Proyecto lesiones en jugadores de futbol universitarios

El origen de los datos que utilizaremos para el proyecto son provenientes de un dataset en Kaggle.Vienen en formato csv(Datos estructurados).

Nuestros dataset relaciona variables de los jugadores,tales como:

-Edad -Altura -Peso -Posicion -Horas de entrenamiento por semana -Partidos jugados la temporada pasada -Numero de lesiones previas -Puntuacion de fuerza de rodilla -Puntuacion de flexibilidad de isquiotibiales -Tiempo de reaccion -Puntuacion en la prueba de equilibirio -Puntuacion de agilidad -Velocidad en sprint de 10 m/s -Promedio de horas de sueño por noche -Puntuacion de nivel de stress -Puntuacion de la calidad de alimentacion -Porcentaje de aderencia al calentamiento -Lesion en la proxima temporada -Indice de masa corporal

```
import pandas as pd
import os

tabla_datos=pd.read_csv("data.csv")
tabla_datos.head()
```

```
   Age  Height_cm  Weight_kg    Position  Training_Hours_Per_Week  \
0   22        173         64  Midfielder                11.575308
1   18        170         67  Midfielder                12.275869
2   22        186         75     Forward                12.254896
3   20        172         62    Defender                 9.006678
4   18        172         94  Midfielder                12.683668

   Matches_Played_Past_Season  Previous_Injury_Count
Knee_Strength_Score  \
0                          36                      1
77.460279
1                          37                      2
72.634442
2                          12                      2
77.064490
3                          11                      1
82.810232
4                          10                      2
76.772859

   Hamstring_Flexibility  Reaction_Time_ms  Balance_Test_Score  \
0              79.115738        284.487853           91.212476
```

```
1               82.541688        250.579249            87.294078
2               75.943631        269.119918            83.440688
3               73.878324        226.376412            87.591894
4               76.653043        229.021042            83.125161

   Sprint_Speed_10m_s  Agility_Score  Sleep_Hours_Per_Night  \
0            5.874630      77.599705               8.238293
1            5.796269      94.418987               8.983737
2            5.731209      70.179176               7.229193
3            6.220212      83.473824               7.681029
4            5.385958      87.037256               6.728091

   Stress_Level_Score  Nutrition_Quality_Score
Warmup_Routine_Adherence  \
0           46.616415                81.472206
1
1           49.368037                81.056677
1
2           43.132808                64.877457
0
3           51.528529                89.824744
1
4           52.379718                71.569197
0

   Injury_Next_Season        BMI
0                   0  21.383942
1                   0  23.183391
2                   1  21.678807
3                   0  20.957274
4                   1  31.773932

tabla_datos.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 800 entries, 0 to 799
Data columns (total 19 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         800 non-null    int64
 1   Height_cm                   800 non-null    int64
 2   Weight_kg                   800 non-null    int64
 3   Position                    800 non-null    object
 4   Training_Hours_Per_Week     800 non-null    float64
 5   Matches_Played_Past_Season  800 non-null    int64
 6   Previous_Injury_Count       800 non-null    int64
 7   Knee_Strength_Score         800 non-null    float64
 8   Hamstring_Flexibility       800 non-null    float64
 9   Reaction_Time_ms            800 non-null    float64
 10  Balance_Test_Score          800 non-null    float64
```

```
 11  Sprint_Speed_10m_s          800 non-null    float64
 12  Agility_Score               800 non-null    float64
 13  Sleep_Hours_Per_Night       800 non-null    float64
 14  Stress_Level_Score          800 non-null    float64
 15  Nutrition_Quality_Score     800 non-null    float64
 16  Warmup_Routine_Adherence    800 non-null    int64
 17  Injury_Next_Season          800 non-null    int64
 18  BMI                         800 non-null    float64
dtypes: float64(11), int64(7), object(1)
memory usage: 118.9+ KB
```

```
print(f"Nuestra fuennte de datos cuenta con 800 datos por columna ,por lo que como nuestro Datafream tiene 19 columnas el volumen de datos es de {19*800} datos ")
```

```
Nuestra fuennte de datos cuenta con 800 datos por columna ,por lo que como nuestro Datafream tiene 19 columnas el volumen de datos es de 15200 datos
```

```
tabla_datos.describe()
```

```
                Age   Height_cm   Weight_kg  Training_Hours_Per_Week  \
count  800.000000  800.000000  800.000000               800.000000
mean    21.135000  177.407500   73.235000                 9.951150
std      1.991037    7.148974    9.929276                 2.610395
min     18.000000  154.000000   45.000000                 5.000000
25%     19.000000  173.000000   66.000000                 8.127151
50%     21.000000  177.000000   73.000000                 9.895710
75%     23.000000  182.000000   80.000000                11.535140
max     24.000000  200.000000  105.000000                18.866608

       Matches_Played_Past_Season  Previous_Injury_Count
Knee_Strength_Score  \
count                  800.000000             800.000000
800.000000
mean                    22.332500               1.536250
74.933249
std                     10.311516               1.292584
6.672704
min                      5.000000               0.000000
52.391351
25%                     13.000000               1.000000
70.432656
50%                     22.000000               1.000000
74.997933
75%                     32.000000               2.000000
79.632391
max                     39.000000               8.000000
93.900051
```

```
       Hamstring_Flexibility  Reaction_Time_ms  Balance_Test_Score  \
count             800.000000        800.000000          800.000000
mean               79.154123        249.423244           83.832337
std                 6.782332         22.532387            6.931657
min                58.180381        180.000000           60.059484
25%                74.495959        234.089585           79.044910
50%                79.187909        249.127328           84.156236
75%                83.813179        265.105082           88.877902
max               100.000000        306.730851          100.000000

       Sprint_Speed_10m_s  Agility_Score  Sleep_Hours_Per_Night  \
count          800.000000     800.000000             800.000000
mean             5.949025      78.341311               7.417124
std              0.329133       8.775418               0.793183
min              4.862435      50.000000               5.000000
25%              5.732552      72.675392               6.850062
50%              5.937692      78.340973               7.424618
75%              6.159205      84.093489               7.988100
max              6.898228     100.000000               9.860553

       Stress_Level_Score  Nutrition_Quality_Score
Warmup_Routine_Adherence  \
count          800.000000               800.000000
800.000000
mean            54.039342                74.382174
0.597500
std             11.421143                 9.324899
0.490708
min             21.561186                50.000000
0.000000
25%             45.775371                67.809084
0.000000
50%             54.047208                74.363878
1.000000
75%             61.910804                80.502613
1.000000
max             87.065007               100.000000
1.000000

       Injury_Next_Season         BMI
count          800.000000  800.000000
mean             0.500000   23.377364
std              0.500313    3.673279
min              0.000000   14.346326
25%              0.000000   20.786644
50%              0.500000   23.130560
75%              1.000000   26.024182
max              1.000000   36.262439
```

A continuacion se presentaran los tipos de datos que contiene cada columna del datafream

```
tabla_datos.dtypes
```

```
Age                             int64
Height_cm                       int64
Weight_kg                       int64
Position                       object
Training_Hours_Per_Week       float64
Matches_Played_Past_Season      int64
Previous_Injury_Count           int64
Knee_Strength_Score           float64
Hamstring_Flexibility         float64
Reaction_Time_ms              float64
Balance_Test_Score            float64
Sprint_Speed_10m_s            float64
Agility_Score                 float64
Sleep_Hours_Per_Night         float64
Stress_Level_Score            float64
Nutrition_Quality_Score       float64
Warmup_Routine_Adherence        int64
Injury_Next_Season              int64
BMI                           float64
dtype: object
```