

ANALISIS DEMOGRÁFICO Y ACADÉMICO DE ESTUDIANTES

NOMBRE CURSO/ASIGNATURA:

ELECTIVA DICIPLINAR I

AUTORES:

JULIO BONIFACIO

JORGE BLANQUICETT

VALENTINA BLANDON

LUISA CHACÓN

EDGAR CERVANTES

GEOVANY MEDRANO

UNIVERSIDAD DEL SINÚ

18/03/24

INTRODUCCION

Este proyecto se enfoca en analizar un conjunto de datos de estudiantes universitarios para comprender las dinámicas demográficas, académicas y de preferencias educativas. A través de la recopilación y análisis de variables como género, edad, carrera universitaria, asistencias, preferencias de recursos educativos, semestre, promedio y nivel socioeconómico, se busca identificar patrones y tendencias que puedan influir en las políticas educativas y estrategias de apoyo estudiantil.

Es importante destacar que la data presentada en este análisis se basa en hipótesis y datos recopilados de fuentes disponibles al momento de la investigación. Si bien se ha realizado un esfuerzo exhaustivo para garantizar la precisión y fiabilidad de la información, es crucial reconocer que estos resultados no son oficiales y deben ser utilizados con respeto y discreción.

ANALISIS DE LA DATA

| | |
|--------------------------------|---------|
| Estudiantes | object |
| ID | int64 |
| genero | object |
| edad | int64 |
| carrera Universitaria | object |
| Asistencias | object |
| Recursos educativos preferidos | object |
| Unnamed: 7 | float64 |
| Semestre | object |
| Promedio | float64 |
| Nivel socioeconomico | object |
| | float64 |

TIPOS DE DATOS

Los datos object. Indican que son variables categóricas con valores de tipo texto.

Los datos int64. Indican que son variables numéricas enteras de 64 bits.

Y los datos float64. Indican que son variables numéricas de punto flotante de 64 bits.

Análisis:

Con este análisis podemos darnos cuenta de que:

1. La mayoría de las variables son de tipo categórico, lo que indica que se trata de datos cualitativos.
2. Hay algunas variables numéricas, como la edad, el ID, el promedio y las dos variables sin nombre, que se pueden utilizar para realizar análisis cuantitativos.
3. La variable "Asistencias" es de tipo object, pero es probable que se trate de una variable numérica que representa la cantidad de asistencias a clases de un estudiante.

| Index | ID | edad | Unnamed: 7 | Promedio | |
|-------|--------------------|--------------------|------------|--------------------|-----|
| count | 44.0 | 44.0 | 0.0 | 44.0 | 0.0 |
| mean | 51262.068181818184 | 20.704545454545453 | NaN | 3.001628680910123 | NaN |
| std | 36006.192155056764 | 1.836495766335701 | NaN | 1.064293834664962 | NaN |
| min | 9758.0 | 19.0 | NaN | 0.2 | NaN |
| 25% | 15920.25 | 20.0 | NaN | 2.327839445408255 | NaN |
| 50% | 37118.5 | 20.0 | NaN | 2.980921363062465 | NaN |
| 75% | 90863.25 | 21.0 | NaN | 3.7992212086375177 | NaN |
| max | 99004.0 | 26.0 | NaN | 4.89208940505698 | NaN |

Gracias a la información de la tabla probablemente contiene 44 filas (basado en el valor count para la mayoría de las columnas). ¿Qué observaciones específicas encontramos gracias a los resultados?

ANÁLISIS POR COLUMNA:

Edad:

¹ Lider y encargado de buscar, limpiar estandarizar y perfeccionar la data para el análisis: Luisa Chacon

Tipo de dato: int64

Descripción: Edad de los estudiantes.

Estadísticas:

- count: 44 (No hay valores nulos)
- mean: 20.704545 años (Promedio de edad)
- std: 1.636496 años (Desviación estándar de la edad)
- min: 19 años (Edad mínima)
- 25%: 20 años (Percentil 25)
- 50%: 20 años (Mediana)
- 75%: 21 años (Percentil 75)
- max: 26 años (Edad máxima)

Unnamed: 7:

Tipo de dato: float64

Descripción: Parece ser una columna sin nombre.

Estadísticas:

- count: 0 (Todos los valores son nulos)
- mean: NaN (No se puede calcular el promedio)
- std: NaN (No se puede calcular la desviación estándar)
- min: NaN (No se puede calcular el valor mínimo)
- 25%: NaN (No se puede calcular el percentil 25)
- 50%: NaN (No se puede calcular la mediana)

- 75%: NaN (No se puede calcular el percentil 75)
- max: NaN (No se puede calcular el valor máximo)

Promedio:

Tipo de dato: float64

Descripción: Promedio académico de los estudiantes.

Estadísticas:

- count: 44 (No hay valores nulos)
- mean: 3.0 (Promedio general)
- std: 1.064 (Desviación estándar del promedio)
- min: 0.2 (Promedio mínimo)
- 25%: 2.32 (Percentil 25)
- 50%: 2.98 (Mediana)
- 75%: 3.79 (Percentil 75)
- max: 4.89 (Promedio máximo)

Observaciones:

- En este caso gracias a el análisis podemos darnos cuenta de que la columna Unnamed: 7 parece no contener datos relevantes, ya que todos sus valores son nulos.
- La columna Promedio tiene un valor medio de 3.318, lo que podría ser un indicador del rendimiento académico en el curso evaluado (según que se trata de calificaciones numéricas en un rango de 0 a 5).

| # | Column | Non-Null Count | Dtype |
|----|--------------------------------|----------------|---------|
| 0 | Estudiantes | 44 non-null | object |
| 1 | ID | 44 non-null | int64 |
| 2 | genero | 44 non-null | object |
| 3 | edad | 44 non-null | int64 |
| 4 | carrera Universitria | 44 non-null | object |
| 5 | Asistencias | 44 non-null | object |
| 6 | Recursos educativos preferidos | 44 non-null | object |
| 7 | Unnamed: 7 | 0 non-null | float64 |
| 8 | Semestre | 44 non-null | object |
| 9 | Promedio | 44 non-null | float64 |
| 10 | Nivel Socioeconomico | 44 non-null | object |
| 11 | | 0 non-null | float64 |

dtypes: float64(3), int64(2), object(7)

Si observamos la imagen podemos darnos cuenta de que esta data cuenta con un total de 12 columnas en las cuales tenemos que El tipo de dato de la columna ID es int64, lo que significa que los valores son números enteros, El tipo de dato de las columnas género, carrera Universitaria, Asistencias, Semestre, Unnamed: 7 y Nivel socioeconomico que es object, lo que significa que los valores son cadenas de texto. Pero que Unnamed: 7 parece no tener datos relevantes ya que todos sus valores son nulos, El tipo de dato de las columnas edad y Promedio es float64, lo que significa que los valores son números decimales.

También pudimos darnos cuenta que:

La columna 12 no tiene nombre y está vacía. El tipo de dato de la columna es float64, lo que significa que los valores deberían ser números decimales. Sin embargo, la columna está vacía, por lo que no hay valores que analizar.

Posibles explicaciones:

- La columna 11 se reservó para datos futuros que aún no se han recopilado.

- La columna 11 se eliminó accidentalmente o se dejó vacía por error.
- La columna 11 no se utiliza en el análisis y se puede eliminar.

Tomamos la decisión de eliminar estas columnas. ¿Por qué?

Columna 11:

- Tipo de dato: float64 (Número decimal)
- Descripción: Vacía, sin valores relevantes.
- Motivo para eliminar: No aporta información útil al análisis.
- Efecto en la data: Ninguno, ya que no hay valores que eliminar.

Columna Unnamed: 7:

Tipo de dato: float64 (Número decimal)

- Descripción: Vacía, sin valores relevantes.
- Motivo para eliminar: No aporta información útil al análisis.²
- Efecto en la data: Ninguno, ya que no hay valores que eliminar.

Tomamos la decisión de eliminar las columnas 11 y Unnamed: 7 ya que no afectará la data de ninguna manera, ya que ambas están vacías y no contienen información relevante. Eliminarlas puede mejorar la claridad y legibilidad de la Data ya que solo se mostrarán las columnas que son relevantes para el análisis.

Beneficios de eliminar las columnas:

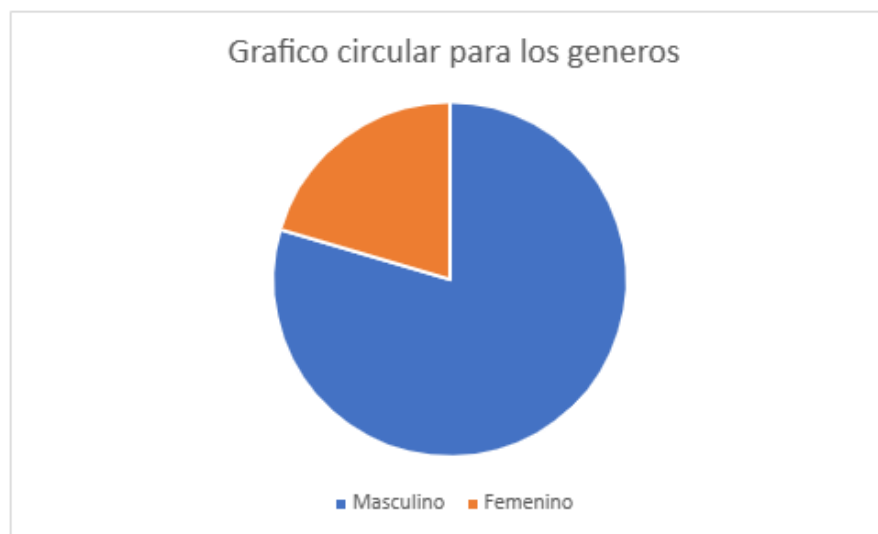
² Encargados del análisis Geovany Medrado y Julio Bonifacio

- Mejora la claridad y legibilidad de la Data.
- Reduce el tamaño de la Data, lo que puede mejorar el rendimiento.
- Evita confusiones al interpretar los datos.

TABLAS Y GRÁFICOS QUE LO DEMUESTRAN

| Genero | Frecuencia |
|-----------|------------|
| Masculino | 35 |
| Femenino | 9 |

Tabla 1 Tabla de frecuencia para los generos



Basados en este analisis llegamos a la conclusión que el género que predomina en este curso es el masculino con un 80%.

| asistencias | Frecuencia |
|-------------|------------|
| regular | 22 |
| irregular | 22 |

Tabla 2 Tabla de frecuencia para las asistencias



Para las asistencias llegamos a la conclusión que la mitad de los estudiantes son tanto regulares como irregulares esto puede deberse a su nivel socioeconómico o por otras causas

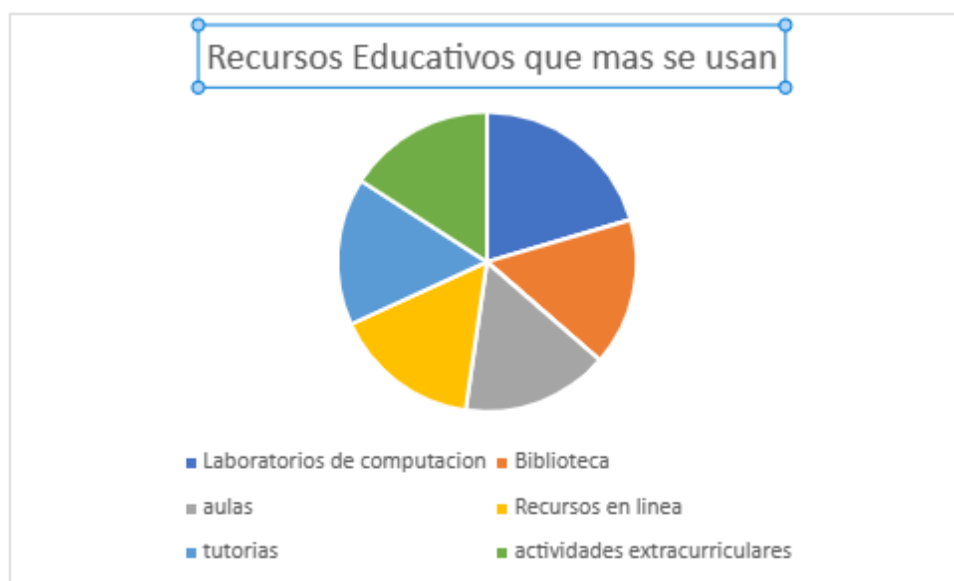
| Nivel Socioeconómico | Frecuencia |
|----------------------|------------|
| bajo | 14 |
| medio | 24 |
| alto | 6 |

Tabla 3 Tabla de frecuencia para Nivel socioeconómico



| Recursos educativos preferidos | Frecuencia |
|--------------------------------|------------|
| Laboratorios de computacion | 9 |
| Biblioteca | 7 |
| aulas | 7 |
| Recursos en linea | 7 |
| tutorias | 7 |
| actividades extracurriculares | 7 |

Tabla 4 Tabla de frecuencia para Recursos educativos preferidos de los estudiantes.



| Semestre | Frecuencia |
|----------|------------|
| IV | 34 |
| V | 10 |

Tabla 5 Tabla de frecuencia para los semestres



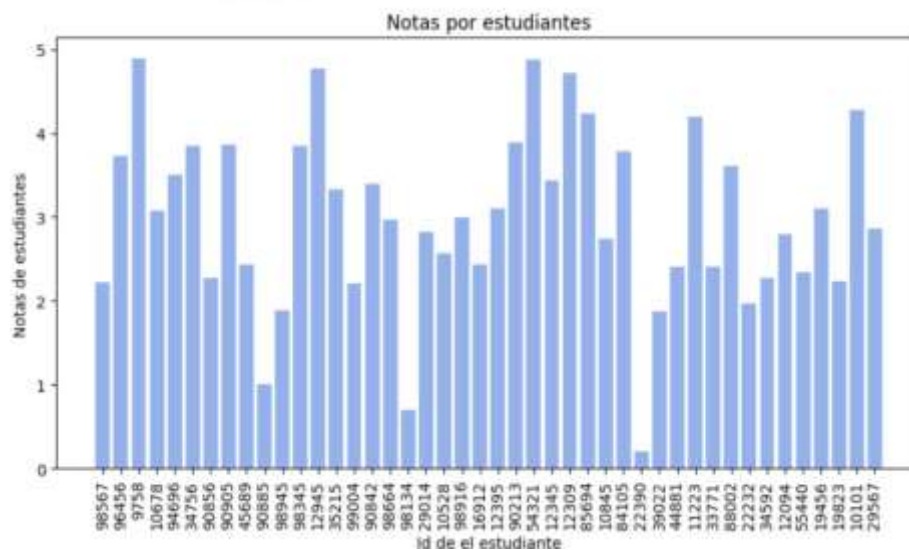
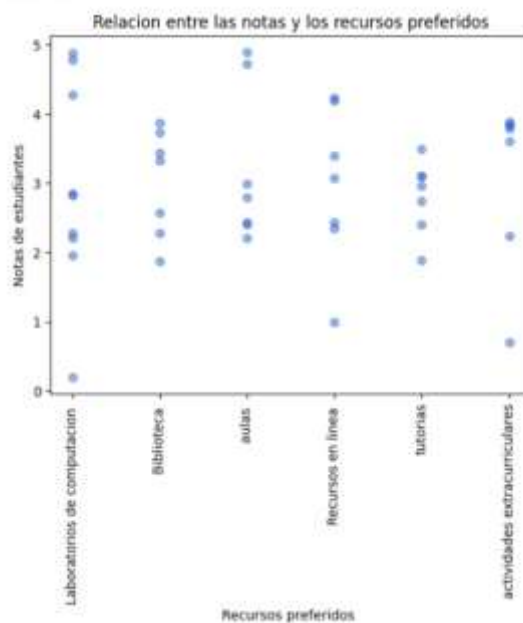


Ilustración 1 evidencia del sesgo de el dato atípico



| Index | Estudiantes | ID | genero | edad | Asistencias | Recursos educativos preferidos | Semestre | Promedio | Nivel Socioeconomico |
|-------|--------------------|-------|-----------|------|-------------|--------------------------------|----------|----------|----------------------|
| 2 | Cleiver Quiroz | 9758 | masculino | 19 | regular | aulas | IV | 5.0 | alto |
| 8 | Miguel Salcedo | 45689 | masculino | 23 | irregular | aulas | IV | 2.0 | bajo |
| 14 | Juliana Fernandez | 99004 | femenino | 19 | regular | aulas | IV | 2.0 | medio |
| 20 | Luisa Chacon niÃ±o | 98916 | femenino | 19 | regular | aulas | IV | 3.0 | bajo |
| 26 | Issabella Polanco | 12309 | femenino | 21 | irregular | aulas | V | 5.0 | alto |
| 32 | Juan Castillo | 44881 | masculino | 21 | irregular | aulas | V | 2.0 | medio |
| 38 | Nestor Ospino | 12094 | masculino | 20 | regular | aulas | IV | 3.0 | bajo |

Ilustración 2 este conjunto de estudiantes prefieren las aulas como recurso educativo preferido

| Index | Estudiantes | ID | genero | edad | Asistencias | Recursos educativos preferidos | Semestre | Promedio | Nivel Socioeconomico |
|-------|------------------|-------|-----------|------|-------------|--------------------------------|----------|----------|----------------------|
| 3 | Andres Rodriguez | 10678 | masculino | 25 | regular | Recursos en línea | IV | 3.0 | medio |
| 9 | Daniel otmos | 90885 | masculino | 23 | irregular | Recursos en línea | V | 1.0 | bajo |
| 15 | Julio Bonifacio | 90842 | masculino | 20 | irregular | Recursos en línea | IV | 3.0 | alto |
| 21 | Raul Puerta | 16912 | masculino | 20 | regular | Recursos en línea | IV | 2.0 | bajo |
| 27 | Jamie Ortega | 95694 | masculino | 22 | irregular | Recursos en línea | V | 4.0 | alto |
| 33 | Kenneth Morales | 11223 | masculino | 19 | irregular | Recursos en línea | IV | 4.0 | alto |
| 39 | Paulo PÃ eres | 55440 | masculino | 21 | regular | Recursos en línea | V | 2.0 | medio |

Ilustraci3n 3 este otro prefieren los recursos en l nea

| Index | Estudiantes | ID | genero | edad | Asistencias | Recursos educativos preferidos | Semestre | Promedio | Nivel Socioeconomico |
|-------|------------------|-------|-----------|------|-------------|--------------------------------|----------|----------|----------------------|
| 4 | Alejandra L belo | 94696 | femenino | 23 | regular | tutor as | IV | 4.0 | medio |
| 10 | Dilan Julio | 98945 | masculino | 21 | irregular | tutor as | V | 2.0 | medio |
| 16 | Luis avila | 98664 | masculino | 20 | irregular | tutor as | IV | 3.0 | medio |
| 22 | Geovany De Avila | 12395 | masculino | 19 | regular | tutor as | IV | 3.0 | bajo |
| 28 | Jairo Restrepo | 10845 | masculino | 22 | irregular | tutor as | V | 3.0 | medio |
| 34 | Luis Mercado | 33771 | masculino | 19 | regular | tutor as | IV | 2.0 | bajo |
| 40 | Orivaldo Tamayo | 19456 | masculino | 26 | irregular | tutor as | V | 3.0 | medio |

Ilustraci3n 4 estos prefieren las tutor as

| Index | Estudiantes | ID | genero | edad | Asistencias | Recursos educativos preferidos | Semestre | Promedio | Nivel Socioeconomico |
|-------|-------------------|-------|-----------|------|-------------|--------------------------------|----------|----------|----------------------|
| 1 | Juan Mosquera | 99456 | masculino | 21 | regular | Biblioteca | IV | 4.0 | medio |
| 7 | Alfredo Cerra | 90905 | masculino | 20 | irregular | Biblioteca | IV | 4.0 | bajo |
| 13 | Juan Fernandez | 35215 | masculino | 19 | regular | Biblioteca | IV | 1.0 | bajo |
| 19 | Joseph Suarez | 10528 | masculino | 21 | irregular | Biblioteca | IV | 3.0 | bajo |
| 25 | Hansel Galvis | 12345 | masculino | 20 | irregular | Biblioteca | IV | 3.0 | medio |
| 31 | Juan Ariza | 39022 | masculino | 20 | irregular | Biblioteca | V | 2.0 | medio |
| 37 | Valentina Salgado | 34592 | femenino | 20 | regular | Biblioteca | IV | 2.0 | medio |

Ilustraci3n 5 prefieren la biblioteca

| Index | Estudiantes | ID | genero | edad | Asistencias | Recursos educativos preferidos | Semestre | Promedio | Nivel Socioeconomico |
|-------|------------------|-------|-----------|------|-------------|--------------------------------|----------|----------|----------------------|
| 0 | Jesus Mulford | 99567 | masculino | 20 | regular | Laboratorios de computacion | IV | 2.0 | bajo |
| 8 | Gabriel Vallejo | 90856 | masculino | 22 | regular | Laboratorios de computacion | IV | 2.0 | bajo |
| 12 | Geraldine Puerta | 12945 | femenino | 20 | regular | Laboratorios de computacion | IV | 5.0 | medio |
| 18 | Jorge Blanquicet | 29014 | masculino | 21 | irregular | Laboratorios de computacion | IV | 3.0 | medio |
| 24 | Cristian Santana | 54321 | masculino | 23 | regular | Laboratorios de computacion | IV | 5.0 | medio |
| 30 | Jorge Guzman | 22390 | masculino | 19 | irregular | Laboratorios de computacion | IV | 0.2 | bajo |
| 36 | Sara Jimenez | 22232 | femenino | 19 | regular | Laboratorios de computacion | IV | 2.0 | medio |
| 42 | Michelle Nova | 10101 | femenino | 21 | irregular | Laboratorios de computacion | IV | 4.0 | alto |
| 43 | Juan Barron | 29567 | masculino | 23 | regular | Laboratorios de computacion | IV | 3.0 | medio |

laboratorio de computacion

| | Estudiantes | ID | genero | edad | Asistencias | Recursos educativos preferidos | Semestre | Promedio | Nivel Socioeconomico | genero |
|----|--------------------|-------|-----------|------|-------------|--------------------------------|----------|----------|----------------------|-----------|
| 5 | elder ruiz | 34756 | masculino | 20 | regular | actividades extracurriculares | IV | 4.0 | medio | masculino |
| 11 | Edgar Cervantes | 90345 | masculino | 19 | irregular | actividades extracurriculares | IV | 4.0 | medio | masculino |
| 17 | Sebastian Barragan | 98134 | masculino | 20 | irregular | actividades extracurriculares | IV | 0.7 | bajo | masculino |
| 23 | Oscar Vergara | 90213 | masculino | 20 | regular | actividades extracurriculares | IV | 4.0 | medio | masculino |
| 29 | Jesus Espitia | 84105 | masculino | 21 | irregular | actividades extracurriculares | IV | 4.0 | medio | masculino |
| 35 | Mauricio Ricaurte | 88002 | masculino | 20 | regular | actividades extracurriculares | V | 4.0 | medio | masculino |
| 41 | Alejandra Alfaro | 19823 | femenino | 20 | irregular | actividades extracurriculares | IV | 2.0 | medio | femenino |

Ilustración 6 y por ultimo. actividades extracurriculares

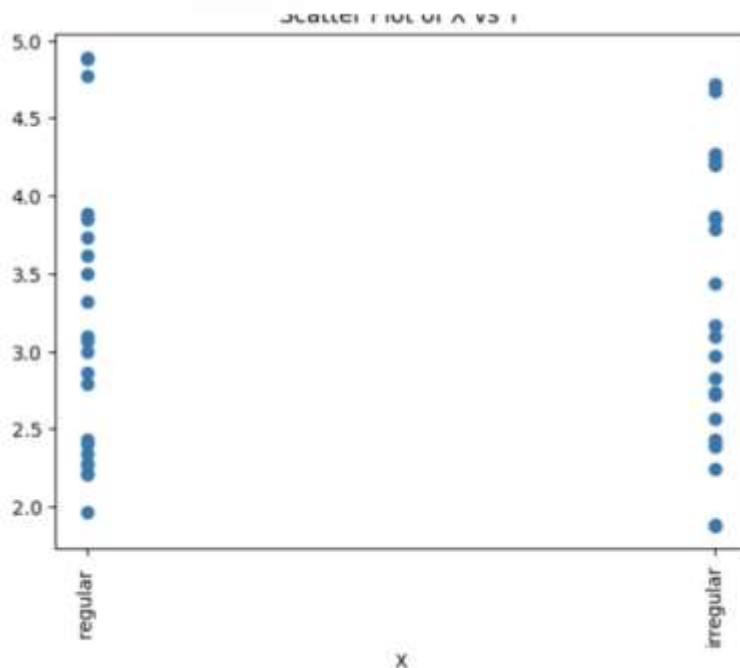


Ilustración 7 diagrama de dispercion comparando las notas vs la asistencia (regular o irregular)

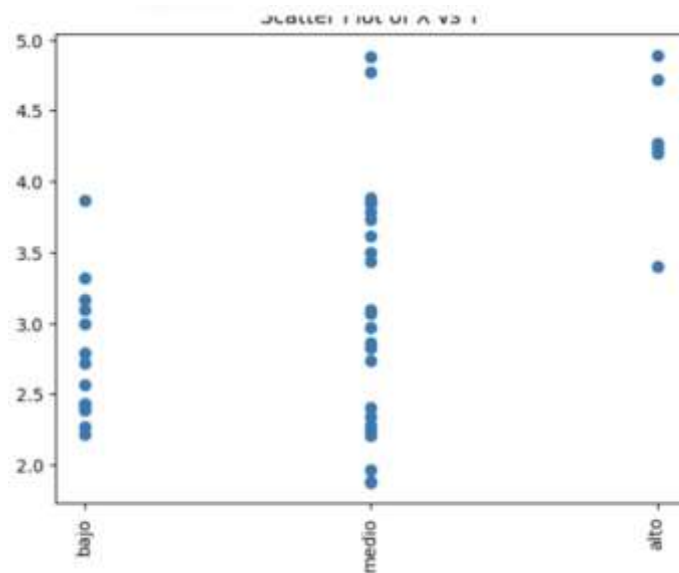


Ilustración 8 notas según el nivel socioeconómico

RECOMENDACIÓN Y PUNTUALIDAD A TENER EN CUENTA CON CADA DIAGRAMA ENTREGADO

Cada diagrama es acompañado de una descripción y justificación, resaltando la relevancia de los datos visualizados para el análisis. Los diagramas contribuyen a una mejor comprensión de las dinámicas estudiantiles, facilitando la interpretación de las tendencias y patrones observados.

CONCLUSIONES

Este análisis proporciona insights valiosos sobre la población estudiantil universitaria, resaltando la importancia de considerar variables demográficas y académicas en el diseño de políticas educativas. Los hallazgos sugieren la necesidad de abordar la disparidad de género, apoyar a estudiantes con asistencia irregular y diversificar los recursos educativos para adaptarse a las preferencias estudiantiles. Eliminar datos irrelevantes mejora la claridad del análisis, permitiendo un enfoque en la información más impactante.

Este trabajo puede servir como un estudio de caso o una referencia para investigaciones futuras, pero se recomienda encarecidamente a los usuarios que verifiquen la información y realicen sus propias evaluaciones antes de sacar conclusiones definitivas o tomar medidas basadas únicamente en los resultados presentados aquí.

En última instancia, la discreción y la prudencia son fundamentales al utilizar cualquier dato o análisis presentado en este documento, y se anima a los usuarios a buscar asesoramiento profesional o realizar investigaciones adicionales según sea necesario. Esto ha sido todo.

REFERENCIAS BIBLIOGRAFICAS

<https://tudashboard.com/sesgo-en-el-analisis-de-datos/#:~:text=Tambi%C3%A9n%20hay%20que%20resaltar%20los%20datos%20que%20respalden,puede%20conducir%20a%20un%20an%C3%A1lisis%20de%20datos%20incorrecto.>

<https://colab.research.google.com/>