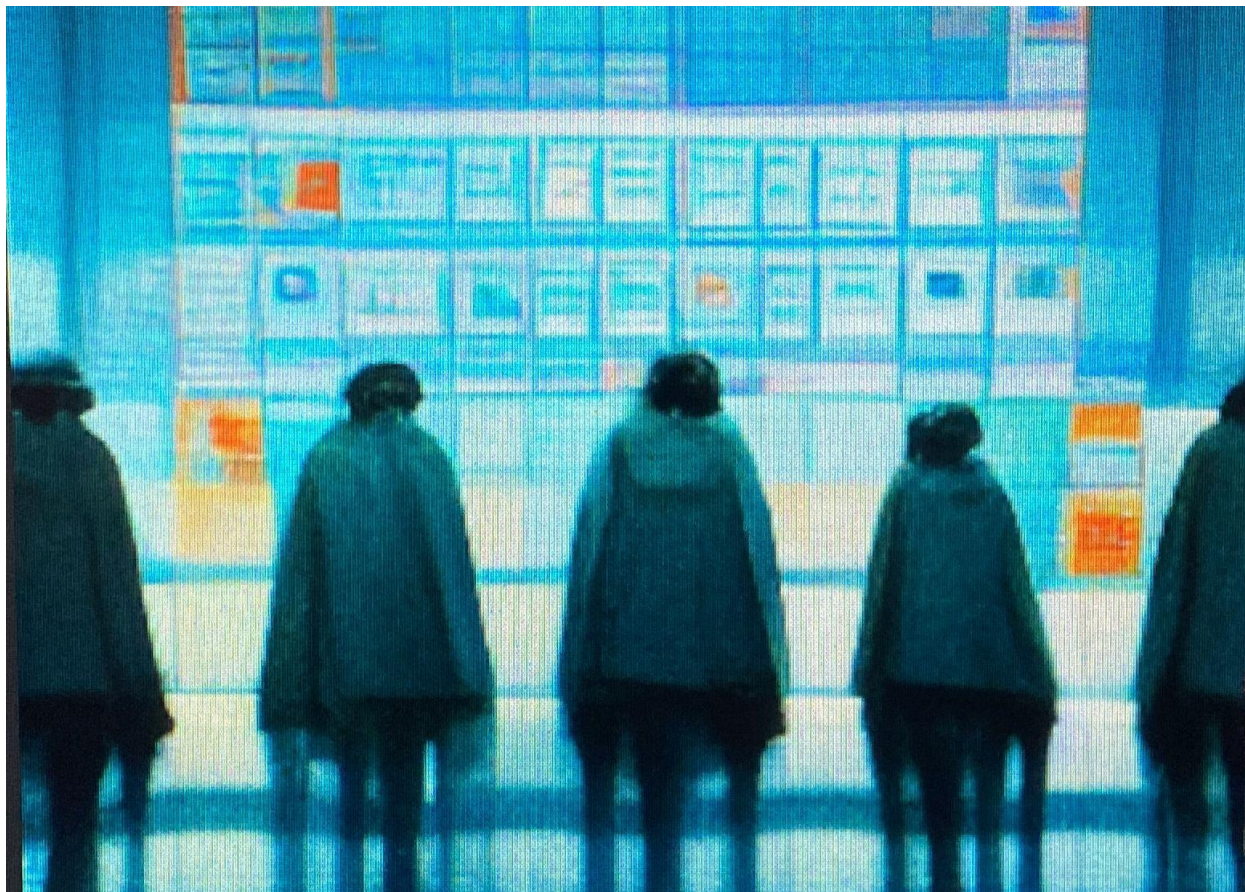


# Progetto FIFA 19

## ANALISTI ANONIMI

---



### Introduzione

Questo logo rappresenta gli "Analisti Anonimi". E' stato generato grazie ad un' Intelligenza Artificiale chiamata [Midjourney](#) che permette di generare immagini dando in input delle caratteristiche.

---

---

## Descrizione delle modalità di lavoro

L'organizzazione del progetto è partita da un **brainstorming** iniziale di tutti i componenti degli **"Analisti Anonimi"**, andando a definire come sarebbe stato svolto il progetto.

Ogni componente del gruppo ha scelto i grafici che avrebbe voluto creare, in modo tale da non fare gli stessi. Abbiamo successivamente creato un **server Discord**, dove abbiamo avuto modo di coordinarci tramite riunioni vocali e presentazioni video per la realizzazione del progetto. Inoltre, per rimanere in costante contatto tra di noi, abbiamo creato anche un **gruppo Whatsapp**, dove ci siamo tenuti quotidianamente aggiornati sull'andamento del progetto.

Abbiamo creato una **parte comune**, che è quella rispettivamente di `fifa_import` e `fifa_clean`, dove tutti i membri del gruppo hanno dato il proprio contributo per rendere il dataset pulito ed utilizzabile nella fase di visualizzazione.

Ognuno di noi, per prima cosa, ha portato a termine la creazione dei propri grafici. Successivamente gli abbiamo uniti in unico file, prestando attenzione a fattori come: l'omologazione le variabili, efficienza dei processi e testing di corretto funzionamento. Tutto ciò coordinandoci nella scrittura del codice e collegandoci quotidianamente sul server **Discord**.

Per la creazione del file **Word**, abbiamo lavorato in sincronia coordinandoci tramite il server Discord, proprio come per l'importazione, la pulizia e la visualizzazione dei dati. Abbiamo utilizzato un servizio fornito da Google: **Google Docs**. Quest'ultimo ci ha permesso di lavorare simultaneamente al progetto, in modo da completare il proprio lavoro singolarmente per poi cooperare successivamente nella parte in comune.

---

## Data Import

Abbiamo cominciato importando il pacchetto **Tidyverse** per poter manipolare il dataset. Abbiamo fatto lo stesso anche nel **data\_clean** e nel **data\_viz**. In quest'ultimo, ovviamente, abbiamo avuto bisogno anche di molte altre librerie per poter eseguire le nostre analisi. Le librerie richieste sono elencate all'inizio di ogni script.

Convertiamo il dataset **fifa.csv** in un database e lo salviamo su una variabile.

Notiamo che non ci sono righe NA ne iniziali ne finali.

Diamo un'occhiata al nostro dataset e controlliamo le prime e le ultime righe per essere certi che non vi siano valori anomali

Salviamo il dataset in formato dati .rda per poterlo poi importare nel successivo script.

## Data Clean

Carichiamo dunque i dati in formato .rda che ci eravamo salvati prima.

**Rimuoviamo le righe che contengono NA** per fare in modo che non diano problemi nella fase di analisi dei dati.

Rimuoviamo inoltre alcune variabili che non ci interessano e che appesantirebbero solamente il dataset.

Abbiamo poi deciso di rinominare le modalità della variabile Position, scrivendole in forma estesa, così da facilitarne la lettura.

Trasformiamo infine tutte le variabili che sono stringhe e quindi non numeriche in fattori per poterle trattare come dei fattori

Controlliamo il dataset e se non sono stati introdotti nuovi errori lo salviamo in un altro file .rda

---

## 1 | Daniele Mariani

Dopo aver svolto insieme agli altri membri del gruppo la parte comune, ho creato il mio script di visualizzazione dei dati. Nel mio caso ho fatto 4 grafici.

I grafici che ho realizzato sono rispettivamente:

1. **Media dell'Overall per Età:** in questo grafico realizzato con la libreria **plot\_ly()**, che permette di realizzare **grafici interattivi e dinamici**, ho creato uno **Scatter Plot**, andando a rappresentare la **media dell'Overall secondo l'età dei giocatori**. La curva ci mostra che con l'aumentare dell'età, e quindi dell'esperienza, l'overall dei giocatori aumenta, andando ad appiattirsi intorno ai 30 anni, per poi iniziare a scendere.
2. **Conteggio dei Giocatori per Nazionalità:** in questo grafico ho contato il **numero di giocatori per nazionalità**, per poi andare a visualizzare il risultato su un **planisfero**. Essendo il grafico interattivo ci si può muovere con il cursore, per vedere con esattezza il numero di giocatori per ogni nazione. A mio avviso questa visualizzazione risulta molto interessante, dato che seguendo la scala dei colori, si capisce bene che **il maggior numero dei giocatori proviene da zone in cui c'è una forte cultura calcistica**. Di fatti i colori giallo e verde si concentrano nell'America del Sud, in particolare Brasile e Argentina, e nel centro Europa, come in Italia, Spagna, Francia e Germania.
3. In questo grafico ho voluto fare un **confronto tra le skills** dei due giocatori che si contendono il podio da sempre, **Cristiano Ronaldo e Lionel Messi**. Le skills su cui sono andato a confrontarli sono **passaggi corti, precisione nei colpi di testa, conclusione, e cross**. I due ovviamente se la contendono, infatti sulla conclusione ed i cross sono alla pari, mentre **Cristiano Ronaldo**

---

**risulta più forte nella precisione nei colpi di testa, mentre Lionel Messi risulta più abile nei passaggi corti.**

4. Nel quarto ed ultimo grafico ho voluto rappresentare la **correlazione tra età e stipendio** (che nel grafico è espresso come migliaia di euro settimanali). Analizzandolo si intuisce che le due variabili inizialmente sono correlate positivamente, vale a dire che all'aumentare dell'una aumenta anche l'altra. Nella fascia d'età tra i 25 e i 30 anni abbiamo il picco dello stipendio dei giocatori. Superata questa fascia **per la maggior parte dei giocatori** lo stipendio inizia a diminuire, infatti abbiamo una **correlazione negativa**, dato che all'aumentare dell'età lo stipendio diminuisce.

## 2 | Stefano de Saraca

Le funzioni da me elaborate sono volte alla visualizzazione in modo rapido e intuitivo dei dati che possono risultare utili sia alle squadre stesse, che agli appassionati, ma anche a dei possibili investitori per i club sportivi.

1. **Top N di un determinato parametro numerico:** questa funzione si occupa dell'elaborazione di una classifica costituita da N elementi di un certo parametro. Ad esempio una top 10 dei giocatori più pagati (stipendio), ma anche dei giocatori più giovani, forti (strength), ecc.
2. **Rapporto tra la forza e il BMI:** l'intuizione e la curiosità per lo sviluppo di questa funzione sono nate dall'idea che un giocatore con un BMI alto possa avere di conseguenza una forza maggiore rispetto a quelli con un BMI inferiore. In effetti, visualizzando i risultati e con l'ausilio di un grafico, possiamo notare come al crescere del BMI aumenta anche la forza seppure con un po' di dispersione. Curioso notare come il giocatore più forte con 97 punti strength sia anche quello con il BMI più alto in assoluto.

- 
3. **Correlazione tra valore del giocatore e stipendio:** sarà capitato a tutti di chiedersi quanto guadagni un giocatore in particolare, ma forse meno volte quanto lo stipendio sia rapportato al suo valore di mercato. Con questa funzione osserviamo, anche con l'ausilio di un grafico, l'andamento dello stipendio correlato al valore di mercato di un giocatore. In effetti, come intuibile all'aumentare del valore del giocatore segue anche un incremento della retribuzione. Curioso notare che il giocatore più "costoso" sul mercato non sia però il più retribuito per una sola posizione.
  4. **Età mediana per club sportivi:** ho pensato fosse interessante vedere quali sono i club sportivi più o meno giovani. Soprattutto da un punto di vista economico credo possa risultare utile sapere dove andare ad investire e quali potrebbero essere le squadre da dove "attingere" a nuovi giocatori professionisti giovani che possano essere integrati in club con età media maggiore. Possiamo osservare che alcune società sportive hanno età mediane molto basse, ad esempio 19 o 20 anni.

### 3 | Valerio Valentini

Nel primo grafico che ho realizzato, ovvero "relazione tra abilità e tra piede destro e sinistro", possiamo osservare la relazione che c'è tra i giocatori che preferiscono il piede destro o il piede sinistro e alcune abilità strettamente legate all'azione dei piedi durante il gioco, ovvero passaggio lungo e corto, crossing, dribbling, curvatura e controllo palla.

Per farlo prima ci isoliamo queste variabili insieme al piede preferito che fungerà da chiave primaria nella trasformazione del dataset in formato long, in modo da poter visualizzare più variabili sullo stesso asse. Per ogni abilità vengono realizzati due boxplot e i grafici vengono divisi in base all'abilità

Nel secondo grafico, ovvero "grafico di dispersione del salario settimanale in milioni in relazione all'Overall", possiamo osservare uno scatterplot ottenuto tramite la correlazione tra salario e punteggio generale, il tutto diviso in base alla posizione.



---

Nel terzo grafico, ovvero "relazione tra BMI e Agilità ", costruito in forma di funzione per permettere di scegliere l'abilità in base al quale realizzare questa analisi, ci creiamo un'altra variabile, ovvero il BMI, calcolato come il rapporto tra peso e altezza al quadrato, per ogni giocatore e lo mettiamo in relazione con uno scatterplot sovrapposto a un geo\_smooth, ad un'altra abilità a nostra scelta. Dividiamo i casi in base alla posizione e al working Rate, ovvero l'impegno dimostrato in campo. Nell'esempio sono stati messi in relazione il BMI e l'abilità di curvare la palla per gli attaccanti.

Nell'ultimo grafico, ovvero "scatterplot tra coppie di abilità", possiamo osservare la relazione tra coppie di variabili tramite un grafico ggpairs. Qui possiamo vederla tra sprint speed dribbling, ball control e work rate. Il risultato è una matrice in cui ogni cella contiene un grafico in base alla tipologia di variabili che riceve in input: boxplot( continua e discreta), scatterplot(continua continua), barplot( discreta, discreta), Grafico di correlazione (visualizza i coefficienti di correlazione tra le variabili), densità (mostra la distribuzione dei dati per ogni variabile continua, sulla diagonale principale della matrice) e Grafico a matrice di densità (visualizza le distribuzioni di frequenza incrociate tra le variabili categoriche).

## 4 | Emanuele Corradi

La relazione che ho elaborato riguarda l'analisi di 4 tipi di dati, di cui 3 sono relativi alla variabile "Stamina", mentre l'ultimo riguarda la media degli stipendi per nazionalità.

→ Inizialmente ho analizzato la correlazione tra la stamina e l'altezza dei giocatori, e ho notato che la fascia in cui la stamina è più alta è quella dei giocatori intorno al metro e ottanta. Inoltre, la percentuale di confidenza aumenta all'aumentare dell'altezza dei giocatori fino a raggiungere il massimo per i giocatori intorno al metro e ottanta e tende a decrescere una volta superata questa soglia di altezza. Quindi ho potuto definire che la stamina diminuisce per i giocatori che tendono ad essere più alti e più bassi della media.

---

→ In seguito ho analizzato la correlazione tra la stamina dei giocatori. Dall'analisi è emerso che i giocatori che hanno un peso medio intorno agli 80 chili presentano una resistenza maggiore rispetto alle altre fasce di peso. possiamo anche notare che per quanto riguarda la percentuale di confidenza c'è molta più incertezza nelle fasce di peso sotto i 70 Kg e sopra i 90 Kg. Tuttavia è importante sottolineare che la relazione tra il peso e la stamina dei giocatori non è lineare, poiché sia i giocatori troppo pesanti che quelli troppo leggeri tendono a mostrare una minore resistenza.

→ Come terzo tipo di analisi ho confrontato la stamina con l'età dei giocatori. In questo caso ho notato che la stamina dipende molto dall'età dei giocatori, e quelli intorno all'età di 28 anni risultano avere maggior resistenza. Inoltre la percentuale di confidenza è sempre alta in tutte le fasce d'età.

→ Infine, ho analizzato la media degli stipendi per nazionalità. Ho deciso di calcolare la media anziché fare la somma perchè essendoci un numero diverso di giocatori per ogni nazionalità non avremmo avuto dei dati chiari e veritieri. Come si può vedere dal grafico i giocatori di nazionalità spagnola risultano essere i più retribuiti.

→ In conclusione, le analisi effettuate mostrano che la stamina dei giocatori è influenzata dall'altezza, dal peso e dall'età, e che i giocatori di nazionalità spagnola hanno la media di stipendi settimanali più alta. Tuttavia, ci tengo a sottolineare che queste analisi sono limitate ai dati considerati e potrebbero non essere generalizzabili a tutti i giocatori di fifa 19.

## 5 | Aurelio Caccese

Nella mia parte di analisi del dataset ho realizzato sei grafici; 2 grafici che analizzano principalmente la correlazione tra due variabili del dataset in relazione al Work Rate 3 grafici di calore rappresentano le distribuzioni congiunte tra il Finishing ed altre



---

variabili, ed infine l'ultimo grafico rappresenta il miglior giocatore per ogni posizione.

- Il primo grafico, composto da 9 sotto-grafici divisi per l'impegno dei giocatori in campo, rappresenta la Correlazione tra *Peso* e *Altezza* diviso in base al *Work Rate*. In sintesi, il codice crea un grafico interattivo che mostra la relazione tra altezza e peso, con una curva di regressione per rappresentare la tendenza generale. Il colore delle curve è basato sulla variabile "Work.Rate" e viene visualizzata una rappresentazione separata per ogni valore univoco di "Work.Rate".
- il secondo grafico, composto anch'esso da 9 sotto-grafici divisi per l'impegno dei giocatori in campo, rappresenta la Correlazione fra *Overall* e *Età* diviso in base al *Work Rate*. In sintesi, il codice crea appunto un grafico interattivo che mostra la relazione tra l'età e l'overall della carta. Il colore delle curve è basato sulla variabile "Work.Rate" e viene visualizzata una rappresentazione separata per ogni valore univoco di "Work.Rate".
- Il terzo, quarto e quinto grafico rappresentano, tramite un grafico di calore, le distribuzioni congiunte tra il "Finishing" e altre variabili continue ("Shot Power", "Curve" e "Vision") per diversi valori della variabile "Work.Rate". Questo tipo di grafici utilizzano aree di densità 2D per visualizzare la distribuzione dei dati, tramite la funzione `geom_density_2d()`. In questo caso il grafico fornisce informazioni sulla probabilità relativa di finalizzazione dei giocatori, in base alla potenza di tiro, curva di tiro e visione di gioco; suddivisa in base alle 9 categorie (High, Medium, Low) di impegno in campo dei giocatori.
- Il sesto grafico rappresenta, tramite un grafico a barre (`geom_bar`), i migliori giocatori per ogni posizione creando così una sorta di squadra di "top player". Inizialmente ho creato una lista dei migliori giocatori per ogni posizione utilizzando la funzione (`find_best_player`) e tramite il grafico

---

interattivo creato con (ggplotly) possiamo vedere il nome dei giocatori ed il relativo overall della carta.