**#codigo**

Texto

Descripción generada automáticamente  
  
**#Resultado**

**Análisis de Datos y Modelo de Regresión Lineal**

**Introducción**

En este análisis, exploramos cómo el ingreso y el nivel socioeconómico influyen en los puntajes de lectura crítica de los estudiantes en las "Pruebas Saber". Luego, desarrollamos un modelo de regresión lineal para predecir estos puntajes.

**Datos**

* **Variables:**
  + Período de la prueba.
  + Puntaje de lectura crítica.
  + Ingreso individual del estudiante.
  + Nivel socioeconómico individual del estudiante.

**Exploración de Datos**

Observamos las relaciones entre las variables y descubrimos que el ingreso y el nivel socioeconómico están asociados con los puntajes de lectura crítica.

**Preparación de Datos**

Eliminamos columnas irrelevantes y dividimos los datos en conjuntos de entrenamiento y prueba.

**Modelo de Regresión Lineal**

Entrenamos un modelo y evaluamos su rendimiento utilizando el Error Cuadrático Medio (MSE).

* **Errores MSE:**
  + Entrenamiento: 942.98
  + Prueba: 961.38

**Interpretación**

* El ingreso tiene un impacto positivo en los puntajes.
* El nivel socioeconómico tiene un impacto negativo.
* El período de la prueba parece tener poca relevancia.

**Conclusiones**

El modelo nos ayuda a comprender cómo el ingreso y el nivel socioeconómico afectan los puntajes de lectura crítica. Sin embargo, recordemos que hay otros factores que también pueden influir en estos puntajes.  
  
Gráfico, Histograma

Descripción generada automáticamente  
Imagen que contiene Gráfico

Descripción generada automáticamente

**#Resultados en crudo.**

**Unnamed: 0 PERIODO MOD\_LECTURA\_CRITICA\_PUNT ESTU\_INSE\_INDIVIDUAL ESTU\_NSE\_INDIVIDUAL**

**0 0 20231 164 61.403944 4.0**

**1 1 20231 132 53.902622 2.0**

**2 2 20231 150 65.881886 4.0**

**3 3 20231 141 71.978823 4.0**

**4 4 20231 147 67.067764 4.0**

**<class 'pandas.core.frame.DataFrame'>**

**RangeIndex: 139288 entries, 0 to 139287**

**Data columns (total 5 columns):**

**# Column Non-Null Count Dtype**

**--- ------ -------------- -----**

**0 Unnamed: 0 139288 non-null int64**

**1 PERIODO 139288 non-null int64**

**2 MOD\_LECTURA\_CRITICA\_PUNT 139288 non-null int64**

**3 ESTU\_INSE\_INDIVIDUAL 139288 non-null float64**

**4 ESTU\_NSE\_INDIVIDUAL 139288 non-null float64**

**dtypes: float64(2), int64(3)**

**memory usage: 5.3 MB**

**None**

**Unnamed: 0 PERIODO MOD\_LECTURA\_CRITICA\_PUNT ESTU\_INSE\_INDIVIDUAL ESTU\_NSE\_INDIVIDUAL**

**count 139288.000000 139288.000000 139288.000000 139288.000000 139288.000000**

**mean 69643.500000 20231.015026 145.865107 55.023014 2.652569**

**std 40209.126485 0.121658 31.279078 9.063076 1.098924**

**min 0.000000 20231.000000 0.000000 20.235238 1.000000**

**25% 34821.750000 20231.000000 123.000000 48.952170 2.000000**

**50% 69643.500000 20231.000000 145.000000 55.275513 2.000000**

**75% 104465.250000 20231.000000 168.000000 61.255557 4.000000**

**max 139287.000000 20232.000000 300.000000 87.275103 4.000000**

**Forma de X\_train: (111430, 4)**

**Forma de X\_test: (27858, 4)**

**Forma de y\_train: (111430,)**

**Forma de y\_test: (27858,)**

**Error cuadrático medio en conjunto de entrenamiento: 942.980527077797**

**Error cuadrático medio en conjunto de prueba: 961.3774688239469**