

Reporte de proyecto de python

Estudiante: Guillermo Cáceres

Profesor: Joel Orellana

Conjunto de datos utilizado:

StudentsPerformance.csv

Link a kaggle:

<https://www.kaggle.com/datasets/spscientist/students-performance-in-exams/data>

Descripción

Este proyecto analiza un conjunto de datos de estudiantes ('StudentsPerformance.csv') para comprender cómo diferentes factores como el género, grupo étnico, tipo de almuerzo y curso de preparación afectan el rendimiento académico en matemáticas, lectura y escritura.

1. Analisis Exploratorio EDA

- Se usó `ydata-profiling` para explorar el dataset y entender su estructura.
- Se verificó que no existen valores nulos.
- Se analizaron distribuciones, correlaciones y características generales.

2. Limpieza de Datos

- Se validó que no había valores faltantes.
- Se revisaron los tipos de datos.
- Se transformaron algunas variables categóricas.

3. Visualización

- Histogramas de las calificaciones.
- Comparación de calificaciones por género.
- Relación entre curso de preparación y desempeño.
- Comparaciones por grupo étnico.
- Relación entre almuerzo y desempeño.

Observaciones clave

- Se observa en el gráfico que las mujeres tienden a tener mejores calificaciones en materias de escritura y lectura, mientras que los hombres tienen calificaciones más altas en matemáticas, aunque la diferencia no es significativa.
- También se observa que la mayoría de las calificaciones se encuentran en el intervalo de entre 6 y 8.5 aproximadamente, sin embargo también hay datos con calificaciones más bajas o más altas.
- Observamos también que los que tienen el curso de preparación completado obtuvieron mejores calificaciones.
- Y a pesar de que hay ciertas diferencias entre las calificaciones según grupo étnico, las diferencias no son drásticas.
- Se observa que aquellos estudiantes con un almuerzo standard obtienen mejores calificaciones.

Conclusiones

1. El análisis evidencia que factores externos (como preparación o almuerzo) pueden incidir en el rendimiento escolar.
2. Las visualizaciones refuerzan la idea de que el apoyo educativo adicional (preparación) mejora el desempeño.
3. Las visualizaciones también refuerzan la idea de que una buena alimentación también juega un papel importante en el desempeño de los estudiantes.

4. Este análisis podría usarse para identificar áreas de mejora en programas escolares.

Próximos pasos

1. Aplicar modelos predictivos para predecir el rendimiento.
2. Ampliar el análisis con otros datasets relacionados a educación.
3. Encontrar otra fuente de datos o ampliar ciertas variables como preparación o almuerzo, para que den información más específica.