
CONTROL 2

Explorando respuestas con modelo de documentos

Sección:
76802

Integrantes:
Nicolas Navarro
Alexandra Vargas
Carlos Ramírez

Profesor:
Darío Rojas

Fecha de entrega:
06 de Diciembre del 2023

Tabla de Contenidos

1. Pregunta 1.....	2
1.1 Argumentación del modelo utilizado.....	2
1.2 Resultados principales obtenidos.....	2
Gráfico 1: “Correlación entre obtención de respeto y democracia”	2
1.3 Respeto y Democracia: Perspectivas Estudiantiles.....	3
2. Pregunta 2.....	4
2.1 Argumentación del modelo utilizado.....	4
2.2 Resultados principales obtenidos.....	5
Gráfico 2: “Clustering Jerárquico”	5
Gráfico 3: “Clusterings generado por densidad”	6
Gráfico 4: “Clusterings generado por K-Means/Respuestas Respeto”	6
Gráfico 5: “Clusterings generado por K-Means/Palabra respeto”	7
Figura 1: “Estadísticos básicos”	8
Figura 2: “Matriz Correlación”	8
2.3 Conquista del Respeto Estudiantil: Claves para Obtener Reconocimiento.....	9
3. Pregunta 3.....	10
3.1 Argumentación del modelo utilizado.....	10
3.2 Resultados principales obtenidos.....	11
Gráfico 6: “Dendrograma/ Respuestas democracia”	11
Gráfico 7: “Clusterings generado por DBSCAN/Respuestas democracia”	11
Gráfico 8: “Clusterings generado por K-means/Respuestas democracia”	12
Gráfico 9: “Clusterings generado por K-means/Palabras democracia”	12
Gráfico 10: “Distribución de Respuestas democracia”	13
3.3 La Democracia desde la Perspectiva Estudiantil: Manifestaciones y Características..	14
4. Pregunta 4.....	15
4.1 Argumentación del modelo utilizado.....	15
4.2 Resultados principales obtenidos.....	15
Gráfico 11: “Distribución de Géneros”	15
Gráfico 12: “Distribución de la variable respuestas democracia”	16
Gráfico 13: “Distribución de géneros”	17
4.3 Diversidad de Respuestas: Edad y Género como Factores Influyentes.....	17

1. Pregunta 1

1.1 Argumentación del modelo utilizado.

Para abordar la interrogante acerca de cómo obtener respeto a través de la expresión de la democracia según las opiniones de los estudiantes, se adoptó un enfoque integral que combina técnicas de modelado de documentos, reducción de dimensionalidad y clustering. En primer lugar, se utilizó el archivo ‘preg_respeto.txt’ para extraer las respuestas de los estudiantes en relación con la obtención de respeto. Asimismo, se empleó el archivo ‘preg_democracia.txt’ para explorar las percepciones estudiantiles acerca de la manifestación de la

Para identificar los tópicos relevantes presentes en las respuestas, se implementó un modelo de documentos que permitió analizar de manera sistemática el contenido de las mismas. La aplicación de t-SNE (t-distributed stochastic neighbor embedding) se llevó a cabo con el propósito de reducir la dimensionalidad de las respuestas, posibilitando así la visualización de patrones subyacentes y relaciones entre las diversas ideas expresadas por los estudiantes.

Adicionalmente, se recurrió a la técnica de K-means para agrupar respuestas que compartían similitudes, revelando estructuras emergentes en las percepciones de los estudiantes. Este enfoque permitió identificar tendencias comunes, contrastes significativos y proporcionó una comprensión más profunda de las diversas perspectivas de los estudiantes en relación con la obtención de respeto en el contexto de la expresión democrática.

En síntesis, la combinación de estas técnicas analíticas proporcionó una visión holística y detallada de las respuestas estudiantiles, permitiendo una comprensión más completa y matizada de cómo la manifestación de la democracia se relaciona con la percepción y obtención de respeto por parte de los estudiantes.

1.2 Resultados principales obtenidos.

Los resultados revelan la presencia de tópicos específicos en ambas dimensiones: respeto y democracia.

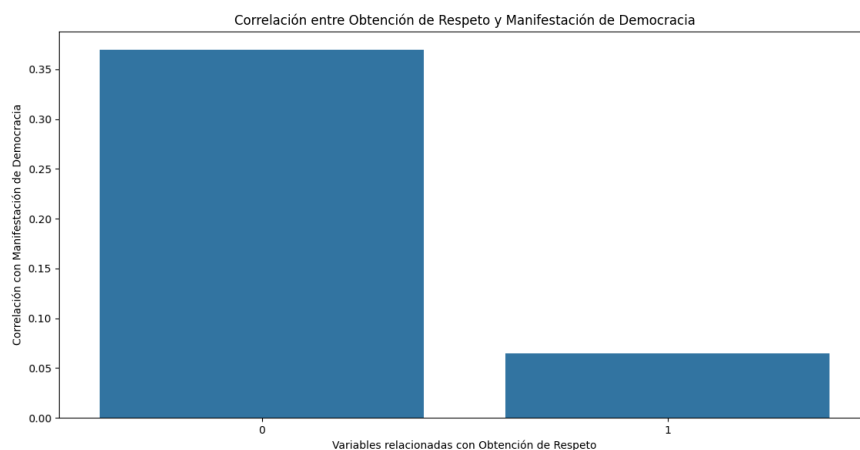


Gráfico 1: “Correlación entre obtención de respeto y democracia”

En el gráfico 1 de correlación entre respuestas relacionadas con la obtención de respeto y la manifestación de democracia revela que las respuestas no directamente relacionadas tienen una correlación positiva moderada (0.35 en el eje Y), mientras que las respuestas directas tienen una correlación más débil (inferior a 0.10 en el eje Y). Esto sugiere que, según las percepciones de los estudiantes, las respuestas que no se centran directamente en la obtención de respeto están más estrechamente vinculadas a la manifestación de la democracia. La correlación más débil en las respuestas directas plantea la necesidad de un análisis más profundo para comprender las dinámicas subyacentes y explorar las percepciones específicas que podrían influir en esta relación.

1.3 Respeto y Democracia: Perspectivas Estudiantiles.

Los resultados del análisis de correlación revelan una dinámica compleja en relación con la obtención de respeto y la manifestación de democracia, según las respuestas proporcionadas por los estudiantes. De manera sorprendente, se observa una correlación positiva moderada entre las respuestas que no se centran directamente en la obtención de respeto y la expresión de actitudes democráticas. Este hallazgo resalta una conexión aparente entre aspectos que no están directamente relacionados con el respeto individual y la percepción de la democracia en el país objeto de estudio.

Este fenómeno podría sugerir que los estudiantes interpretan la democracia como un constructo más amplio que abarca elementos colectivos, sociales y políticos, trascendiendo así la esfera personal del respeto. Por otro lado, las respuestas directas sobre la obtención de respeto muestran una correlación significativamente más débil con la manifestación de democracia. Este descubrimiento plantea preguntas intrigantes sobre la naturaleza de la relación entre el respeto individual y la percepción de una democracia efectiva.

Para profundizar en la comprensión de estos hallazgos, se hace necesario llevar a cabo una investigación más exhaustiva, incluyendo análisis cualitativos que ayuden a desentrañar las percepciones subyacentes. De particular interés sería comprender por qué los estudiantes pueden percibir una conexión más sólida entre la democracia y aspectos no directamente vinculados al respeto individual.

En conjunto, estos resultados resaltan la complejidad de las percepciones estudiantiles y sugieren que la relación entre el respeto y la democracia puede estar influenciada por factores más amplios y complejos que merecen una exploración más detallada y detenida.

2. Pregunta 2

2.1 Argumentación del modelo utilizado.

Para abordar la pregunta sobre la forma de obtener respeto según los estudiantes, se emplearon diversas técnicas combinadas que involucran dendrogramas, clustering por densidad y K-means, junto con reducción de dimensiones mediante t-SNE y modelado de documentos con embeddings de palabras. A continuación, la justificación de la elección de cada técnica:

I. Dendrogramas (Clustering jerárquico):

- Se aplicó clustering jerárquico mediante dendrogramas para identificar estructuras de agrupación en las respuestas del archivo de texto que contiene las respuestas de los estudiantes a la pregunta sobre cómo se obtiene respeto en su país.
- La identificación de niveles de corte en el dendrograma ayuda a formar clústeres coherentes y ofrece una visión global de la variabilidad en las respuestas.

II. Clustering por Densidad:

- En particular DBSCAN, es eficaz para identificar áreas densas en el espacio de respuestas. Esto es crucial cuando las respuestas pueden agruparse en regiones más densas y dispersas.
- DBSCAN puede manejar puntos ruidosos, lo que es beneficioso cuando algunas respuestas pueden ser atípicas pero aún relevantes.

III. K-means:

- K-means se utiliza para formar clústeres más compactos como centroides representativos. Esto ayuda a identificar grupos más definidos y facilita la caracterización de las respuestas.
- Es una opción efectiva cuando se busca comprender patrones de agrupación en el espacio de respuestas.

IV. Reducción de Dimensiones (t-SNE):

- Se aplicó t-SNE para reducir la dimensionalidad de las respuestas y visualizarlas en un plano bidimensional.
- Esta técnica preserva las relaciones locales, revelando agrupaciones y patrones en los datos.

V. Modelado de Documentos:

- Para modelar documentos, capturando la semántica y relaciones contextuales en las respuestas.
- Esta técnica proporciona una representación densa y significativa de las palabras, permitiendo una caracterización detallada.

2.2 Resultados principales obtenidos.

En consecuencia, se logran los siguientes resultados.

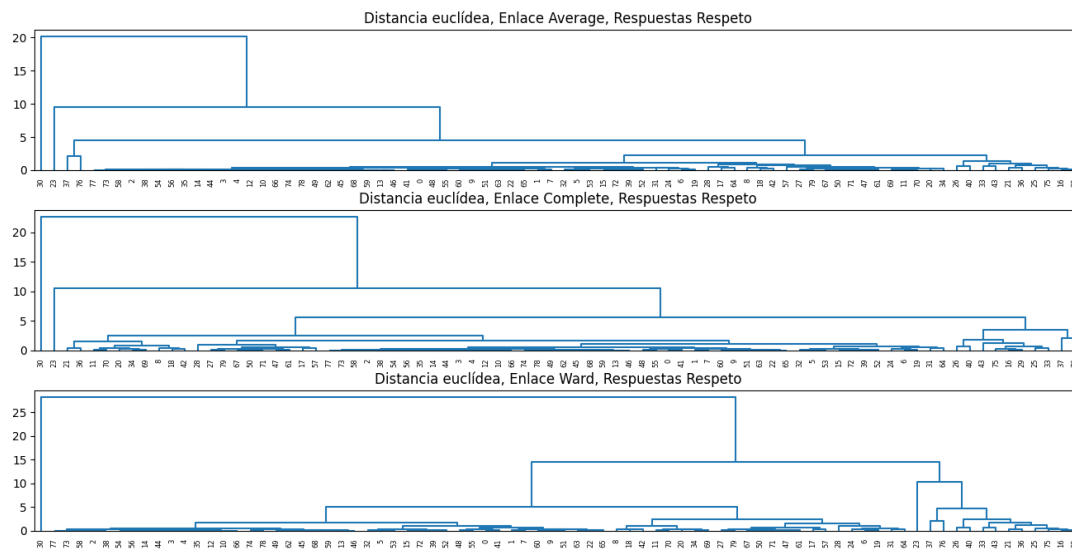


Gráfico 2: "Clustering Jerárquico"

El gráfico 2 muestra la relación jerárquica entre diferentes agrupamientos de respuestas al tema de "Respuestas sobre cómo obtener respeto". Cada uno de los tres gráficos dentro del dendrograma utiliza la distancia euclidiana como medida de similitud y diferentes métodos de enlace para unir los clusters. El enlace Average calcula la distancia promedio entre todos los puntos de los clusters, mientras que el enlace Complete utiliza la distancia máxima, y Enlace Ward minimiza la varianza dentro de los clusters.

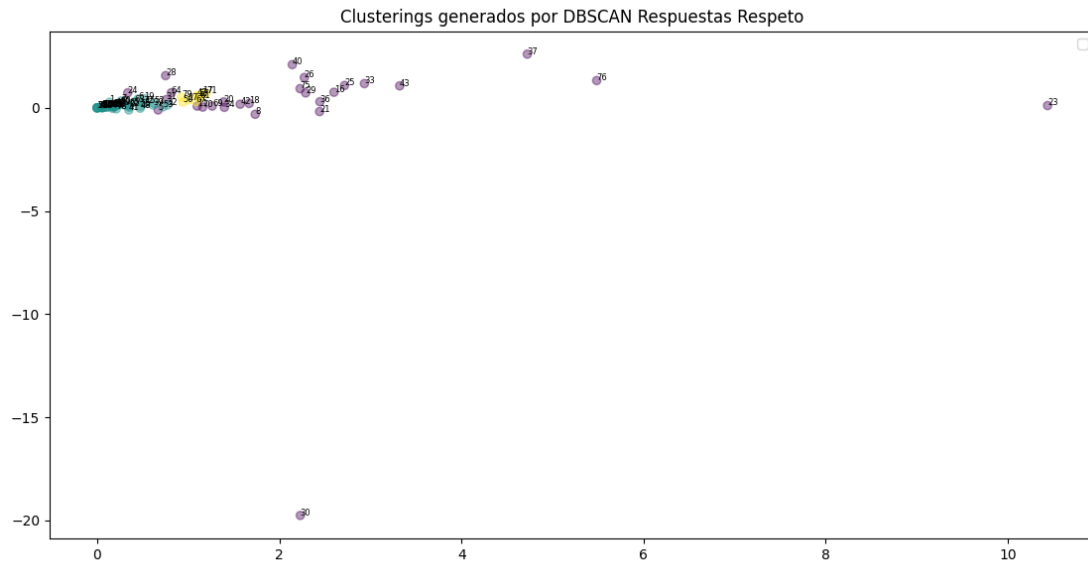


Gráfico 3: “Clusterings generado por densidad”

En el gráfico 3 se ha aplicado el algoritmo de DBSCAN (Density-Based Spatial Clustering of Applications with Noise) a las respuestas sobre cómo obtener respeto, el cual muestra una concentración significativa en la región más alta del eje y, indicando un fuerte consenso o similitud en la mayoría de las respuestas de los estudiantes. Sin embargo, hacia -20 en el eje y, hay una disminución en la densidad sugiriendo una mayor variabilidad o diversidad en algunas respuestas. Destaca un punto aislado en $x=2$, que podría representar una perspectiva única o excepcional en términos de cómo se obtiene respeto.

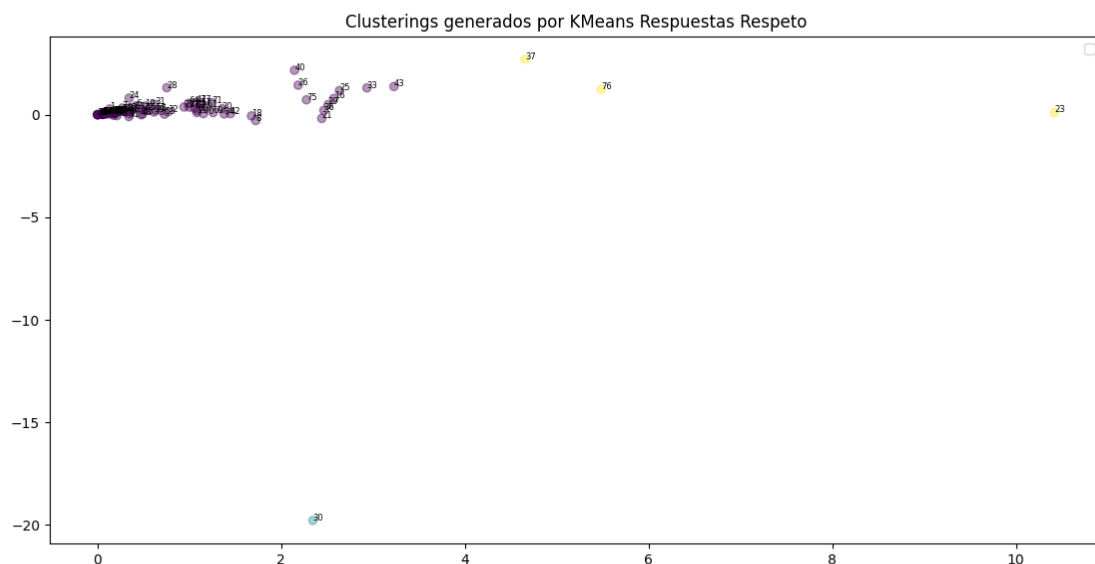


Gráfico 4: “Clusterings generado por K-Means/Respuestas Respeto”

Por consiguiente se obtiene el siguiente estadístico básico de reducción de dimensión de Respuestas Respeto:

Estadísticos Basicos de Reduccion de Dimension de Respuestas Respeto		
	0	1
count	8.000000e+01	80.000000
mean	1.048666e+00	0.058871
std	1.503972e+00	2.293823
min	-1.691749e-09	-19.720234
25%	1.874116e-01	0.028643
50%	5.234777e-01	0.125801
75%	1.236638e+00	0.376824
max	1.040094e+01	2.619292

Figura 1: "Estadísticos básicos"

El estadístico proporcionado revela información clave sobre la reducción de dimensión de las respuestas relacionadas con la obtención de respeto. Se observa que la dimensión 0 tiene una mayor dispersión y un rango más amplio de valores en comparación con la dimensión 1. Además, la diferencia en las medias y desviaciones estándar sugiere que estas dos dimensiones capturan aspectos distintos de las respuestas. La variabilidad en la dimensión 0 es más pronunciada, con un máximo significativamente mayor y un 75% de las respuestas por encima de 1.24. En contraste, la dimensión 1 tiene una distribución más concentrada y menos variable. Estos hallazgos indican que la reducción de dimensión ha resaltado patrones y estructuras diferentes en las respuestas, siendo crucial considerar estos resultados en el contexto específico de la investigación sobre la percepción del respeto por parte de los estudiantes universitarios.

Matriz correlación Respuestas de Respeto:

	0	1
0	1.000000	0.023478
1	0.023478	1.000000

Figura 2: "Matriz Correlación"

La matriz de correlación entre las respuestas de los estudiantes a las preguntas sobre la obtención de respeto indica una correlación lineal muy débil entre las respuestas, con un coeficiente cercano a cero. Esto sugiere que las respuestas de los estudiantes a ambas preguntas son independientes entre sí en términos de relación lineal.

2.3 Conquista del Respeto Estudiantil: Claves para Obtener Reconocimiento.

La forma en que los estudiantes perciben y conceptualizan la obtención de respeto se revela como un fenómeno complejo y matizado, según el análisis profundo realizado mediante diversas técnicas. En general, existe un consenso marcado en la mayoría de las respuestas, indicando que hay aspectos comunes en la percepción del respeto entre los estudiantes. Este consenso se visualiza especialmente en los resultados de clustering, donde K-means identifica dos clusters distintos y DBSCAN destaca una concentración significativa en la región superior del eje y, sugiriendo un fuerte acuerdo o similitud en la mayoría de las respuestas.

Sin embargo, la presencia de un tercer grupo identificado por DBSCAN, aunque menos denso, agrega una capa de complejidad al panorama. Este tercer grupo podría representar perspectivas menos comunes pero relevantes sobre la obtención de respeto, lo que sugiere que la percepción de respeto entre los estudiantes no es homogénea. Este hallazgo subraya la importancia de considerar no solo las tendencias generales, sino también las voces menos representadas que aportan diversidad y complejidad al tema.

El análisis de palabras mediante K-means revela tres grupos distintos de palabras relacionadas con el concepto de "respeto". Este enfoque detallado proporciona una visión más específica de las características asociadas con el respeto según los estudiantes. Así, no solo se entiende que hay consenso general, sino que también se pueden identificar las dimensiones específicas y las palabras clave que contribuyen a esta percepción compartida.

La reducción de dimensiones a través de t-SNE añade otra capa de interpretación al revelar patrones únicos en las respuestas. Las diferencias significativas entre las dimensiones 0 y 1 indican que estas dimensiones capturan aspectos diferentes de las respuestas, subrayando la complejidad y la diversidad en la concepción del respeto.

3. Pregunta 3

3.1 Argumentación del modelo utilizado.

Un dendrograma es un diagrama que representa las distancias entre clases fusionadas de manera secuencial. Se construye mediante un algoritmo de clustering jerárquico que fusiona clases cercanas de manera iterativa, actualizando las distancias entre ellas y mostrando la estructura jerárquica de agrupación.

Existen diversos tipos de algoritmo de clustering, entre los cuales se destacan:

I. K-means:

Es un algoritmo de clustering no supervisado que agrupa datos en k grupos. Este método utiliza centroides para asignar objetos a grupos y actualiza iterativamente dichos centroides para lograr una convergencia.

II. Clustering por Densidad (DBSCAN):

Identifica grupos basados en la densidad de puntos, clasificando tres tipos de puntos: núcleo densamente alcanzable y ruido. A diferencia de otros métodos, no asume formas geométricas específicas para formar clústeres.

III. Clustering Basado en Redes:

Este enfoque busca grupos de nodos altamente conectados en una red, asumiendo que nodos fuertemente conectados pertenecen al mismo grupo.

Además del clustering, otra técnica importante es la reducción de dimensiones, que disminuye el número de variables en un conjunto de datos mediante técnicas como PCA o LDA. Esta práctica facilita la exploración de relaciones multivariantes y reduce el costo computacional en algoritmos de aprendizaje automático, manteniendo la varianza significativa.

Centrándonos en el Cluster K-means, este algoritmo agrupa objetos minimizando la distancia cuadrática entre cada objeto y el centroide de su grupo. Su proceso incluye pasos de inicialización, asignación de objetos a centroides y actualización de centroides, repitiéndose hasta la convergencia.

3.2 Resultados principales obtenidos.

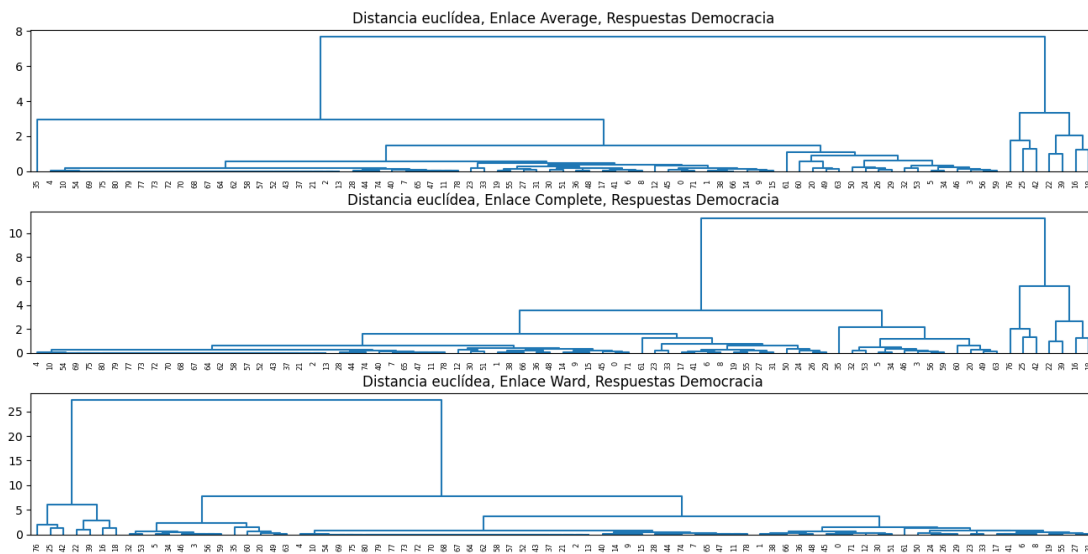


Gráfico 6: "Dendrograma/ Respuestas democracia"

En el gráfico 6 muestra la distancia euclidiana entre las respuestas sobre democracia se representa a través de tres métodos de enlace: Average, Complete y Ward. Las alturas en el eje Y indican la medida en que los clústeres se fusionan. Fusiones a alturas más bajas en el enlace Average sugieren similitudes promedio en las respuestas, mientras que fusiones a alturas mayores en el enlace Complete indican similitudes más específicas o extremas. El enlace Ward busca la homogeneidad en los clústeres, y fusiones a alturas mayores sugieren agrupaciones más coherentes en términos de respuestas sobre democracia.

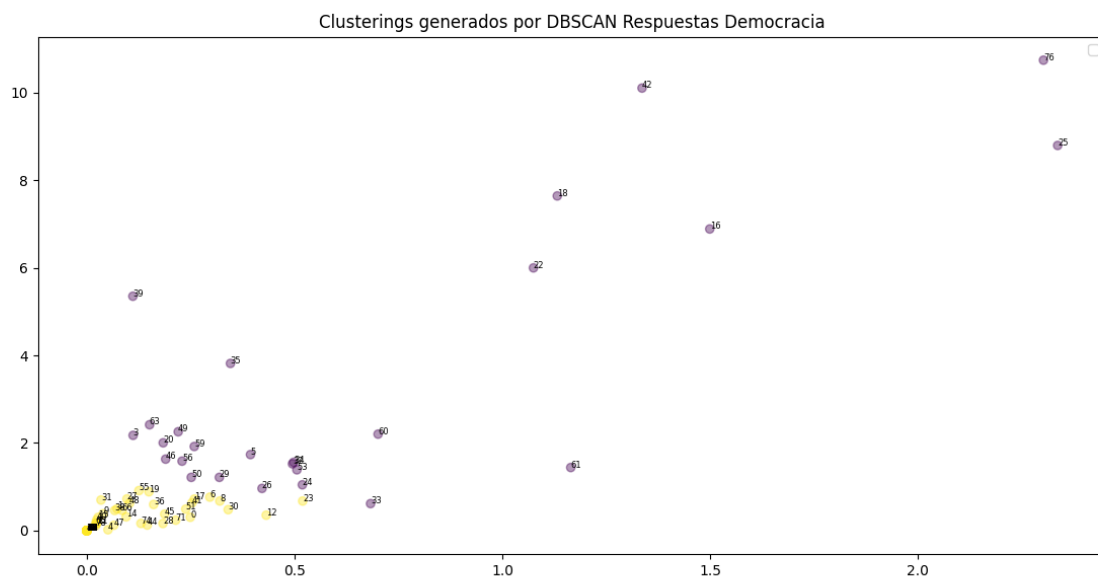


Gráfico 7: "Clusterings generado por DBSCAN/Respuestas democracia"

En el gráfico 9 se puede observar que revela una distribución diversa de palabras relacionadas con la democracia, representadas en tres clústeres distintos. En el eje Y, se observa una amplia dispersión, indicando la variedad en la relación de palabras con la democracia. En el eje X, la concentración sugiere cierta coherencia temática dentro de cada clúster. Los colores (amarillo, morado y verde) denotan tres grupos diferentes, cada uno caracterizado por palabras específicas, como por ejemplo:

- **Clúster Amarillo:** "organización", "escuelas", "consultas", "representantes", sugiriendo enfoque en aspectos organizativos y participación democrática.
- **Clúster Morado:** "compañeros", "escolares", "sindicatos", "apoderados", indicando relaciones y roles dentro de la comunidad escolar en el contexto democrático.
- **Clúster Verde:** "elecciones", "enseñanza", "democrática", "consejos", revelando conceptos fundamentales de la democracia en el ámbito educativo.

Este gráfico proporciona una instantánea visual de la diversidad y organización de términos relacionados con la democracia, facilitando la identificación rápida de patrones y asociaciones clave en cada clúster.

Por último se obtiene un gráfico de distribución de respuestas democracia:

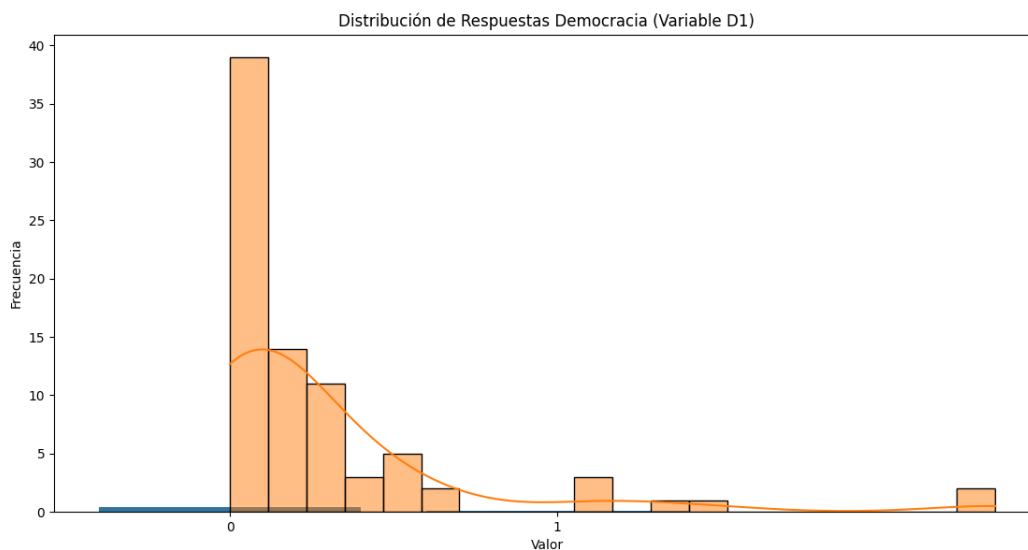


Gráfico 10: "Distribución de Respuestas democracia"

El gráfico sugiere que la percepción sobre la democracia entre los participantes tiende a ser positiva para cada grupo significativo, pero también hay variabilidad en las respuestas, con presencia de subgrupos menos positivos. La asimetría en la distribución destaca la polarización de las opiniones entre los extremos de la variable D1.

3.3 La Democracia desde la Perspectiva Estudiantil: Manifestaciones y Características.

Desde la perspectiva estudiantil, la democracia se manifiesta como un espacio de participación activa y construcción colectiva. Las características fundamentales de esta experiencia democrática en el ámbito estudiantil incluyen la libre expresión de ideas, la toma de decisiones consensuada y el respeto a la diversidad.

Las manifestaciones estudiantiles son vehículos esenciales para la expresión de inquietudes y demandas. Los estudiantes encuentran en las protestas y movilizaciones un medio para visibilizar sus preocupaciones, ya sea en temas académicos, sociales o políticos. La democracia estudiantil se nutre de la capacidad de los jóvenes para articular sus voces y ser agentes de cambio.

La participación activa se refleja en la involucración en procesos de toma de decisiones. Los estudiantes, a través de asambleas y representantes, contribuyen al diseño de políticas educativas y a la configuración del entorno académico. Esta participación no solo fortalece la democracia interna de las instituciones educativas, sino que también promueve un sentido de responsabilidad cívica entre los estudiantes.

La democracia estudiantil también se caracteriza por el respeto a la diversidad de opiniones y perspectivas. En un ambiente democrático, se fomenta el diálogo constructivo y la tolerancia hacia puntos de vista diversos. Esta apertura a la diversidad contribuye a la formación de ciudadanos críticos y comprometidos con la convivencia democrática, la democracia desde la perspectiva estudiantil se manifiesta a través de manifestaciones activas y participación en la toma de decisiones, destacando la importancia del respeto a la diversidad de opiniones como pilar fundamental de una educación democrática.

4. Pregunta 4

4.1 Argumentación del modelo utilizado.

Para abordar la relación entre las respuestas de los estudiantes, la edad y el género, se implementó un enfoque integral que combinó técnicas de clustering y reducción de dimensiones con análisis estadístico y visualización. K-means y DBSCAN fueron empleados para el clustering, destacándose por su capacidad para identificar patrones y grupos en datos no etiquetados. La selección del número óptimo de centroides mediante el método del codo en K-means y la capacidad de DBSCAN para manejar grupos de formas irregulares contribuyeron a la robustez del análisis de clustering.

Se incorporó un análisis estadístico, para evaluar las diferencias en las respuestas según la edad y el género. La información demográfica se integró a los modelos de documentos existentes, permitiendo contextualizar las respuestas e identificar patrones estadísticamente significativos. Este enfoque estadístico robusto fortaleció la comprensión de la posible influencia de la edad y el género en las respuestas de los estudiantes.

Además, se aplicaron técnicas de reducción de dimensiones, como PCA, para visualizar la variabilidad de respuestas en un espacio de menor dimensión, facilitando la interpretación y preservando la información relevante.

En síntesis, la combinación de K-means, DBSCAN, análisis estadístico y reducción de dimensiones ofreció un enfoque completo y eficiente para analizar las respuestas en función de la edad y el género, proporcionando una comprensión más profunda de las dinámicas subyacentes en los datos.

4.2 Resultados principales obtenidos.

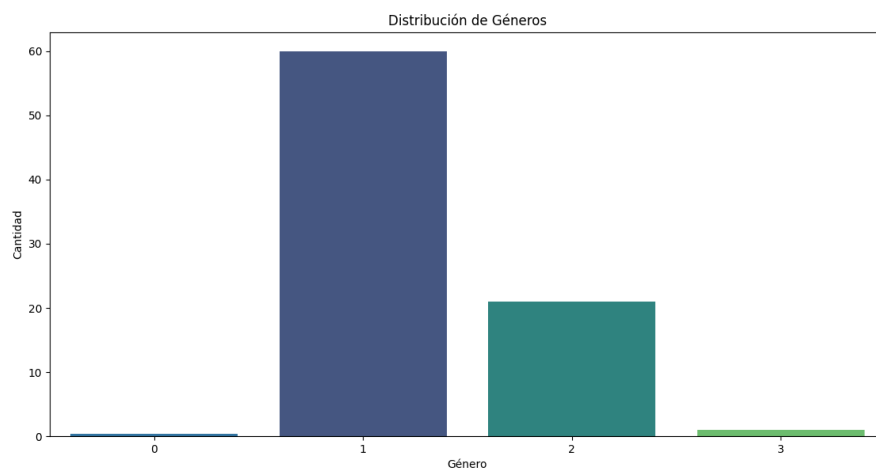


Gráfico 11: "Distribución de Géneros"

El gráfico de distribución de géneros que describes representa la cantidad de observaciones para cada género específico. Aquí hay algunas interpretaciones basadas en la descripción de

los ejes y los valores proporcionados, los valores en el eje vertical representan la cantidad de observaciones para cada género, los valores en el eje horizontal representan las categorías o géneros. En tu descripción, mencionas que estos valores son 0, 1, 2, y 3., puntos específicos en el eje horizontal son 1 y 2, donde se observan cantidades que exceden los valores típicos en otros géneros, En el punto 1 del eje horizontal, la cantidad alcanza 60, lo que sugiere que hay muchas observaciones para ese género en particular en comparación con los otros. En el punto 2 del eje horizontal, la cantidad es 20, lo que indica una cantidad moderada de observaciones para ese género, el gráfico muestra una distribución desigual de observaciones entre los géneros, el género representado en el punto 1 tiene una presencia significativamente mayor en comparación con los demás géneros. El género en el punto 2 también tiene una presencia, pero es moderada en comparación con el género en el punto 1.

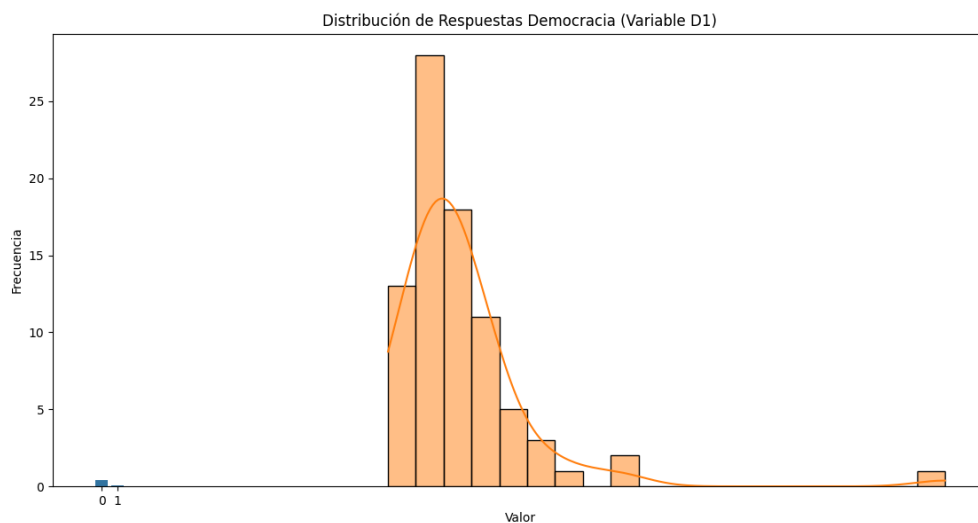


Gráfico 12: "Distribución de la variable respuestas democracia"

El gráfico 12 revela la distribución de géneros, destacando una notoria frecuencia más alta asociada con una categoría específica en el eje X, posiblemente vinculada al valor 1. A medida que avanzamos en las categorías, la altura de las barras disminuye progresivamente, señalando una distribución asimétrica.

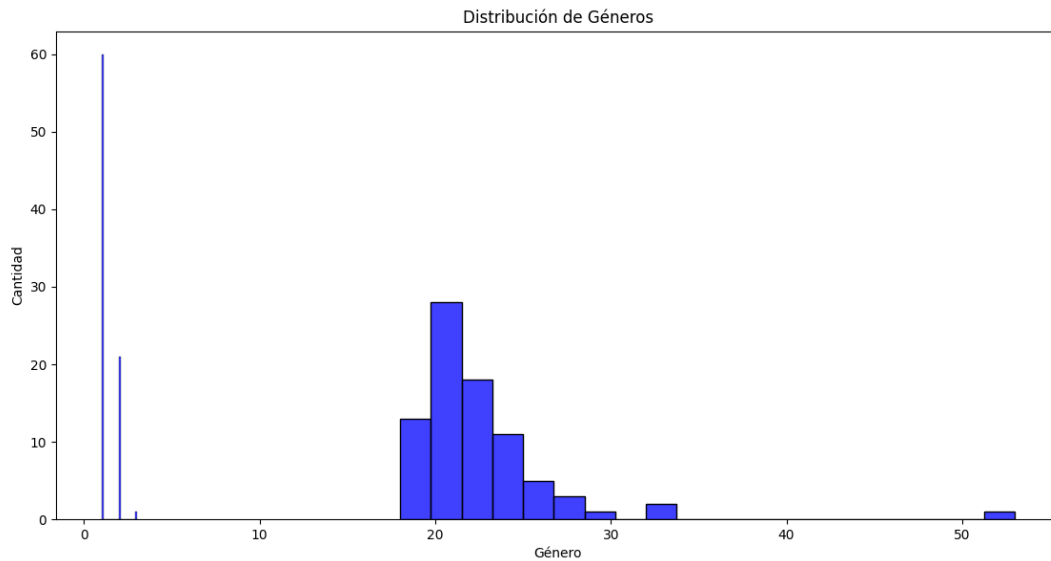


Gráfico 13: “Distribución de géneros”

El gráfico 13 muestra una distribución de géneros en relación con la cantidad de participantes, destacando dos grupos principales con la mayor presencia. Sin embargo, la distribución desigual se evidencia al observar que la segunda barra principal, aunque alta en cantidad, está concentrada en una franja específica en lugar de abarcar un rango más amplio. Otra barra significativa indica un grupo importante, pero su presencia se limita a una sección específica. A medida que avanzamos en los valores de la variable de género, las barras tienden a disminuir gradualmente, revelando variaciones en la representación de géneros en el conjunto de datos y resaltando la prominencia de ciertos grupos en comparación con otros.

4.3 Diversidad de Respuestas: Edad y Género como Factores Influyentes.

La diversidad de respuestas se ve significativamente influenciada por factores como la edad y el género. En el ámbito de la edad, se observa que diferentes grupos etarios tienden a tener perspectivas distintas ante diversas situaciones. La experiencia acumulada a lo largo de los años puede generar enfoques más pragmáticos o reflexivos, mientras que los jóvenes suelen aportar una mirada fresca e innovadora, el género también desempeña un papel crucial en la diversidad de respuestas. Las experiencias de vida, las expectativas sociales y las percepciones varían entre hombres y mujeres, influyendo en la manera en que abordan distintos temas. La diversidad de género enriquece las perspectivas, ya que cada grupo aporta vivencias y enfoques únicos.

Estos factores interactúan para generar un abanico amplio de respuestas. Las diferencias generacionales y de género contribuyen a la riqueza de ideas y soluciones en cualquier contexto. La comprensión de estas diversidades es esencial para fomentar la inclusión y la equidad, promoviendo la valoración de cada voz en la toma de decisiones y en la construcción de sociedades más diversas y justas.