EDWIN GIOVANNY CUENCA MORENO

GESTION DE DATOS

PUJ – MAESTRIA IOT

PARCIAL 1

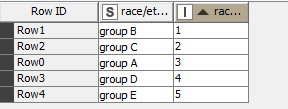
1. ¿Qué tipo de variables tiene el dataset? Detalle el tipo de variable de cada columna.

* Género: Cualitativa Ordinal
* Raza / Etnia: Cualitativa Ordinal
* Nivel de escolaridad de los padres: Cualitativa Ordinal
* Tipo de Alimentación: Cualitativa Ordinal
* Preparo el Examen: Cualitativa Ordinal
* Puntaje Matemáticas: Cuantitativa Discreto
* Puntaje Lectura: Cuantitativa Discreto
* Puntaje Escritura: Cuantitativa Discreto

1. ¿Qué tipo de problemas de calidad de datos logra identificar? Defina e implemente las estrategias de limpieza de datos que correspondan.

Se valido que no existen errores de formato o de escritura para cada una de las variables cargadas en el dataset, fue posible cargar sin ningún inconveniente dichos archivos en el software Knime. Con el fin de poder realizar un análisis un poco más practico, se procedió a realizar el cambio de las variables de tipo cualitativo a cuantitativo para “Raza /Etnia” y “Nivel de Escolaridad”.

* Normalización Raza / Etnia



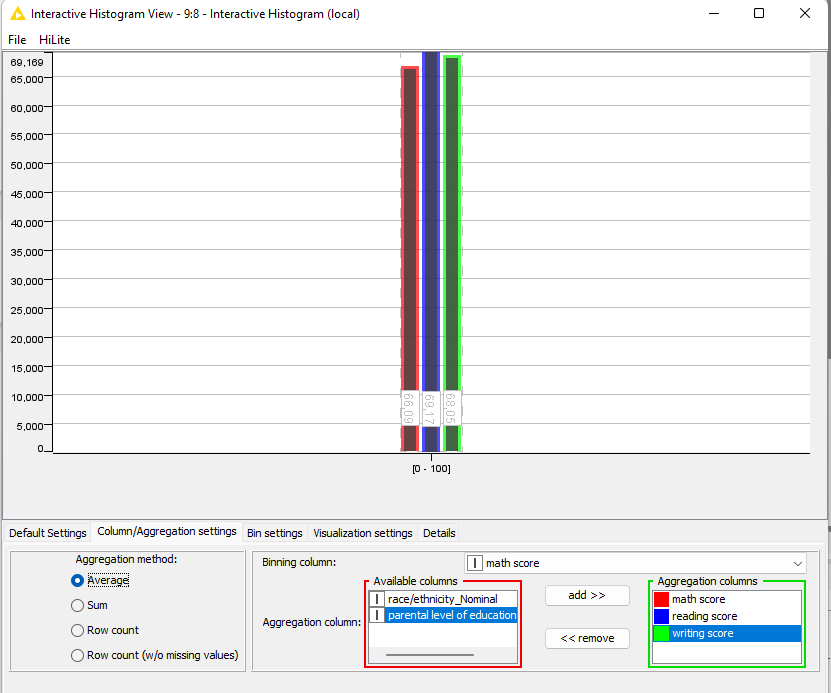
* Normalización Niveles de Escolaridad

Texto, Tabla

Descripción generada automáticamente

1. ¿En qué asignatura en promedio los estudiantes obtuvieron un mejor puntaje? ¿Hay evidencia de algún sesgo en la distribución de dichos puntajes?

La asignatura en la cual los estudiantes obtuvieron el mayor puntaje fue lectura con un promedio de 69.17, en comparación con escritura 68.05 y matemáticas 66.09:



1. ¿Hay alguna diferencia observable en los puntajes de la asignatura de matemáticas entre géneros? ¿Qué género obtuvo en promedio los mejores puntajes?

Para los puntajes de matemáticas se evidencia que el género masculino obtuvo un puntaje superior al del género femenino de 68.73 y 63.63 respectivamente. El genero que obtuvo los mayores puntajes fue el genero femenino con un porcentaje de 69.57 respecto al genero masculino de 65.83

Gráfico

Descripción generada automáticamente con confianza media