UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

ESCUELA DE ESTUDIOS DE POSTGRADO

MAESTRÍA EN INGENIERÍA CON ESPECIALIDAD EN CIENCIAS DE LA COMPUTACIÓN CON ESPECIALIZACIÓN EN CIENCIA DE DATOS

INTRODUCCIÓN AL ANÁLISIS DE DATOS

PROYECTO INICIAL ANÁLISIS DE CONJUNTO DE DATOS FIFA 2015- 2020

Jesfrin Jesser González Fuentes

999014909

Introducción	1
Objetivos	2
General	2
Específicos	2
Desarrollo del proyecto	3
Discovery	3
Proceso de análisis	5
Presentación de resultados	6
Diferenciación de términos	6
Análisis general de variables	7
Análisis de atributos	9
Atributos principales	9
Atributos de ataque	10
Atributos de defensa	12
Atributos en portería	13
Análisis de correlación	15
Preguntas y soluciones	17
Conclusiones y recomendaciones	25

Introducción

Uno de los videojuegos que ha mantenido su popularidad desde sus inicios es FIFA. Gracias a esta popularidad, ha sido capaz de reinventarse a lo largo de los años y agregar distintos modos de juego que captan la atención de los gamers. En la actualidad, es cada vez más necesario para los gamers estar informados sobre qué futbolistas y clubes deben considerar en los diferentes modos de juego para que la experiencia sea la mejor posible. Para los jugadores profesionales, esta información es aún más crucial. Con base en esto, la siguiente investigación busca presentar el proceso de análisis de datos de un dataset de FIFA, que abarca del año 2015 al 2020, y exponer los resultados de dicho análisis junto con recomendaciones para los gamers.

Objetivos

General

Realizar un análisis exhaustivo del dataset de FIFA del 2015 al 2020 para extraer información relevante que permita generar recomendaciones útiles para gamers interesados en optimizar su experiencia de juego al seleccionar jugadores y clubes en distintos modos de FIFA.

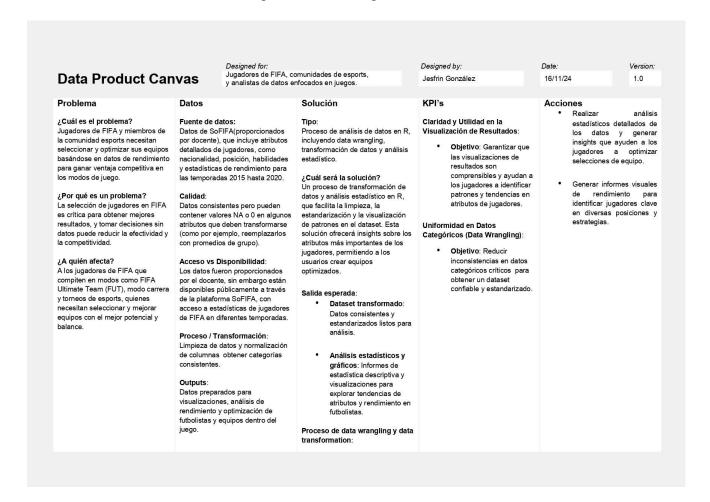
Específicos

- Identificar y analizar las características de los futbolistas, como atributos de rendimiento y
 habilidad, que podrían influir en su desempeño en el videojuego FIFA, facilitando la
 selección de jugadores en función de sus fortalezas.
- Explorar y comparar clubes y nacionalidades para determinar cuáles ofrecen los jugadores con mejores calificaciones y atributos en el juego, ayudando a los gamers a construir equipos competitivos.

Desarrollo del proyecto

Discovery

Para la fase de Discovery se realizó un Data Product Canvas, en el cual se abordaron distintos puntos para poder tener una visualización amplia de lo que se quería lograr con el análisis de datos. A continuación se presenta una imagen del Data Product Canvas realizado:



Data Wrangling: Creación Actores Hipótesis Los datos de rendimiento y atributos físicos en SoFIFA permiter realizar un análisis de nuevas variables y Cliente: Jugadores de FIFA y miembros de la comunidad de esports. reestructuración de datos para responder preguntas estadístico y de agrupamiento que ayuda a los jugadores de FIFA a optimizar sus equipos, identificando importantes. Stakeholders: Comunidades de análisis de datos jugadores con atributos específicos alineados a las estrategias deseadas. en videojuegos, empresas de videojuegos, y ligas de esports. Data Transformation: Estandarización y normalización de atributos Usuarios de la solución: Jugadores de FIFA, analistas de esports, y creadores de contenido. numéricos, creación de nuevas variables y conversión de variables Impacto: La solución proporciona una guía categóricas por medio de de datos para mejorar el rendimiento en el juego mediante análisis estadísticos y agrupamiento de jugadores. Valores / Riesgos Rendimiento / Impacto Valores Impacto: Tamaño del problema: Mejora en el rendimiento de los jugadores de FIFA al seleccionar Optimizar las selecciones de equipo en FIFA es un desafío constante para los jugadores y puede influir en equipos basados en análisis de datos. el éxito en competiciones de esports. Medición del impacto: Línea base: Análisis de la retroalimentación de usuarios y comparación del Antes de esta solución, los jugadores seleccionan equipos con criterios subjetivos o basados en opiniones rendimiento en el juego antes y después de aplicar el análisis de datos. de la comunidad. • Incremento de valor / ahorro: Visualización de la mejora: Mejora en el rendimiento y éxito en el juego mediante decisiones informadas basadas en análisis de datos y Informes de comparación y gráficos de rendimiento en diferentes atributos, resaltando la mejora en la selección y uso de jugadores en visualización. Riesgos Datos desactualizados que podriían no reflejar el rendimiento real en el videojuego. Cambiosen las actualizaciones de FIFA que modifiquen los valores o el impacto de ciertos atributos.

De igual manera se comparte el pdf del Data Product Canvas en la entrega del análisis para una mejor visualización.

Proceso de análisis

Para poder llegar a tener resultados en base al análisis se llevó a cabo el siguiente proceso:

- 1. **Data cleaning:** En la limpieza de datos se exploró el dataset e indago sobre lo que representa cada una de las variables del dataset. Se buscó generar un dataset con la menor cantidad de valores nulos, y que si existieran fueran intrascendentes o tuvieran una razón lógica del porqué se tomó la decisión de dejarlos con valores nulos. Se estandarizaron las variables categóricas pasándolas todas a mayúsculas y eliminando espacios. también optó por llenarlas con el dato "NO DATA" si la variable tenía un valor nulo.
- 2. Data analysis: Se realizó un análisis del dataset una vez limpiado para generar estadísticas descriptivas básicas. También se realizaron distintos histogramas para poder entender las habilidades de los futbolistas en el videojuego, asi como una matriz de correlación para comprender cómo interactúan ciertas variables.
- 3. **Data wrangling:** Se crearon nuevas variables y se formularon preguntas y sus soluciones para poder ayudar a los gamers a escoger jugadores o clubes en el videojuego.
- 4. Data transformation: Se normalizaron atributos numéricos, se aplicó One-Hot a variables categóricas que no fueran demasiado extensas, todo con el objetivo de posteriores análisis.

Presentación de resultados

Diferenciación de términos

Es importante diferenciar entre los siguientes términos para comprender mejor el análisis:

- 1. Gamers o jugadores de FIFA: usuarios del videojuego que buscan optimizar sus elecciones de equipos y jugadores en el juego.
- 2. Futbolista o Futbolistas del dataset: futbolistas de la vida real, cuyos atributos y habilidades están registrados en el dataset y reflejan sus habilidades físicas y técnicas en el mundo real.

Análisis general de variables

Se encontraron datos interesantes en varias variables. Se pueden resaltar los siguientes:

Nacionalidades: Hay un total de 179 nacionalidades representadas en el dataset.

Equipos: Se incluyen 918 equipos diferentes.

Posiciones de los futbolistas: Los jugadores pueden ocupar diversas posiciones, y

algunos están capacitados para jugar en más de una.

Pie preferido: Algunos jugadores cambian su pie dominante entre años (derecho o

izquierdo). Esto afecta a 849 jugadores de un total de 36,559, representando un 2.32%, lo

que sugiere posibles inconsistencias en los datos.

Tipo de cuerpo: Se identifican 3 tipos de cuerpo principales, además de tipos especiales

para jugadores excepcionales.

Cara real: 2,431 jugadores tienen una representación facial real en al menos un año del

dataset.

Préstamos: 5,310 jugadores fueron cedidos en préstamo en algún momento entre 2015 y

2020.

Edad:

Mínima: 16 años.

Máxima: 47 años.

o Promedio: 25 años.

Desviación estándar: 4.63 años.

Altura (cm):

Mínima: 154 cm.

o Máxima: 205 cm.

7

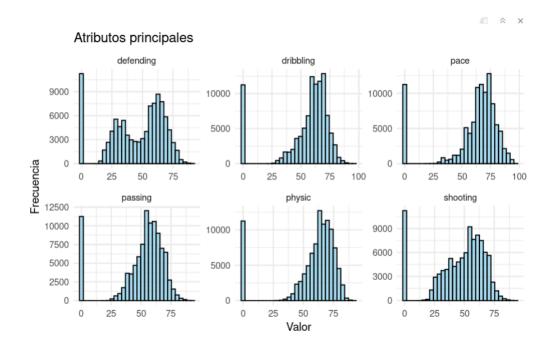
- o Promedio: 181 cm.
- Desviación estándar: 6.69 cm.
- Peso (kg):
 - o Mínimo: 49 kg.
 - o Máximo: 110 kg.
 - o Promedio: 75 kg.
 - o Desviación estándar: 6.98 kg.
- Overall (valoración general):
 - o Mínimo: 40.
 - o Máximo: 94.
 - o Promedio: 65.
 - Desviación estándar: 7.
- Potential (potencial):
 - Mínimo: 40.
 - Máximo: 95.
 - o Promedio: 70.
 - o Desviación estándar: 6.32.
- Valor (€):
 - Mínimo: 0 euros.
 - Máximo: 123 millones de euros.
 - Promedio: 2 millones de euros.
 - O Desviación estándar: 5 millones de euros.

Análisis de atributos

Una de las principales formas en las que un gamer se basa para la selección de jugadores y equipos es los atributos que poseen los futbolistas, por lo que se agrupó y se analizaron los siguientes atributos:

Atributos principales

Incluyen atributos como la defensa, el dribling, velocidad, pase, físico y disparo. Los histogramas generados son los siguientes:



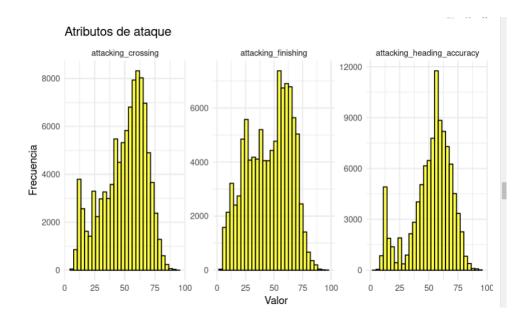
De forma general, se nota que varios futbolistas tienen una habilidad de 0 o cercana a 0 en los atributos de rendimiento y habilidad. Esto ocurre porque son porteros, y se les asignó 0 en estos aspectos debido a que no son relevantes para su posición.

 Defending: Hay una cantidad significativa de jugadores con habilidades defensivas en niveles medios y altos.

- Dribbling: La mayoría de los jugadores tienen una calificación de dribbling entre 50 y 80, lo que indica que la habilidad para regatear es relativamente buena en la mayoría de ellos.
- 3. **Pace:** La velocidad se concentra principalmente entre 50 y 80, sugiriendo que el promedio de velocidad de los jugadores es bastante alto.
- 4. **Passing:** La habilidad para pasar se encuentra principalmente entre 50 y 75, reflejando una capacidad razonable de pase en la mayoría de los jugadores, con pocos en los niveles más bajos y algunos alcanzando calificaciones destacadas.
- 5. **Physic:** La capacidad física de la mayoría de los jugadores está en un rango de 50 a 75, algo esperado en un deporte de alto rendimiento como el fútbol. Aunque hay algunos jugadores con poca capacidad física, son una minoría.
- 6. **Shooting:** La habilidad para disparar se encuentra principalmente entre 50 y 75, lo que indica que la mayoría tiene una capacidad moderada para disparar a gol, con algunos jugadores de élite destacándose en este aspecto.

Atributos de ataque

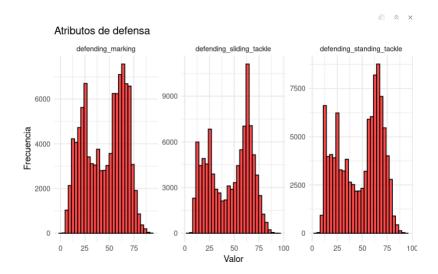
Incluyen atributos como la habilidad de centro, finalización y precisión en el cabeceo. Los histogramas generados son los siguientes:



- 1. **Attacking Crossing:** La habilidad de centro muestra que la mayoría de los jugadores tienen calificaciones entre 40 y 70. Esto sugiere que, en promedio, los jugadores tienen una capacidad moderada para realizar centros, mientras que pocos alcanzan niveles de élite en esta habilidad.
- 2. Attacking Finishing: La finalización presenta una distribución similar, concentrándose entre 40 y 70. Esto indica que muchos jugadores tienen una habilidad decente para finalizar jugadas, con una menor cantidad de jugadores en niveles más altos, probablemente los goleadores de élite.
- 3. **Attacking Heading Accuracy:** La precisión en el cabeceo tiene una distribución más sesgada, con la mayoría de los jugadores ubicados entre 40 y 70, pero también con un pico en valores bajos. Esto sugiere que, aunque algunos jugadores son hábiles en el juego aéreo, esta no es una habilidad tan común.

Atributos de defensa

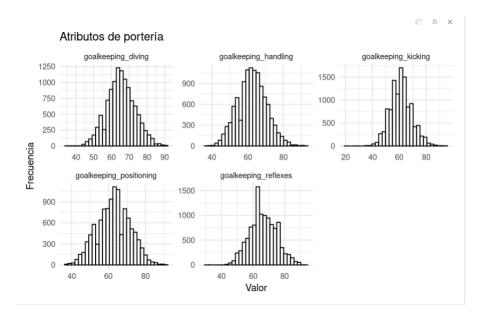
Incluyen atributos como la habilidad de marcaje, de barrida o entrada y de entrada en pie. Los histogramas generados son los siguientes:



- 1. **Defending Marking:** La habilidad de marcaje muestra una distribución amplia, con la mayoría de los jugadores en un rango de 25 a 75. Esto indica que muchos poseen habilidades de marcaje moderadas, mientras que son menos los que destacan o tienen un nivel muy bajo.
- 2. Defending Sliding Tackle: La habilidad de barrida o entrada también presenta una distribución amplia, con la mayoría de los jugadores concentrados entre 25 y 75. Se observa un pico notable en la zona media, lo que refleja que esta habilidad es común en niveles moderados.
- 3. **Defending Standing Tackle:** La habilidad de entrada en pie sigue un patrón similar a los otros atributos defensivos, con la mayoría de los jugadores en el rango de 25 a 75. La concentración en valores medios sugiere que, aunque muchos jugadores pueden realizar entradas en pie de forma efectiva, son pocos los que alcanzan un nivel de élite en esta área.

Atributos en portería

Incluyen atributos como la habilidad de estirada, la habilidad del manejo del balón, la patada, el posicionamiento y los reflejos. Los histogramas generados son los siguientes:



De forma general, varios futbolistas tienen habilidades de 0 o cercanas a 0 en los atributos de portería. Esto se debe a que son jugadores de campo, y estos atributos no son relevantes para ellos. Por esta razón, se filtraron únicamente los porteros para obtener un análisis más claro.

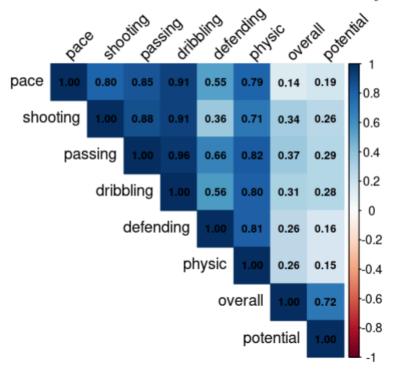
1. **Goalkeeping Diving:** La habilidad de estirada del portero tiene una distribución con una media en el rango de 60-70. Esto indica que los porteros suelen tener un buen desempeño en esta habilidad, aunque los valores oscilan entre 40 y 90.

- Goalkeeping Handling: La habilidad de manejo del balón muestra una media similar, cercana a 60-70, lo que sugiere que, en promedio, los porteros tienen un manejo sólido del balón.
- 3. **Goalkeeping Kicking:** La habilidad de patada del portero se distribuye de manera parecida, con valores centrados entre 50 y 70.
- 4. **Goalkeeping Positioning:** La habilidad de posicionamiento del portero sigue un patrón similar, lo que indica que la mayoría de los porteros tienen habilidades de posicionamiento bastante consistentes.
- Goalkeeping Reflexes: La habilidad de reflejos del portero muestra una fuerte concentración alrededor de 60-70, siendo esta habilidad clave para reaccionar a disparos inesperados.

Análisis de correlación

Se crea una matriz de correlación para ver cómo interactúan atributos de rendimiento y calificación.





La matriz de correlación analizada resalta las relaciones entre los atributos de rendimiento de los jugadores y sus calificaciones generales. Los puntos clave son los siguientes:

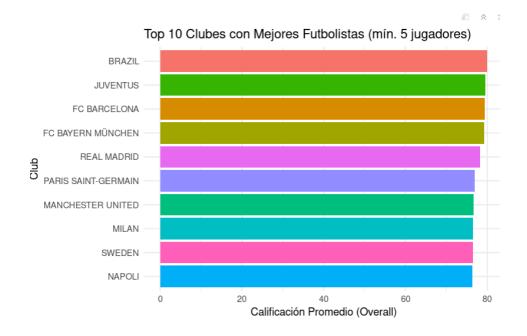
- Altas correlaciones entre atributos de rendimiento:
 - Passing y Dribbling tienen una correlación alta (0.96), lo que indica que los jugadores con buenos valores en pase suelen tener habilidades de dribbling destacadas.

- También se observan correlaciones elevadas entre Shooting y Passing (0.88), así
 como entre Shooting y Dribbling (0.91), lo que sugiere que los jugadores con
 buenas habilidades para disparar suelen ser hábiles en pase y regate.
- Correlación con la calificación general (Overall):
 - La calificación general (Overall) tiene correlaciones moderadas con atributos como Passing (0.37) y Dribbling (0.31). Esto indica que estos atributos contribuyen a la calificación general, aunque no son los únicos factores que influyen.
 - Potential muestra una correlación más fuerte con Overall (0.72), lo que sugiere que los jugadores con una calificación general alta tienden a tener un alto potencial de crecimiento.
- Baja correlación con el potencial (Potential):
 - El Potential tiene correlaciones bajas con atributos como Defending (0.16) y
 Physic (0.15). Esto indica que, aunque estos atributos tienen alguna influencia, el potencial de un jugador está determinado principalmente por factores no directamente relacionados con habilidades físicas o de juego.

En base a lo anterior podemos inferir que si se desea construir un equipo, se puede priorizar jugadores que tengan un buen equilibrio entre Passing y Dribbling si el objetivo es mantener una buena posesión del balón. Además de que la valoración de un futbolista no está completamente relacionada con las habilidades actuales, lo que sugiere que los jóvenes talentos pueden tener un gran potencial de crecimiento incluso si no sobresalen en algunos atributos específicos en el presente.

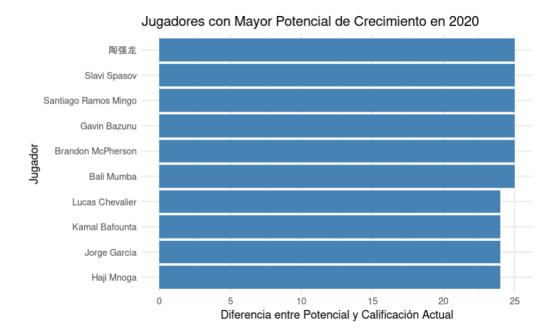
Preguntas y soluciones

1. ¿Cuáles son los jugadores con mejores futbolistas?



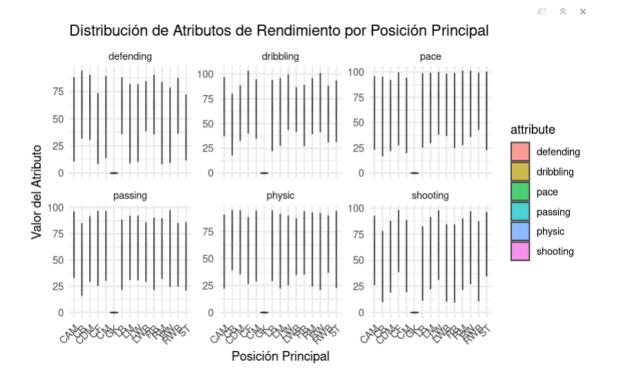
Podemos observar en la gráfica anterior que si un gamer desea escoger un club con los mejores jugadores en su plantilla puede escoger uno de los anteriores. Se resaltan clubes como Juventus, FC Barcelona, FC Bayern y Real Madrid, los cuales justamente coinciden con los clubes que en la vida real poseen los jugadores más desequilibrantes.

2. ¿Cuáles son los jugadores con mayor potencial de crecimiento?



Se puede concluir que existen jugadores con un potencial bastante interesante, de hasta 25 puntos. Estos jugadores serían idóneos para ser utilizados en modo carrera, ya que este modo le permite a los gamers desarrollar al jugador y explotar dicho potencial.

3. ¿Cómo se distribuyen los atributos de rendimiento clave para cada posición principal?



La gráfica anterior ilustra cómo se distribuyen los atributos de rendimiento (defending, dribbling, pace, passing, physic, shooting) en función de la posición principal de los jugadores. Se puede concluir lo siguiente:

Defending

- Los valores más altos en el atributo "defending" pertenecen a posiciones defensivas como defensa central (CB), lateral izquierdo (LB) y lateral derecho (RB).
- Las posiciones ofensivas, como delantero centro (ST) y extremo izquierdo (LW),
 tienen valores considerablemente más bajos, reflejando su menor rol defensivo.

Dribbling

- El atributo "dribbling" destaca en posiciones ofensivas como centrocampista ofensivo (CAM), extremo izquierdo (LW) y extremo derecho (RW), donde el control del balón es esencial para avanzar en el campo.
- Las posiciones defensivas tienden a mostrar valores más bajos en este atributo,
 consistente con su enfoque en tareas defensivas.

Pace

- Las posiciones de ataque y de banda, como extremos (LW, RW) y delanteros (ST), tienen los valores más altos en "pace," necesario para avanzar rápidamente hacia el área rival.
- En contraste, las posiciones centrales defensivas (CB) tienen menores valores, ya
 que su juego depende más del posicionamiento que de la velocidad.

Passing:

- Los mediocampistas, como centrocampistas centrales (CM) y ofensivos (CAM),
 tienen altos valores de "passing," ya que su función implica la creación y
 distribución del juego.
- Las posiciones defensivas y de ataque directo tienden a tener valores más bajos,
 dado que estas habilidades son menos críticas en su rol.

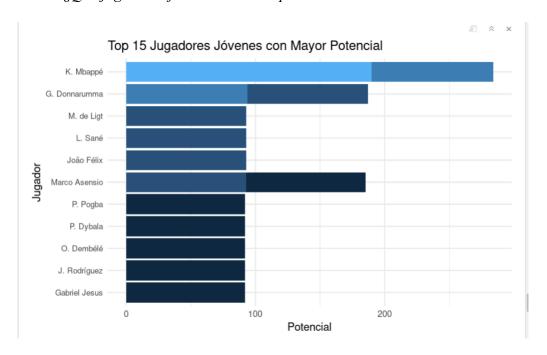
• Physic:

- Las posiciones defensivas (CB) y delanteros (ST) destacan en el atributo "physic,"
 donde la fuerza y resistencia física son esenciales.
- Por el contrario, posiciones técnicas como CAM y extremos (LW) tienen valores menores, reflejando un enfoque más técnico que físico.

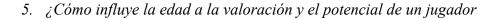
Shooting

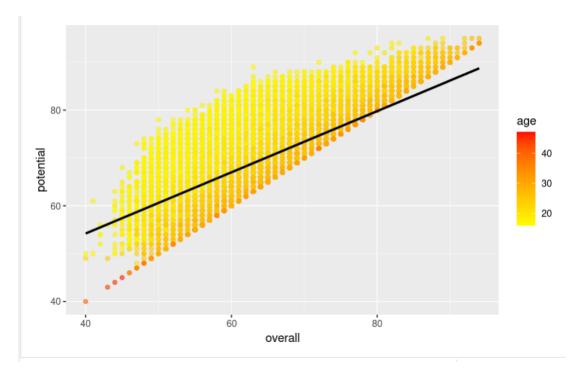
- Los delanteros (ST) sobresalen en "shooting," lo cual es vital para finalizar jugadas y anotar goles.
- Las posiciones defensivas y mediocampistas defensivos presentan los valores más bajos, debido a su menor participación en acciones de ataque.

4. ¿Qué jugadores jóvenes tienen el potencial más alto?



Entre los futbolistas menores de 23 años con mayor potencial podemos observar futbolistas consolidados en la realidad, como lo son el caso de Mbappé, Donnarumma, M. de Light entre otros. Por lo cual es una buena idea tomarlos en cuenta a la hora de armar un equipo.





En base al gráfico anterior se puede concluir lo siguiente:

- Relación entre Calificación Actual y Potencial:
 - Existe una relación positiva entre la calificación actual y el potencial de los jugadores. La gráfica nos indica que a medida que aumenta el overall de un jugador, también tiende a aumentar su potential.
 - La mayoría de los puntos se encuentran por encima de la línea de tendencia, lo que indica que muchos jugadores tienen un potencial superior a su calificación actual, especialmente los más jóvenes.
- Patrón según la Edad:
 - Los jugadores jóvenes (representados en amarillo) tienden a tener un potencial alto en comparación con su overall actual, lo que indica que tienen margen de mejora.

- Los jugadores de mayor edad (en tonos anaranjados y rojos) tienen su overall y
 potential más alineados, lo que sugiere que ya han alcanzado su máximo
 rendimiento o están cerca de él.
- También se observa que a medida que aumenta la edad, es menos común que el
 potencial sea significativamente mayor que el overall. Esto es un comportamiento
 esperado, ya que los jugadores mayores suelen haber alcanzado o superado su pico
 de rendimiento.
- Tomar Decisiones Basadas en el Ciclo de Vida del Jugador:
 - Los jugadores mayores con overall y potential similares pueden ser ideales para roles de liderazgo o para cubrir posiciones importantes a corto plazo. En cambio, los jugadores jóvenes con gran diferencia entre overall y potential pueden ser formados y mejorados para que se conviertan en futuras estrellas del equipo.

Conclusiones y recomendaciones

Además de lo expuesto en las preguntas y sus soluciones se puede concluir y recomendar lo siguiente:

Conclusiones

- Existe una alta correlación entre atributos de rendimiento clave como el pase, el dribbling
 y el disparo, lo que indica que futbolistas con altos valores en uno de estos atributos
 suelen tener también altos valores en los otros.
- La calificación general (overall) está moderadamente correlacionada con el potencial, lo
 que sugiere que los futbolistas con altas calificaciones actuales tienen más probabilidades
 de tener un alto potencial de mejora.
- Los futbolistas jóvenes tienen un margen de mejora significativo en comparación con los futbolistas de mayor edad, cuyo rendimiento y potencial tienden a estar más alineados.
- Los porteros tienen habilidades específicas que los diferencian significativamente de los futbolistas de campo, lo que resalta la necesidad de análisis separados para cada rol.

Recomendaciones

• Selección de jugadores:

 Para equipos enfocados en posesión, priorizar jugadores con altos valores en passing y dribbling.

• Construcción de equipos:

 Los clubes como FC Barcelona, Real Madrid y Juventus tienen plantillas con los mejores promedios generales, siendo ideales para gamers que buscan un equipo competitivo. Para equipos en formación, optar por jugadores jóvenes con alto potencial, ya que pueden ser entrenados para maximizar su rendimiento.

• Foco en atributos clave:

 En modo carrera, enfocar el entrenamiento en atributos con mayor impacto para cada posición, maximizando el crecimiento del jugador.