



Análisis e identificación de sesgos en los estudiantes de la clase
de “Electiva Disciplinar I”

Sara Jiménez, Luis Mercado, Kennet Morales, Juan Mosquera, Jesús Mulford y Néstor
Ospino

Universidad del Sinú seccional Cartagena de Indias

Sede Plaza Colón

Electiva I

Ing. Cristian Camilo Cuadrado Beltrán

Abril 2024

Introducción

Se han registrado una serie de datos de los estudiantes tipo encuesta de la clase de “Electiva Disciplinar I”, con el fin de realizar un debido análisis exploratorio de datos, con el fin de identificar sesgos, para la finalidad de generar una serie de hipótesis en base a estos mismos.

Objetivo del análisis exploratorio

- Realizar el análisis exploratorio del grupo de estudiantes
- Identificar sesgos en el conjunto de datos

Contextualización

Cabe mencionar que la observación y la entrevista fueron utilizadas como métodos de recolección de datos. Es importante tener en cuenta que estos datos pueden no ser 100% verídicos, ya que están sujetos a la percepción y memoria de los participantes, así como a posibles sesgos en la interpretación de los investigadores. Por lo tanto, el análisis exploratorio se realiza reconociendo esta limitación y se centra en identificar patrones y tendencias generales en los datos recopilados, más que en establecer conclusiones definitivas, usando una metodología analítica en el proceso.

Descripción de los datos

El conjunto de datos contiene un total de 272 registros, distribuidos en 34 filas y 8 columnas.

La columna “*Estudiante*” identifica a las personas encuestadas, mientras que la columna “*Género*” indica el género de los estudiantes. La columna “*Foraneo*” registra si los estudiantes viven en la ciudad o solo están estudiando allí temporalmente.

“*Conforme_PC_Institucional*” refleja la conformidad de los estudiantes con los dispositivos de la universidad, mientras que “*Tiene_PC_Personal*” dice qué estudiantes cuentan con un propio dispositivo personal, y “*Trae_PC_Personal*” indica si suelen traer su propio

dispositivo para actividades académicas. La columna “*ActividadesExtracurriculares*” indica si están inscritos en actividades culturales y deportivas extracurriculares, y “*Uniforme_Completo(%)*” muestra qué tan completo llevan el uniforme de la universidad, expresado en porcentaje. Estos datos proporcionan información valiosa sobre las preferencias y comportamientos de los estudiantes en la universidad.

Clasificando los datos, identificamos que *Estudiante*, *Género*, *Foraneos*, *Conforme_PC_Institucional*, *Tiene_PC_Personal*, *Trae_PC_Personal*, *ActividadesExtracurriculares* son variables categóricas mientras que *Uniforme_Completo(%)*, numéricas.

```
Data columns (total 8 columns):
#      Column                                Non-Null Count  Dtype
---  -
0      Estudiante                             34 non-null    object
1      Genero                                   34 non-null    object
2      Foraneos                                 34 non-null    object
3      Conforme_PC_Institucional                34 non-null    object
4      Trae_PC_Personal                        34 non-null    object
5      Actividad Extracurriculares              34 non-null    object
6      Uniforme_Completo(%)                    34 non-null    int64
7      Tiene PC Personal                       34 non-null    object
```

Manejo de datos faltantes

En el proceso de análisis se identificó la ausencia de valores nulos o valores faltantes, es decir, todas las filas de las columnas contienen valores válidos, por lo que el análisis no se verá afectado en la precisión de los cálculos y creación de hipótesis y conclusiones, permitiendo la fiabilidad y precisiones en las visualizaciones presentadas.

Análisis descriptivo

A continuación se presenta un resumen estadístico descriptivo general de las variables que conforman el conjunto de datos, de un total de 34 estudiantes, identificamos la frecuencia de las variables:

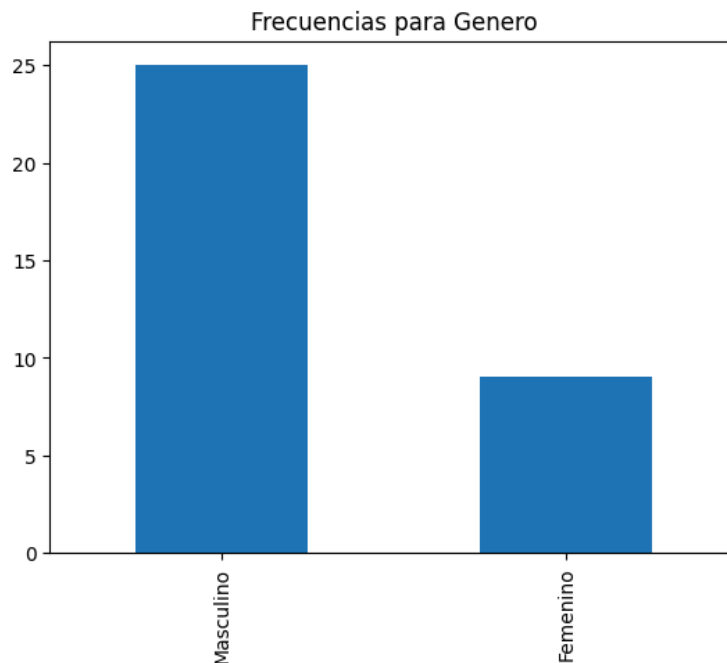
Genero		Conforme_PC		Foraneos		Trae_PC_Personal		Actividad_Extracurricular		Uniforme_Completo(%)	
Masculino	25	Si	10	Si	18	Si	12	No inscrito	17	100%	9
Femenino	9	No	24	No	16	No	22	Futbol	7	50%	11
								Voleibol	5	25%	8
								Baloncesto	4	0%	6
								Orquesta y Musica	1		

Tiene_PC_Personal	
Si	31
No	3

Inmediatamente en un enfoque general identificamos lo siguiente:

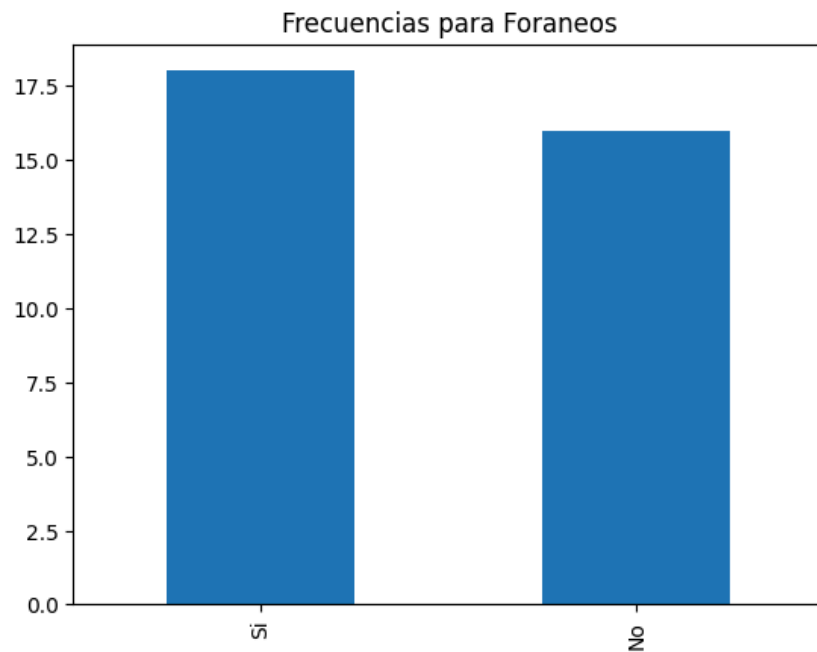
- El número de estudiantes masculinos es mayor que el número de estudiantes femeninos.
- Hay más estudiantes inconformes con los dispositivos de cómputo de las instalaciones.
- La cantidad de estudiantes foráneos es casi la misma que los estudiantes locales.
- Es mayor la cantidad de estudiantes que por lo general no usan su pc personal en la universidad.
- La cantidad de estudiantes que llevan el uniforme completo es muy baja comparada con el total de estudiantes.

Diagrama de frecuencias para Géneros:



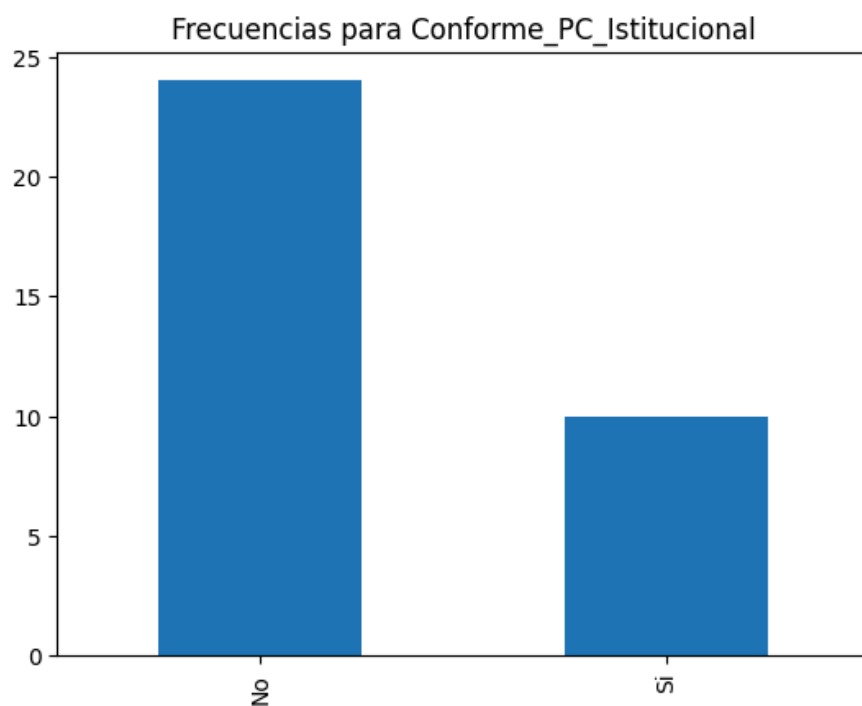
Este diagrama muestra la frecuencia de los géneros Masculino y Femenino, demostrando la predominancia en el género masculino.

Diagrama de frecuencias para Foráneos:



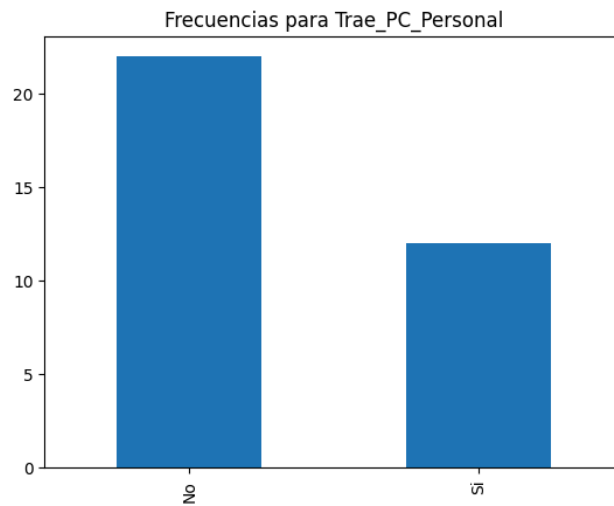
Este diagrama evidencia la frecuencia con la que los estudiantes son foráneos o no en el salón de electiva I, es decir, que provienen de otro lugar.

Diagrama de frecuencia para “Conforme_PC_Institucional”:



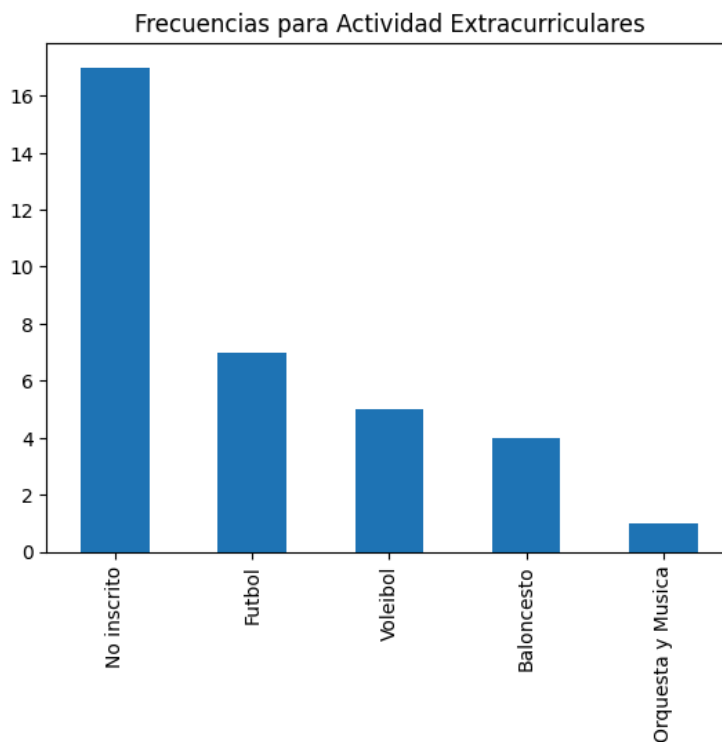
Este diagrama evidencia la cantidad de estudiantes que están conformes con los equipos institucionales del salón de electiva I.

Diagrama de frecuencia para “Trae_Pc_Personal”



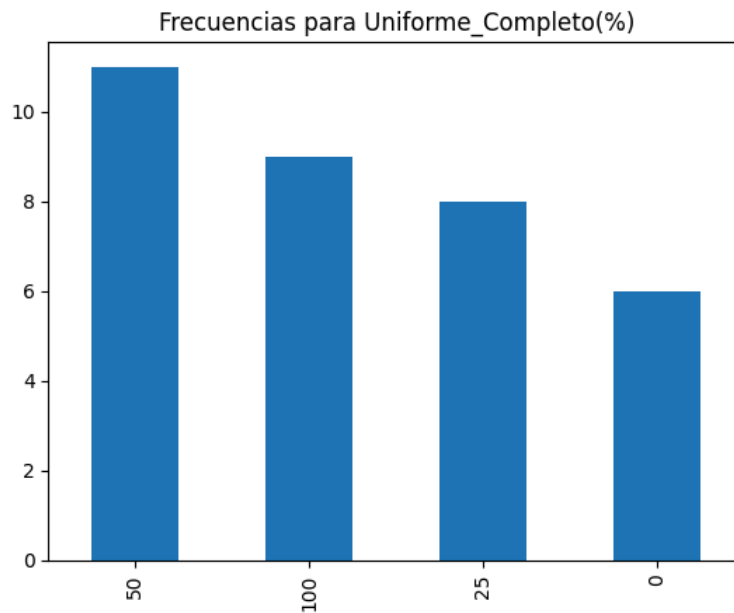
En este diagrama podemos evidenciar la cantidad de estudiantes que lleva su PC personal.

Diagrama para las actividades culturales:



En este diagrama evidenciamos las actividades a las que se inscriben los estudiantes, siendo “No inscrito” la más frecuente.

Diagrama para las frecuencias de uniforme completo:



En este diagrama podemos ver los porcentajes de estudiantes que llevan el uniforme completo requerido a clases.

Diagrama de dispersión sobre el porcentaje de estudiantes que llevan el uniforme completo:

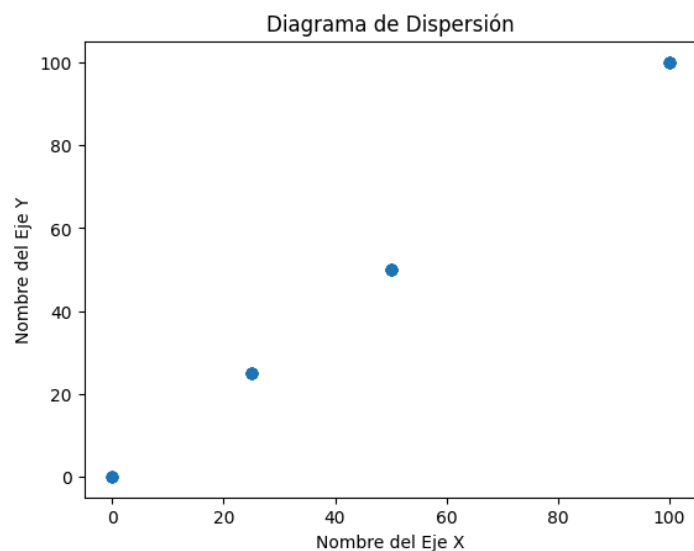
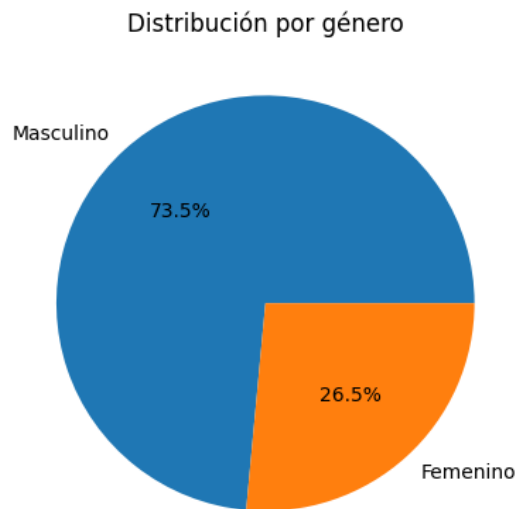
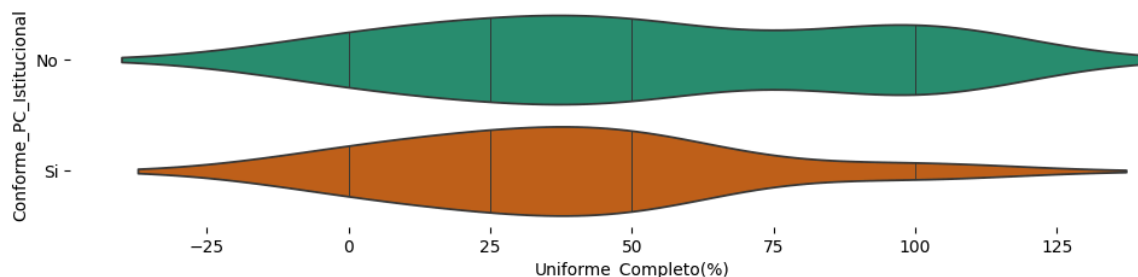


Gráfico de pastel para la distribución de géneros:



Este diagrama muestra la distribución total en géneros de una lista de estudiantes, donde aclara que existe mayor número de estudiantes con sexo masculino que femenino, arrojando una cifra favorable para el sexo masculino de un 73,5%, mientras que el sexo femenino cuenta con un 26,5%.

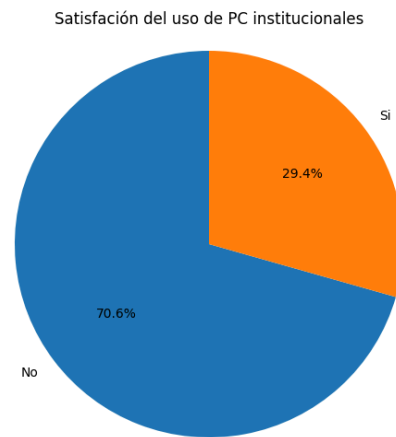
Comparación de estudiantes que tienen uniforme completo y los estudiantes que están conformes con los equipos institucionales:



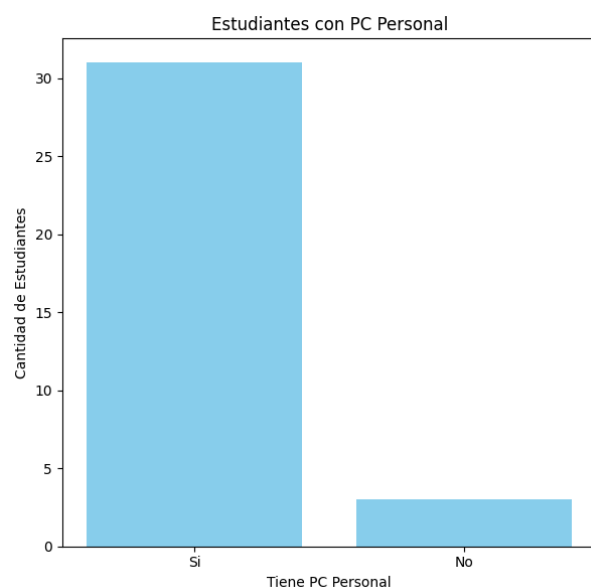
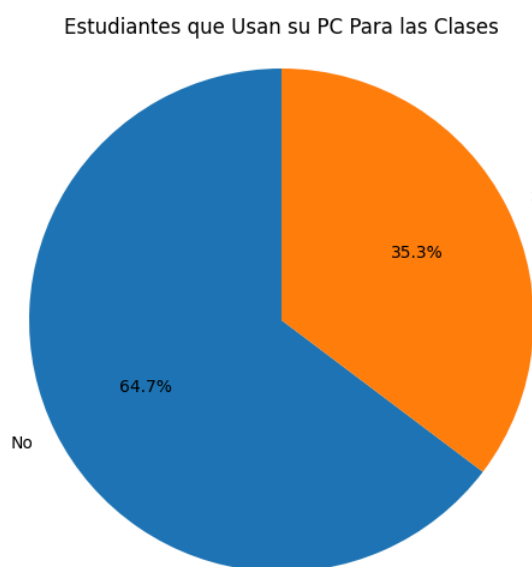
Identificación de sesgos

Después de analizar la información, se identificaron varios sesgos. El primer sesgo identificado podría considerarse un sesgo de negación, ya que inicialmente se consideró que

la insatisfacción de la mayoría de estudiantes (24 de 34, es decir, 70.6% de los estudiantes) con los equipos de cómputo de las salas de la universidad podría ser debido a la lentitud o falencias en los dispositivos de entrada.

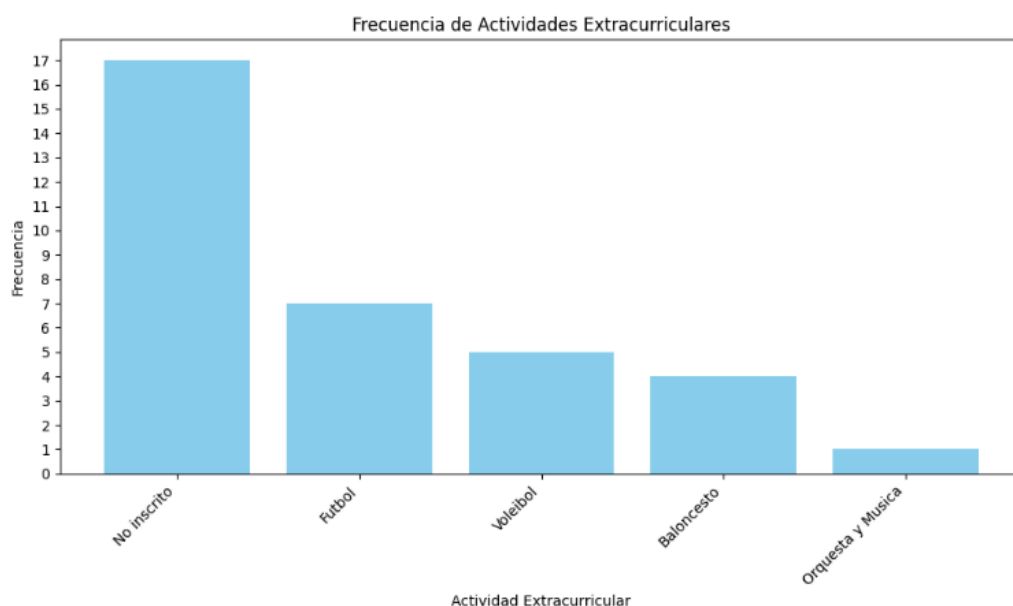


A partir de esto, se supuso que esta insatisfacción llevaría a que un alto número de estudiantes utilizarán sus dispositivos personales en clase, dado que casi todos cuentan con un dispositivo personal (solo 3 de 34 estudiantes no tienen). Sin embargo, los datos recopilados comprobaron que solo el 35.3%, es decir, 12 de 34 estudiantes, efectivamente usan su dispositivo personal en clase, mientras que el 64.7% no.



Lo que sugiere que la insatisfacción con los equipos de la universidad no necesariamente se traduce en un uso generalizado de dispositivos personales en el salón de clases, desmintiendo completamente la hipótesis o creencia inicial.

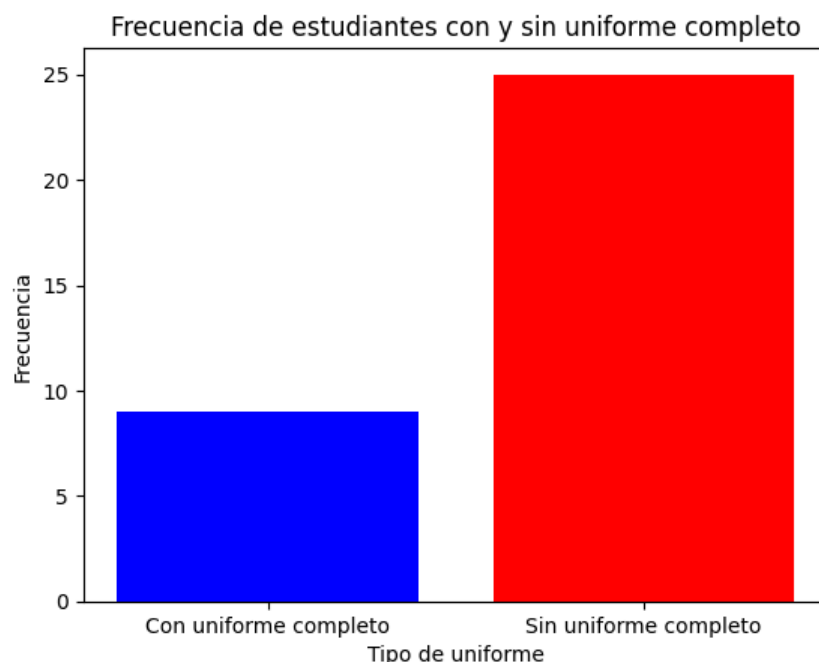
Otro sesgo que identificamos es en las actividades extracurriculares, la “Orquesta y Música” podría considerarse como un outlier, el cual podría usarse para un sesgo de valores atípicos tiene una participación significativamente menor en comparación con las otras actividades, lo cual puede deberse a varios factores como falta de conocimiento sobre esta actividad, falta de interés generalizado, o posiblemente falta de promoción o marketing por parte de la universidad en comparación con otras actividades extracurriculares.



La frecuencia de participación en las actividades extracurriculares muestra que la mayoría de los estudiantes (17 en total) no están inscritos en ninguna actividad extracurricular ("No inscrito"). Sin embargo, la actividad "Orquesta y Música" tiene una frecuencia de solo 1 estudiante participante, lo que la convierte en un valor atípico en el conjunto de datos. Las otras actividades extracurriculares, como "Futbol", "Voleibol" y "Baloncesto", tienen

frecuencias de 7, 5 y 4 estudiantes respectivamente, lo que indica que estas actividades son más comunes entre los estudiantes encuestados.

Por otro lado, se sabe que el porte del uniforme de la universidad es de carácter obligatorio, pero se considera que esta norma no se está cumpliendo, decimos que son muy pocos los estudiantes que portan el uniforme completo en un 100%. Es por ello que identificamos un sesgo de confirmación, en el diagrama muestra claramente que hay más estudiantes que no utilizan el uniforme completo en comparación con aquellos que sí lo usan. Esta diferencia significativa respalda la existencia del sesgo de confirmación, donde los estudiantes podrían estar optando por no cumplir con la normativa del uniforme completo por diversas razones, como no estar conforme con las normas de presentación personal de la universidad, factores económicos de no disponer del uniforme, o de paso prefieren usar otras prendas con las cuales se sienten más cómodos.



La barra azul representa la frecuencia de estudiantes que llevan el uniforme completo, que en este caso son 9 estudiantes. Mientras que la barra roja representa la frecuencia de estudiantes

que no llevan el uniforme completo, que son 25 estudiantes. La diferencia en la altura de las barras muestra claramente que hay más estudiantes que no llevan el uniforme completo en comparación con aquellos que sí lo llevan completo.

Conclusión

En base al análisis exploratorio de datos y la identificación de sesgos en base a los datos, se concluye que el siguiente paso a realizar es hacer una recolección y registro de datos más precisa, para posteriormente realizar la respectiva comparación con las hipótesis planteadas y la primera data recolectada, para así poder plantear una serie de recomendaciones en base al resultado de la comparación y verificación de los datos.