**📑 Informe de Preprocesamiento y Análisis**

**1. Estadísticos descriptivos**

* En **entrenamientos** se observaron valores amplios en métricas como distancia\_total, carga\_explosiva, carga\_sostenida y carga\_regenerativa, con máximos muy elevados respecto a los promedios.
* En **partidos** las variables mantienen un rango más homogéneo, aunque se destacan altas cargas en carga\_regenerativa y rendimiento\_partido.

**2. Correlaciones**

**ENTRENAMIENTOS PARTIDOS**

**Imagen que contiene Texto

El contenido generado por IA puede ser incorrecto. Imagen que contiene Tabla

El contenido generado por IA puede ser incorrecto.**

* Se detectaron **correlaciones superiores al 90%** entre varias métricas, especialmente:
  + hmlD\_m con carga\_sostenida (0.99).
  + hsr\_abs\_m con carga\_explosiva (0.97).
  + distancia\_total con carga\_regenerativa (0.96).
* Esto indica **redundancia**: muchas variables representan la misma dimensión del esfuerzo físico.
* Se recomienda priorizar **carga\_explosiva, carga\_sostenida, carga\_regenerativa y rendimiento\_total/diario**, ya que son más interpretables y reducen la multicolinealidad.

**3. Outliers**

**Entrenamientos**

**Imagen que contiene Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.**

* Variables con más outliers:
  + hsr\_rel\_m (7.7%), carga\_regenerativa (7.5%), distancia\_total (5.6%).
* Posibles causas: sesiones regenerativas específicas, jugadores con perfiles físicos distintos o errores de medición.

**Partidos**

**Imagen que contiene Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.**

* Principales outliers en:
  + velocidad\_prom\_m\_min (7.2%), sprints\_distancia\_m y sprints\_cantidad (3.2%).
* Interpretación: diferencias de posición y estilo de juego (ej. extremos con mayor velocidad y sprints vs. defensores con menos exigencia explosiva).

**4. Interpretación general**

* En entrenamientos, los outliers marcan **variaciones por tipo de sesión** (explosiva, regenerativa, aeróbica).
* En partidos, los outliers señalan **roles de jugadores** o picos de rendimiento.
* Mantener estas diferencias puede ser útil para el modelo, ya que los casos extremos ayudan a identificar riesgos de fatiga o patrones de lesión.

**5. Pasos a seguir para el modelo de ML**

1. **Depurar el dataset** → conservar las variables principales y descartar redundancias.
2. **Tratar los outliers** → escalado robusto o winsorizing para que no dominen el modelo.
3. **Entrenar modelo base** → Random Forest o XGBoost con las 4 variables principales.
4. **Analizar importancia de variables** → verificar qué métrica aporta más a la predicción.
5. **Iterar y comparar** → modelos con todas las variables (aplicando PCA o selección automática) vs. set reducido.