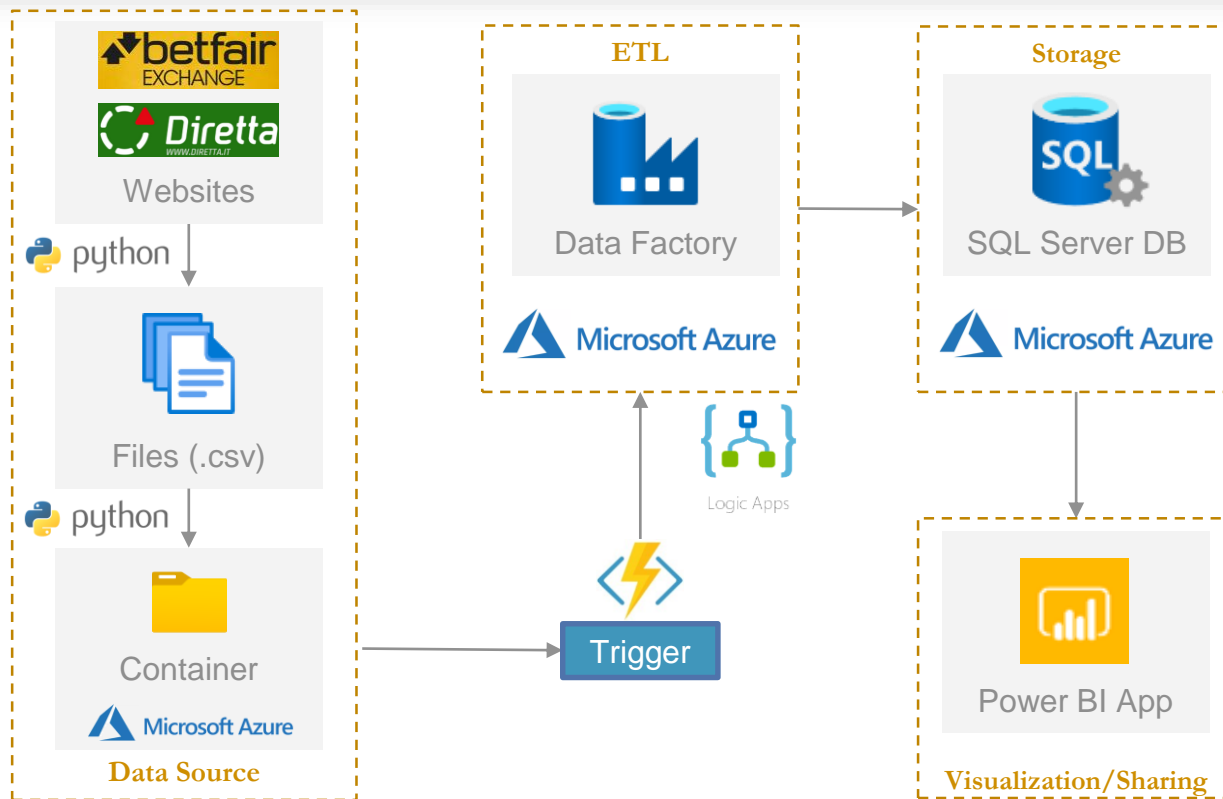
- **Tempo mancante** (trend di lungo)
- **Sviluppo della partita** (oscillazioni di medio)
- **Occasioni pericolose** (oscillazioni di breve)

*Dove si può intervenire?*



- **Ci sono fasi in cui il mercato valuta non correttamente le quote?**
- **Si riesce a prevedere quando ci sono questi eccessi e trarne vantaggio?**

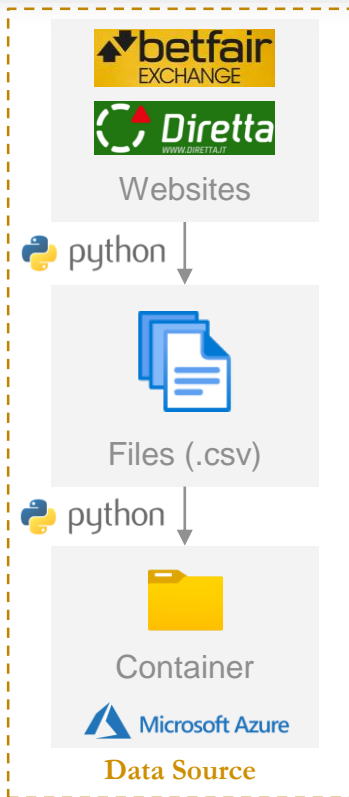
# ARCHITETTURA





# RACCOLTA DATI

Scraper su Notebook Colab



Lettura **real-time** durante i match tramite supporto di **Selenium**

Generazione di **tre file** in formato csv (con id-match come identificativo): **Quote** (Betfair), **Statistiche** e **Commenti** (Diretta.it)

Invio dei file a **container Azure**

# AZURE PIPELINE

## Data Factory



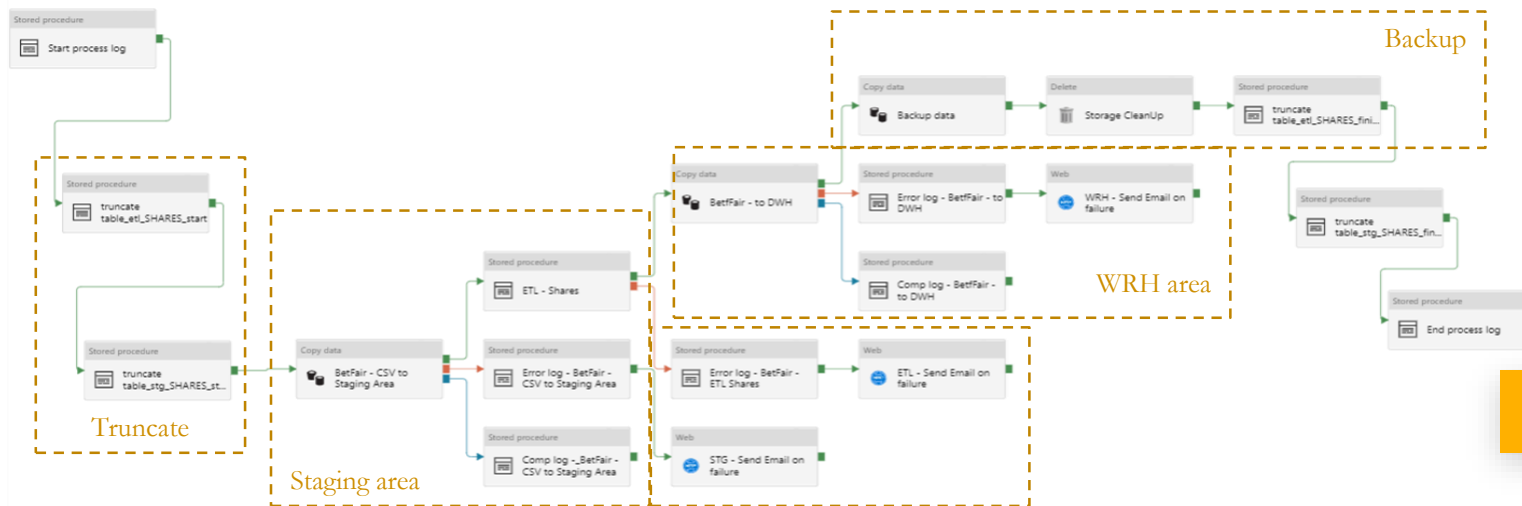
**Trigger:** ogni pipeline ha associato un apposito trigger (nuovo file depositato sul container) che scatena l'esecuzione della stessa.

**Truncate:** all'inizio dell'esecuzione vengono troncate la tabella di staging (\_stg) e quella dove avvengono le trasformazioni (\_etl) per eliminare i dati derivanti da precedenti esecuzioni.

**Staging area:** caricamento dei dati dal csv e trasformazioni (aggregazioni, filtri, pivoting)

**Data warehouse:** popolazione delle tabelle finali con i nuovi dati

**Back-up:** spostamento dei file caricati in un container di back-up e clean up storage

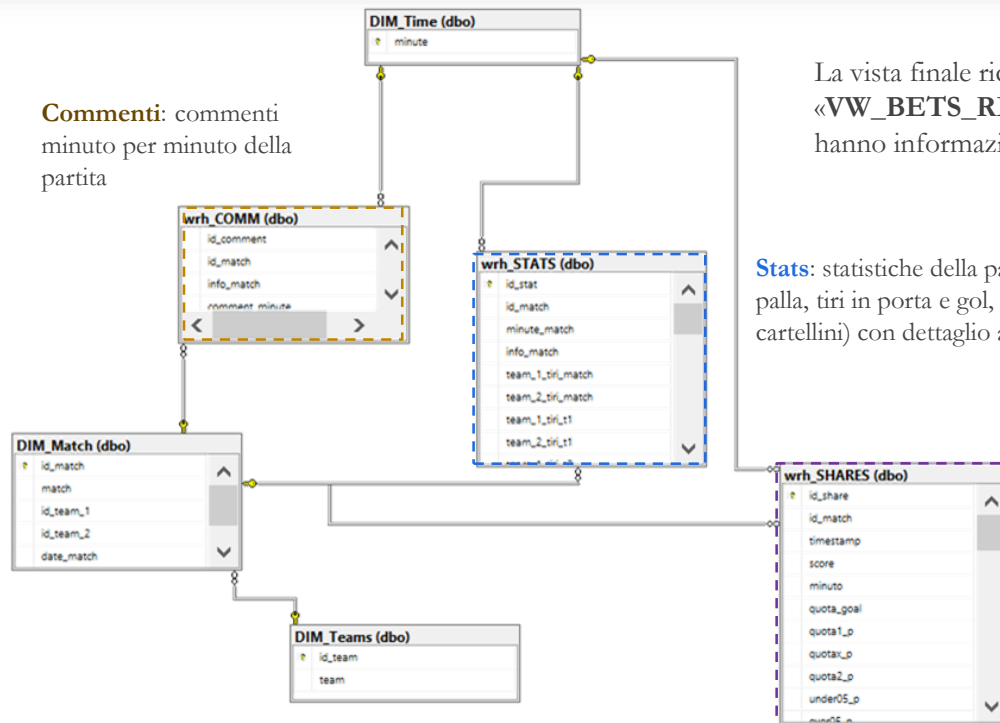


# DATA MODEL

## Diagramma



**Commenti:** commenti  
minuto per minuto della  
partita



La vista finale riconciliata per le analisi è  
«**VW\_BETS\_RECONC**» dove per minuto e match si  
hanno informazioni di Statistiche, Quote e Commenti.

**Stats:** statistiche della partita (possesto  
palla, tiri in porta e gol, calci d'angolo,  
cartellini) con dettaglio al minuto

**Shares:** Quote 1, X, 2,  
Under e Over per ogni  
minuto giocato



# INSIGHT

UEFA Nations League - 6/9/2020



Serbia

0-0

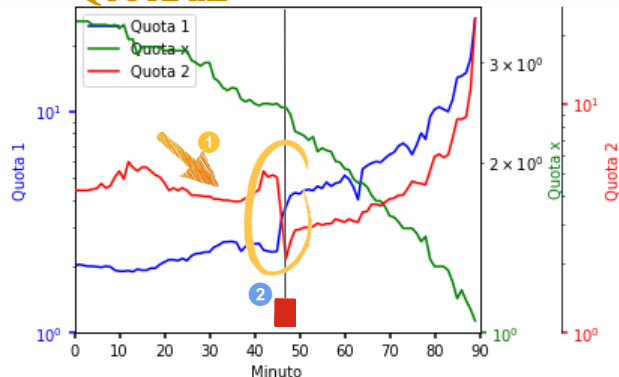
Finale



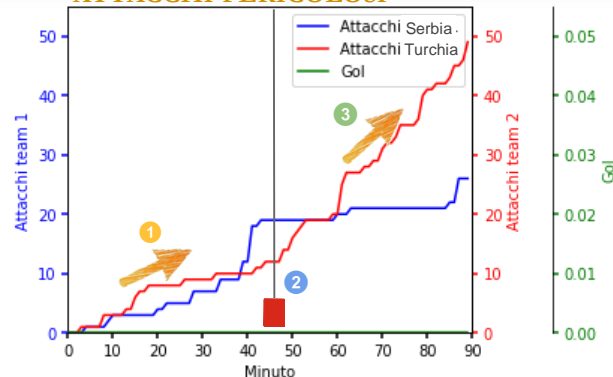
Turchia



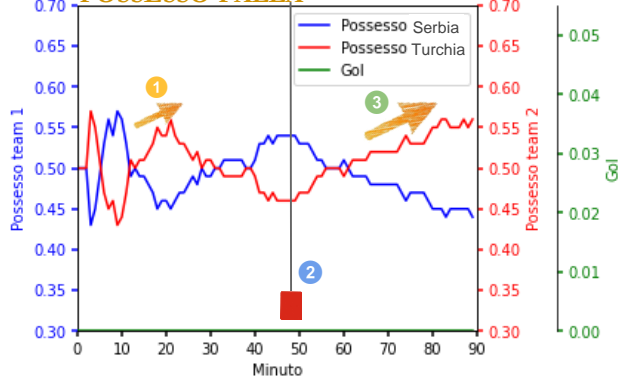
## QUOTE 1X2



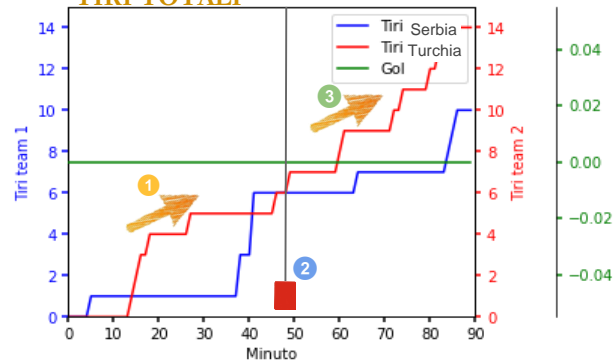
## ATTACCHI PERICOLOSI



## POSSESSO PALLA



## TIRI TOTALI



1 Dal 15' al 40' la Turchia ha il **predominio** → il mercato sconta facendo **calare la sua quota**.

2 Al 49' la Serbia resta in 10: la Turchia, forte dell'**uomo in più**, diviene più appetibile → la sua **quota crolla**.

3 D'ora in poi la Turchia riprende il dominio, sebbene non riesca mai a segnare → il **pareggio converge a 1**.

# INSIGHT

UEFA Nations League - 5/9/2020



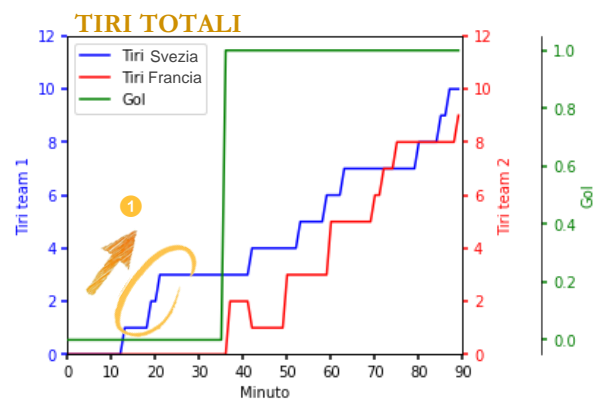
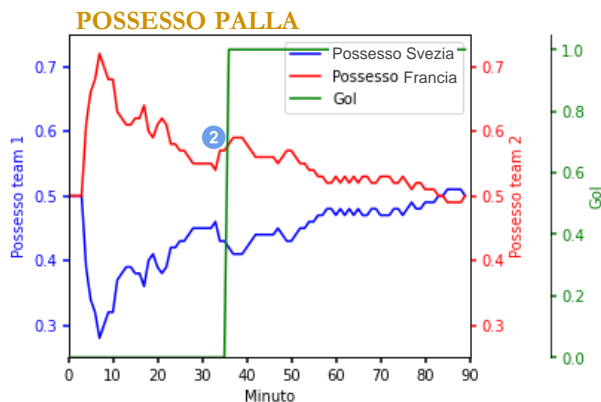
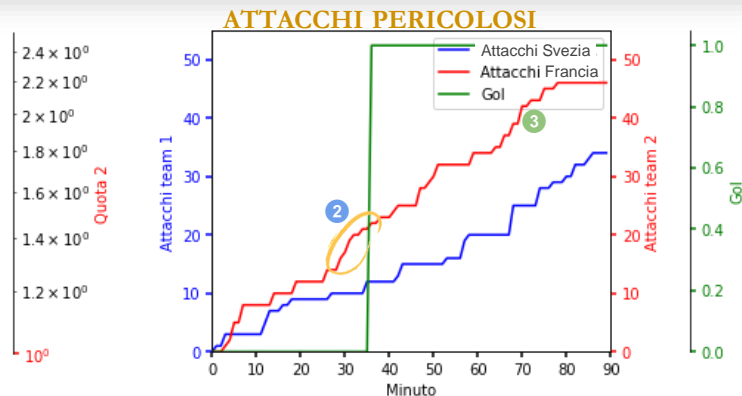
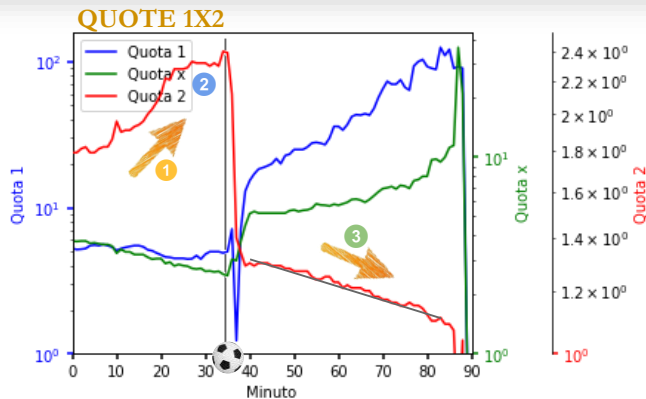
Svezia

0-1

Finale



Francia



1 Da inizio partita fino al 25' la Svezia è più pericolosa e la **quota** della Francia **cresce**

2 Dal 25' al Gol, la Francia aumenta l'intensità e la quota smette di crescere, perché le **aspettative del Gol** crescono. Poco dopo viene segnato il Gol e la quota crolla

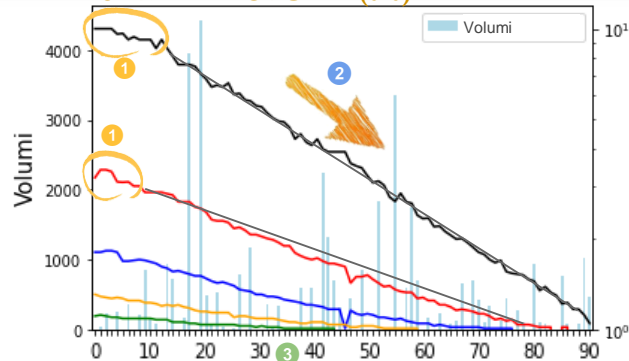
3 La Francia mantiene le redini del gioco e la sua **quota converge linearmente** verso l'unità

# INSIGHT

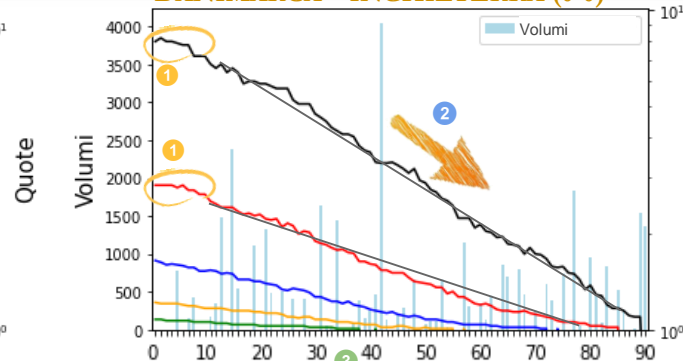
UEFA Nations League – UNDER trend



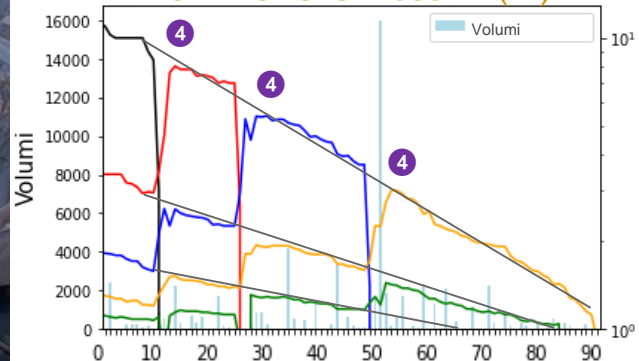
SERBIA – TURCHIA (0-0)



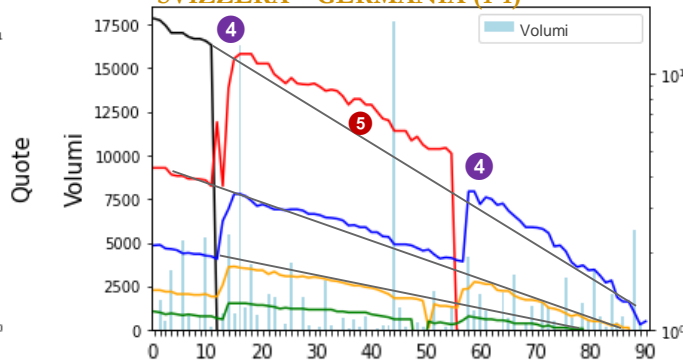
DANIMARCA – INGHILTERRA (0-0)



REPUBBLICA CECA – SCOZIA (1-2)



SVIZZERA – GERMANIA (1-1)



1 Nei primi 10 minuti, dove i gol sono meno frequenti, la curva dell'under rimane flat

2 Dopo inizia la discesa lineare verso quota 1 → aumenta la probabilità della selezione

3 Solo l'under più basso ancora valido converge a 1 a fine partita; gli altri lo fanno progressivamente prima

4 Appena viene segnato il primo gol, ad esempio, viene chiuso il mercato dell'under 0.5 e la quota dell'under 1.5 atterra sulla retta che avrebbe seguito l'under 0.5.

Questo succede con gli altri under ad ogni gol

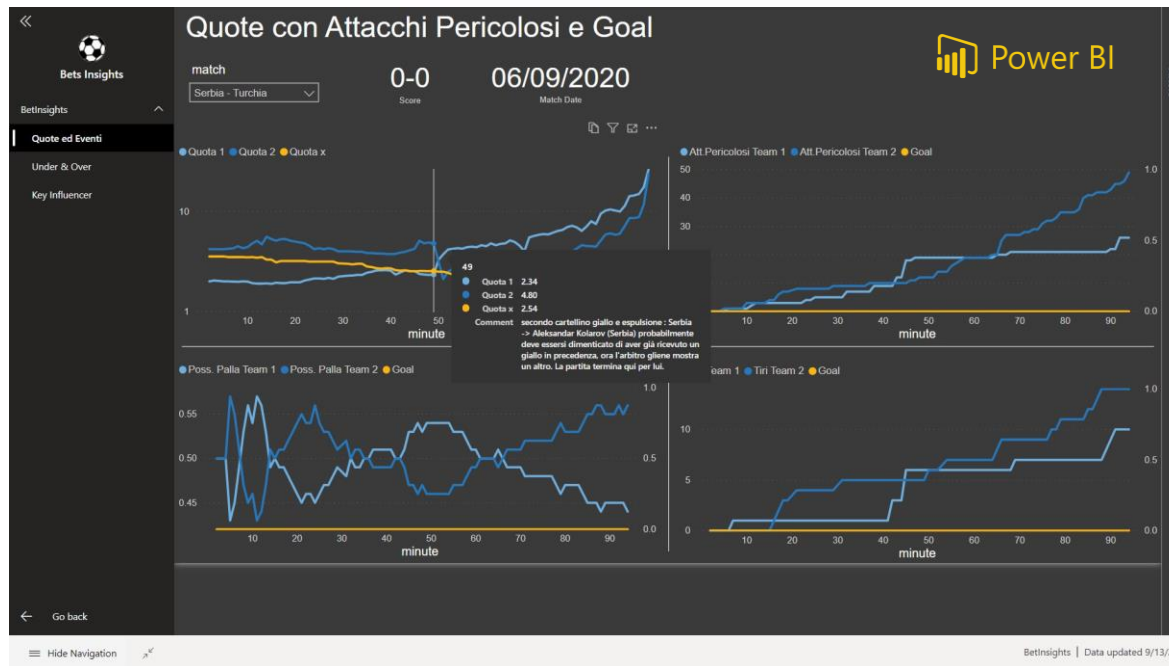
5 Molte inefficienze!

# VISUALIZZAZIONE RISULTATI

Microsoft Power BI App



I risultati sono stati poi visualizzati su **Microsoft Power BI**, connettendosi in **modalità Import** al **Database su Azure**.



# DIFFICOLTÀ RISCONTRATE E LIMITI SOLUZIONE



## Difficoltà

- **Disponibilità dei dati** in intervalli temporali ristretti perché scaricati da eventi real-time (match)
- **Possibilità limitate per testare** gli scraper (solo durante l'evento)
- **Integrazione dei dati** provenienti da siti diversi
- **Variazioni continue dei siti web** che implicano aggiornamenti degli scraper

## Limitazioni

- **Due scraper** diversi in cui i parametri sono **modificati manualmente** a seconda del match
- Su Betfair è necessario lanciare **un web driver per ogni quota**, questo appesantisce lo scarico di dati ma a seconda dei goal vengono chiusi di volta in volta i web driver dei mercati di riferimento



## NEXT STEP



- **Automatizzazione** di tutto il processo di scarico dati
- Creazione di un **modello di machine learning** che stimi le curve delle quote per evidenziare le inefficienze
- Creazione di un **tool automatizzato** che compri e venda quote giocando sulle inefficienze (verificando vincoli legali legati alla soluzione)

# THANKS!

- Davide Airaghi
- Luca Gaddoni
- Daniele Raimondi
- Federico Reggiani
- Chiara Teruzzi

