**Prueba Ingeniero de Datos**

Esta prueba está diseñada para conocer tus habilidades técnicas en análisis de datos. Junto a este documento se enviará una tabla con datos públicos del juego “fifa19”. Para mayor información sobre el dataset te puedes remitir a <https://www.kaggle.com/karangadiya/fifa19> y los nombres de las posiciones en el campo de juego se encuentran en <https://www.fifauteam.com/fifa-ultimate-team-positions-and-tactics/>. Estos datos no contienen errores.

**Caso:**

Actualmente la compañía en la que trabajas está fundando un equipo de futbol con el ánimo de comprar jugadores baratos y venderlos caros. Investigando los datos disponibles, se encontró que los de fifa19 reflejan muy bien las características reales de los jugadores, además de tener el valor y salario actual de cada jugador. Tu eres el científico de datos y te piden hacer un análisis que se divide en 2 partes:

1. Encontrar los jugadores que actualmente estén sub-valorados en el mercado, para comprarlos hoy y venderlos en 2 años.
   1. Usa las variables Value y Wage para entender el costo del jugador.
   2. Usa las características de juego (Crossing, Finishing, HeadingAccuracy…), edad, posición principal y cualquier otra variable que consideres relevante, como variables explicativas.
   3. Encuentra los jugadores sub-valorados en el mercado.
   4. Construye un modelo de Machine Learning para estimar el valor de venta del jugador
   5. Da una sugerencia de los jugadores a contratar para el club

Nota: Para simplificar el ejercicio ten como supuesto se paga el valor de Value y no el de Release Clause. En el mercado no existe inflación, ni cambios salariales y cualquier jugador está dispuesto a aceptar las ofertas.

Para el desarrollo de alguna metodología de Machine Learning debes usar R o Python y el procedimiento debe ser entendible para el calificador. Se evaluará tanto la presentación de las conclusiones como la construcción del modelo.

No se espera que logres llegar a un entregable óptimo, pero sí que demuestres tu capacidad de análisis, conocimientos en herramientas y metodologías de análisis de datos. La prueba está pensada para resolverse de 2 a 3 horas.

Confírmanos la fecha de entrega de la prueba por favor.