

EVALUACIÓN FINAL: ANÁLISIS DE DATOS DE ATLETAS OLÍMPICOS

El Comité Olímpico Internacional ha recopilado información sobre el desempeño de atletas en diversas disciplinas a lo largo de los últimos juegos olímpicos. El objetivo es analizar patrones de rendimiento, identificar factores clave que influyen en el éxito de los atletas y realizar predicciones sobre el desempeño futuro.

El equipo de análisis de datos necesita que explores, transformes y modelas los datos para extraer información relevante sobre la relación entre características de los atletas (edad, altura, peso, entrenamientos semanales, país) y su desempeño en las competiciones.

Objetivo

Evaluar la capacidad del estudiante para aplicar análisis exploratorio de datos, estadística descriptiva, correlación, regresión lineal y visualización de datos mediante Python y las librerías Pandas, Seaborn, Matplotlib y Statsmodels, en el contexto del desempeño de atletas olímpicos en distintas disciplinas deportivas.

INSTRUCCIONES

1. Análisis Exploratorio de Datos (2 puntos)

- Carga el dataset en un DataFrame de Pandas.
- Muestra las primeras 5 filas y usa `.info()` para obtener información sobre los datos.
- Calcula estadísticas descriptivas con `.describe()`.
- Genera un histograma del número de entrenamientos semanales.

2. Estadística Descriptiva (2 puntos)

- Determina el tipo de variable de cada columna.
- Calcula la media, mediana y moda de la cantidad de medallas obtenidas.
- Calcula la desviación estándar de la altura de los atletas.
- Identifica valores atípicos en la columna de peso utilizando un boxplot.

3. Análisis de Correlación (2 puntos)

- Calcula la correlación de Pearson entre entrenamientos semanales y medallas totales.
- Crea un gráfico de dispersión (scatterplot) entre peso y medallas totales con Seaborn.

- Explica si existe correlación entre estas variables.

4. Regresión Lineal (2 puntos)

- Implementa un modelo de regresión lineal para predecir el número de medallas obtenidas en función del número de entrenamientos semanales.
- Obtén los coeficientes de regresión e interpreta el resultado.
- Calcula el R^2 para medir el ajuste del modelo.
- Usa Seaborn (regplot) para graficar la regresión lineal.

5. Visualización de Datos con Seaborn y Matplotlib (2 puntos)

- Crea un heatmap de correlación entre todas las variables numéricas.
- Crea un boxplot de la cantidad de medallas por disciplina deportiva.
- Personaliza los gráficos con títulos, etiquetas y colores.

INSTRUCCIONES ADICIONALES:

- Puntos totales = 10.
- Descarga el material complementario olimpicos.csv
- Comprimir el archivo completo en formato .zip o .rar.
- Sube el archivo a la plataforma.