



**Universidad del Desarrollo**  
Facultad de Ingeniería

## EL CHILE QUE QUEREMOS

Relación entre las necesidades del país versus las necesidades personales y familiares

POR: JOSÉ IGNACIO ZAMORA CIFRAS

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dra. MARIA PAZ RAVEAU MORALES

noviembre 2023

SANTIAGO

La curiosidad inicio mi camino y la  
perseverancia la transformó en mi destino.  
Gracias a mi familia y pareja por estar siempre  
presentes.

# TABLA DE CONTENIDO

## Contenido

<b>RESUMEN.....</b>	<b>1</b>
<b>1. INTRODUCCIÓN .....</b>	<b>2</b>
<b>2. HIPÓTESIS Y OBJETIVOS.....</b>	<b>3</b>
<b>3. DATOS Y METODOLOGÍA.....</b>	<b>4</b>
3.1. DATOS .....	4
3.2. METODOLOGÍA.....	5
3.2.1. PREPROCESAMIENTO DE LOS DATOS.....	5
3.2.2. LEMATIZACIÓN.....	6
3.2.3. OTROS ASPECTOS DEL PREPROCESAMIENTO.....	7
3.2.4. WORD EMBEDDINGS.....	8
3.2.5. CATEGORIZACIÓN .....	9
<b>4. RESULTADOS .....</b>	<b>10</b>
4.1. CLUSTERING .....	10
4.2. ANÁLISIS EXPLORATORIO .....	13
4.2.1. ANÁLISIS POR CLÚSTER .....	13
4.2.2. ANÁLISIS POR REGIÓN .....	16
<b>5. CONCLUSIONES .....</b>	<b>18</b>
<b>6. BIBLIOGRAFIA.....</b>	<b>19</b>
<b>7. ANEXOS.....</b>	<b>20</b>

## **Resumen**

Este proyecto analiza las percepciones y necesidades de los ciudadanos chilenos, basado en los datos de la iniciativa 'El Chile que Queremos'. Se centra en explorar cómo las necesidades nacionales se correlacionan con las personales o familiares, proporcionando insights sobre el estado socioeconómico y las preocupaciones predominantes en Chile. A través de un análisis detallado de las respuestas, se identifican temas clave y tendencias en las percepciones públicas.

Los resultados revelan variaciones significativas en las prioridades y preocupaciones entre diferentes sectores de la población, reflejando la diversidad de experiencias y expectativas en el país. Además, el estudio destaca la importancia de entender las necesidades personales en el contexto de los desafíos nacionales, subrayando cómo los factores personales y familiares influyen en la percepción de las necesidades nacionales.

## **1. Introducción**

"El Chile que Queremos" es una iniciativa lanzada por el gobierno de Chile con el objetivo de involucrar a los ciudadanos en un proceso de diálogo y reflexión acerca de los desafíos y prioridades del país. Esta iniciativa se enmarcó en un contexto de demandas sociales y políticas crecientes, buscando capturar una amplia gama de opiniones y perspectivas sobre el futuro de Chile.

La recolección de datos para "El Chile que Queremos" se realizó en un período crucial, posterior a las protestas sociales de 2019 que sacudieron el país. Estas manifestaciones, conocidas como el "Estallido Social", reflejaron el descontento generalizado con la desigualdad, el acceso a servicios básicos y la calidad de vida. En respuesta, el gobierno buscó crear un espacio para que los ciudadanos expresaran sus inquietudes, necesidades y propuestas para un Chile más inclusivo y equitativo.

El propósito de esta iniciativa era recoger las opiniones y propuestas de los chilenos sobre diversas cuestiones, desde educación y salud hasta equidad social y reformas políticas. Se utilizó una plataforma digital para permitir que miles de ciudadanos compartieran sus puntos de vista, lo que permitió una amplia participación y una recolección de datos representativa de las diversas realidades del país.

Los datos recolectados a través de "El Chile que Queremos" ofrecen una ventana única a las percepciones y aspiraciones de la población chilena en un momento crítico de su historia reciente. Esta información es invaluable para comprender mejor los desafíos actuales del país y para guiar las políticas y decisiones que moldearán el futuro de Chile.

## **2. Hipótesis y Objetivos**

### **Hipótesis**

Correlación entre Necesidades Nacionales y Personales/Familiares: Las percepciones de necesidades a nivel nacional están significativamente correlacionadas con las necesidades personales o familiares de los ciudadanos chilenos.

Diversidad Regional y Demográfica en Necesidades y Preocupaciones: Las necesidades y preocupaciones de los ciudadanos varían considerablemente en función de su ubicación geográfica y características demográficas.

### **Objetivo General**

Relación entre Percepciones de Necesidades y Factores Demográficos: Explorar y entender cómo las percepciones de necesidades nacionales y personales/familiares en Chile están interrelacionadas y cómo estas varían en diferentes regiones y entre distintos grupos demográficos.

### **Objetivos Específicos**

Categorización de Necesidades: Identificar y clasificar las principales necesidades expresadas por los participantes de 'El Chile que Queremos'.

Análisis de Correlación: Determinar la relación entre las necesidades nacionales y personales/familiares mediante técnicas de análisis de datos.

Estudio de Variaciones Regionales y Demográficas: Examinar cómo las percepciones de necesidades difieren entre distintas regiones y demografías en Chile.

## 3. Datos y Metodología

### 3.1. Datos

El Dataset “El Chile que queremos” cuenta con columnas de ID de grupos de personas y datos demográficos como la Región, Comuna, Lugar y Dirección, Edad, Nivel educacional. Las respuestas son de carácter individual y están organizadas en columnas categorizadas por número de pregunta y número de individuo según el grupo donde:

- P1\_A: Emociones / Sentimientos / Sensaciones
- P1\_B: Explique lo mencionado
- P2\_A: Necesidad que enfrenta el país
- P2\_B: Explique lo mencionado
- P2\_C: Urgencia
- P3\_A: Necesidades que enfrento personalmente o que existen en mi hogar o familia
- P3\_B: Explique lo mencionado
- P4\_A: Necesidades del país identificadas
- P4\_B: Rol del Estado (Describa)
- P4\_C: Actor social (empresa, organizaciones sociales, medios de comunicación, comunidad, etc.)
- P4\_D: Rol del actor social (Describa)
- P5: ¿Cuál es mi contribución personal para construir el Chile que queremos?

Y otras columnas adicionales de menor relevancia.

Este dataset consta de 13.325 registros y 289 columnas, abarcando una amplia gama de aspectos vitales para el estudio.

## **3.2. Metodología**

### **3.2.1. Preprocesamiento de los Datos**

En el inicio del análisis del conjunto de datos, fue esencial realizar una reorganización y depuración de las columnas para optimizar el proceso. Este paso implicó una evaluación detallada de cada columna para determinar que se utilizaría en el análisis, su relevancia en relación con la pregunta de investigación central y subyacentes. La eliminación de datos irrelevantes no solo simplificó el análisis, sino que también aseguró una mayor precisión y enfoque en los aspectos críticos del estudio, alineando los datos disponibles con los objetivos específicos del análisis.

### **3.2.2. Lematización**

En el análisis del texto recogido por "El Chile que Queremos", se implementó la técnica de lematización, un proceso clave en el procesamiento del lenguaje natural. La lematización implica reducir las palabras a su forma base o lema, lo que permite agrupar diferentes formas de una misma palabra para un análisis más consistente. Este paso es crucial para consolidar variantes lingüísticas y facilitar una interpretación precisa del contenido textual.

Posteriormente, y tras la decisión de trabajar con las columnas de datos que responden a las preguntas de investigación (P2\_A; P2\_B; P3\_A; P3\_B), se utilizaron técnicas de lematización. Concretamente, se empleó un modelo preentrenado de spaCy, "es\_core\_news\_md", que es el modelo de tamaño intermedio y demuestra un buen desempeño en comparación con otros modelos preentrenados en español.

#### **Desafíos en la Lematización y Estrategias Adoptadas**

Al principio, se incluyeron todas las categorías gramaticales (Part of Speech) en el proceso de lematización del proyecto "El Chile que Queremos". Sin embargo, se observó que palabras con múltiples usos en diferentes contextos, como 'Mejores', 'Justo', y 'Abandono', presentaban desafíos, ya que aparecían en diversos temas, lo que podía complicar la interpretación de los datos y la agrupación por temas (clustering).

Para abordar este problema, se experimentó restringiendo la lematización solo a sustantivos y auxiliares, este último, dado que palabras como 'afp' que para este análisis son valiosas, spaCy las interpretaba como un auxiliar. Aunque esta decisión implicó una cierta pérdida de información, resultó en una mayor precisión en los datos, logrando un balance satisfactorio para el objetivo del estudio.

### **3.2.3. Otros aspectos del preprocesamiento**

#### **Limpieza y Estandarización del Texto**

Después de la lematización, nos centramos en la limpieza y estandarización del texto, una etapa fundamental para garantizar la calidad de los datos. Este proceso implicó varios pasos críticos:

Eliminación de Puntuaciones y Caracteres Especiales: Se eliminaron todas las puntuaciones y caracteres especiales. Esto es importante porque tales caracteres pueden distorsionar el significado del texto o interferir con la tokenización y el análisis posterior.

Remoción de Espacios Extras y Números: Los espacios adicionales y los números se eliminaron para asegurar una uniformidad en el conjunto de datos. Los números a menudo se eliminan porque generalmente no contribuyen al análisis temático en estudios de este tipo.

#### **Tokenización**

El siguiente paso fue la tokenización, un proceso crítico en el análisis de texto. Se utilizó la función `word_tokenize` de la biblioteca NLTK para descomponer el texto en tokens individuales (palabras). La tokenización es vital porque convierte el texto largo y continuo en unidades más pequeñas y manejables, permitiendo un análisis más detallado.

#### **Eliminación de Stop Words**

Dado que se filtraron todas las palabras que no son verbos o auxiliares en el proceso de lematización, finalmente, se realizó la eliminación de las 'stop words', utilizando la lista de stopwords en español disponible en NLTK. Las 'stop words' son palabras que, aunque frecuentes en el lenguaje, tienen poco valor analítico en el análisis de texto (como preposiciones, conjunciones, etc.) y en este caso nos ayudara a eliminar los verbos auxiliares que no nos son útiles para el análisis.

#### **Importancia de Estos Pasos en el Análisis**

Estos procesos adicionales de preprocesamiento fueron cruciales para asegurar la calidad y precisión de los datos. La limpieza y estandarización del texto, seguida de la tokenización y la eliminación de palabras irrelevantes, sentaron las bases para un análisis textual más efectivo y enfocado, facilitando la identificación de patrones y temas relevantes en las respuestas del dataset.

### 3.2.4. Word Embeddings

Para el word embeddings, se seleccionó el modelo preentrenado FastText, orientado específicamente para el procesamiento del español. Este modelo se destacó por su capacidad de capturar con precisión la semántica y la sintaxis de las palabras, y por su eficiencia al manejar palabras fuera del vocabulario (OOV) a través de subunidades de texto (subpalabras). FastText pondera los vectores considerando no sólo las palabras completas, sino también los n-gramas de caracteres contenidos en ellas. Esto permite que el modelo genere representaciones vectoriales ricas incluso para palabras que no estaban presentes en el corpus de entrenamiento, aprovechando los n-gramas compartidos con palabras conocidas. Este enfoque es especialmente valioso en idiomas como el español, con una rica morfología y una gran cantidad de derivaciones de palabras, garantizando así una cobertura más amplia y una representación semántica más robusta en la vectorización del texto.

#### **Proceso y Aplicaciones**

**Word Embeddings:** El texto lematizado fue convertido en vectores numéricos usando FastText. Esta transformación es esencial para aplicar técnicas analíticas cuantitativas a datos cualitativos.

**Facilitación de Análisis Avanzados:** La vectorización permitió una identificación eficiente de patrones y tendencias en los datos, esencial para técnicas como el clustering.

**Interpretación Matizada de los Datos:** La representación vectorial de las palabras facilitó un análisis más profundo y detallado, crucial para la comprensión de las respuestas en el conjunto de datos.

Este paso de vectorización fue fundamental para preparar los datos para el análisis posterior, permitiendo un enfoque más sofisticado y matizado en el estudio de las respuestas recogidas en 'El Chile que Queremos'.

### **3.2.5. Categorización**

#### **Uso del Algoritmo KMeans**

En la fase de categorización, se optó por emplear el algoritmo KMeans de la biblioteca sklearn, una técnica de clustering ampliamente reconocida por su eficacia en la identificación de grupos homogéneos en un conjunto de datos. Este método fue elegido por su habilidad para manejar grandes volúmenes de datos y por su eficiencia en revelar patrones subyacentes en conjuntos de datos complejos.

#### **Preparación y Combinación de Datos**

Conversión a Arrays de Numpy: Inicialmente, se convirtieron las cadenas de texto vectorizadas en arrays de Numpy para facilitar su manipulación y análisis.

Combinación de Vectores de Preguntas: Los vectores generados para las respuestas a la pregunta 2 (Necesidades del país) y pregunta 3 (Necesidades personales o familiares) se combinaron en un único conjunto de datos. Este nos ayuda a verificar la hipótesis sobre cómo se relacionan las necesidades personales con las del país. Al agrupar las respuestas de ambas preguntas en los mismos clústers, pudimos ver claramente estas conexiones y entender mejor lo que valoran las personas.

#### **Ajuste y Aplicación del Modelo**

Configuración del Modelo KMeans: Configuré el modelo KMeans para identificar 22 clústeres distintos, basándose en la evaluación previa de la estructura de los datos (anexo: imagen 3.1 y 3.2). La elección de 22 clústeres fue guiada por un análisis exploratorio y la necesidad de lograr un equilibrio entre la granularidad y la coherencia de los grupos.

Reproducibilidad mediante Semilla Fija: Se utilizó una semilla fija en el algoritmo para asegurar la reproducibilidad del análisis, un aspecto crucial en la investigación científica.

#### **Asignación de Etiquetas y Almacenamiento de Datos**

Etiquetado de Datos: Tras ajustar el modelo, se asignó las etiquetas de los clústeres generados por KMeans a las filas correspondientes en el DataFrame original. Este paso fue esencial para categorizar y segmentar los datos en grupos significativos, lo que facilitó un análisis detallado y centrado en las respuestas.

Almacenamiento para Análisis Futuros: Los datos categorizados, ahora etiquetados con sus respectivos clústeres, se almacenaron en un nuevo archivo. Este archivo sirve como

base para análisis futuros y exploraciones más profundas de las tendencias y patrones identificados.

## 4. Resultados

### 4.1. Clustering

#### Análisis General:

El análisis de clustering ha proporcionado una visión cuantitativa de las prioridades y preocupaciones de los ciudadanos chilenos a través de la iniciativa "El Chile que Queremos". Los temas de mayor frecuencia y relevancia, reflejados en los clústers (Tabla 1.1 y 1.2) como pensiones, vivienda, salud y educación, resaltan la importancia de estas áreas en la vida cotidiana y la demanda pública para su consideración en agendas políticas y sociales. Otros temas como cambio, pensiones y justicia también emergen, revelando preocupaciones adicionales de significativa importancia. Este análisis de clustering ofrece una visión holística de la convergencia de preocupaciones en la población.

Tabla 1.1.- Clústers Necesidades del País

**Clúster 0:** transporte, diálogo  
**Clúster 1:** salud, salud salud  
**Clúster 2:** paz, abuso  
**Clúster 3:** cambio, cambio ley  
**Clúster 4:** educación, edad  
**Clúster 5:** pensión, pensión pensión  
**Clúster 6:** vivienda, vivienda salud  
**Clúster 7:** justicia, delincuencia  
**Clúster 8:** sueldo, sueldo afp  
**Clúster 9:** seguridad, equidad  
**Clúster 10:** afp, afps  
**Clúster 11:** medicamento, aumento  
**Clúster 12:** recurso, recurso estado  
**Clúster 13:** respeto, confianza  
**Clúster 14:** trabajo, trabajo sueldo  
**Clúster 15:** medio ambiente, sistema  
**Clúster 16:** empleo, desempleo  
**Clúster 17:** agua, educación calidad  
**Clúster 18:** salario, salario salario  
**Clúster 19:** jubilación, sistema pensión  
**Clúster 20:** igualdad, igualdad género  
**Clúster 21:** locomoción

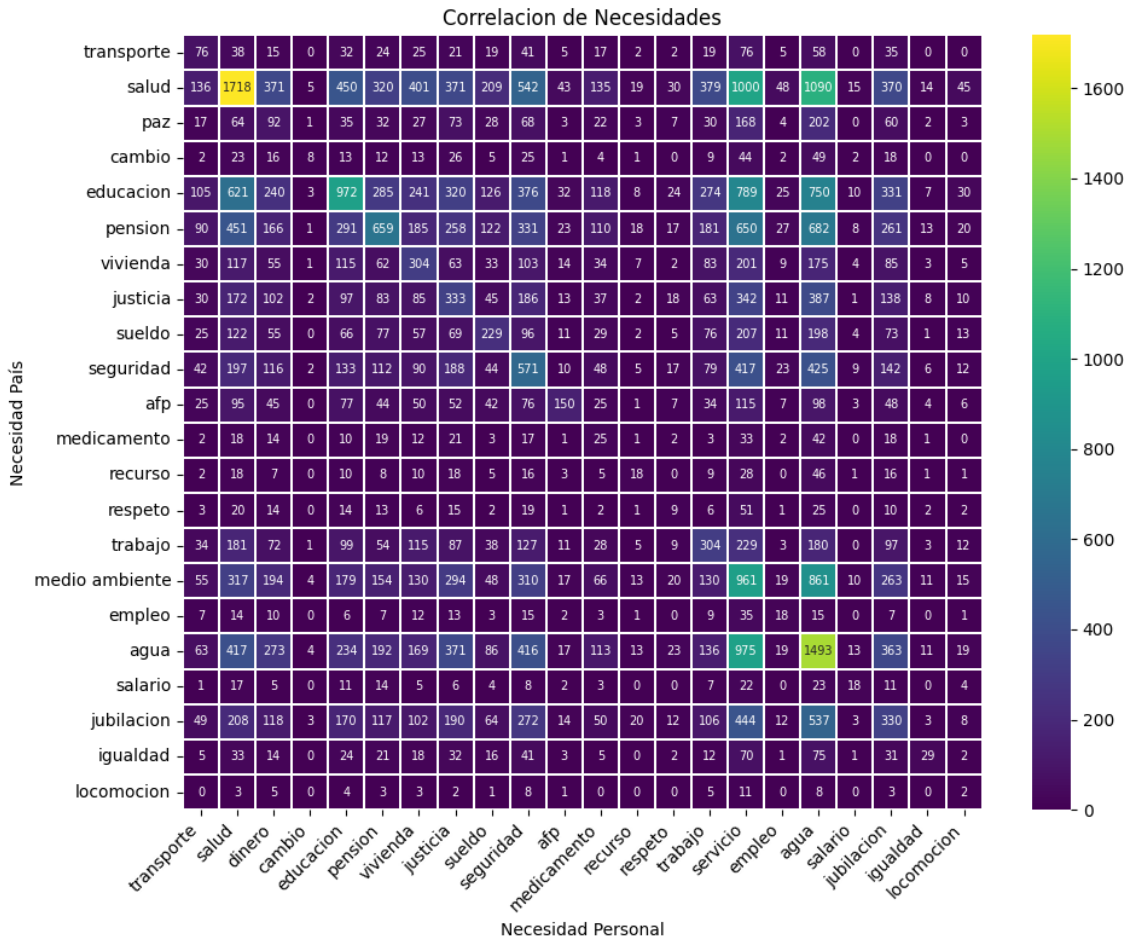
Tabla 1.2.- Clústers Necesidades Personales o Familiares

**Clúster 0:** transporte, diálogo  
**Clúster 1:** salud, salud salud  
**Clúster 2:** dinero, falta trabajo  
**Clúster 3:** cambio, cambio afp  
**Clúster 4:** educacion, edad  
**Clúster 5:** pension, pension pension  
**Clúster 6:** vivienda, viver  
**Clúster 7:** justicia, acceso vivienda  
**Clúster 8:** sueldo, sueldo sueldo  
**Clúster 9:** seguridad, estabilidad  
**Clúster 10:** afp  
**Clúster 11:** medicamento, endeudamiento  
**Clúster 12:** recurso, curso  
**Clúster 13:** respeto, hora  
**Clúster 14:** trabajo, trabajo mujer  
**Clúster 15:** servicio, medio ambiente  
**Clúster 16:** empleo, desempleo  
**Clúster 17:** agua, ingreso  
**Clúster 18:** salario, sala  
**Clúster 19:** jubilación, alimentación  
**Clúster 20:** igualdad, igualdad género  
**Clúster 21:** locomoción, locomoción ser

### Análisis Específico:

El heatmap (Imagen 1.3) generado como parte del análisis de los datos de "El Chile que Queremos" revela varios insights específicos sobre cómo las necesidades personales/familiares se reflejan en las preocupaciones nacionales en los temas principales. La presencia de altos valores en la diagonal principal, especialmente en categorías como salud, educación y seguridad, sugiere una fuerte correlación directa entre las necesidades personales y la percepción de lo que se requiere a nivel nacional. Esto implica que, para muchos chilenos, las preocupaciones personales no son distintas de las nacionales; hay una búsqueda de mejoras en ámbitos que afectan su vida diaria y que también ven como fundamentales para el progreso del país.

Imagen 1.3 Heatmap de coocurrencias Necesidades Personales vs Necesidades País.



Además, la presencia de valores significativos fuera de la diagonal principal indica preocupaciones transversales que conectan diferentes necesidades personales con áreas específicas de enfoque nacional. Por ejemplo, la conexión entre categorías nacionales como 'agua' o 'medio ambiente' con necesidades personales como 'servicio' puede reflejar un reconocimiento de la necesidad de mejora en los servicios públicos entregados por el país.

Este análisis específico del heatmap también destaca la importancia de considerar la heterogeneidad en las percepciones y necesidades a nivel regional y demográfico. La concentración de respuestas en ciertas áreas sugiere puntos de partida para políticas diferenciadas que puedan abordar de manera más efectiva las preocupaciones particulares de diferentes grupos dentro de la población.

Vemos en la tabla (Tabla 1.4) que del total de ocurrencias solo un 10,4% identifican la misma necesidad tanto personal como país, es decir, 1 de cada 10 individuos siente la misma necesidad en ambos ámbitos.

Tabla 1.4.- Porcentaje de coocurrencias

Cluster	Coocurrencias	Total Ocurrencias	Porcentaje
0	76	1.233	6.2%
1	1.718	10.857	15.8%
2	92	2.848	3.2%
3	8	300	2.7%
4	972	7.757	12.5%
5	659	6.217	10.6%
6	304	3.261	9.3%
7	333	4.655	7.2%
8	229	2.369	9.7%
9	571	5.781	9.9%
10	150	1.231	12.2%
11	25	1.098	2.3%
12	18	344	5.2%
13	9	413	2.2%
14	304	3.339	9.1%
15	961	9.978	9.6%
16	18	407	4.4%
17	1.493	11.346	13.2%
18	18	245	7.3%
19	330	5.212	6.3%
20	29	525	5.5%
21	2	267	0.7%
<b>Total</b>	<b>8319</b>	<b>79683</b>	<b>10.4%</b>

## 4.2. Análisis Exploratorio

### 4.2.1. Análisis por Clúster

Luego de haberse generado los clústers haremos un análisis exploratorio de los resultados.

En este primer punto (Imagen 2.1) se ve que los principales clústers identificados, o, dicho de otro modo, los que mayor cantidad de respuestas acumulan a nivel país son los clústers 1 (Salud) y el 2 (Educación), mientras que a nivel personal o familiar son el 15 y el 17, ambos clúster sin ninguna característica distintiva pareciendo más bien una mezcla de temas variados como vemos a continuación:

Necesidades Personales:

Cantidad de Ocurrencias Clúster 15:

['servicio']	272
['medio', 'ambiente']	177
['tiempo']	157
['camino']	120
['espacio']	108
['casa']	96
['gasto']	94
['luz']	94
['falta', 'oportunidad']	94
['impuesto']	92

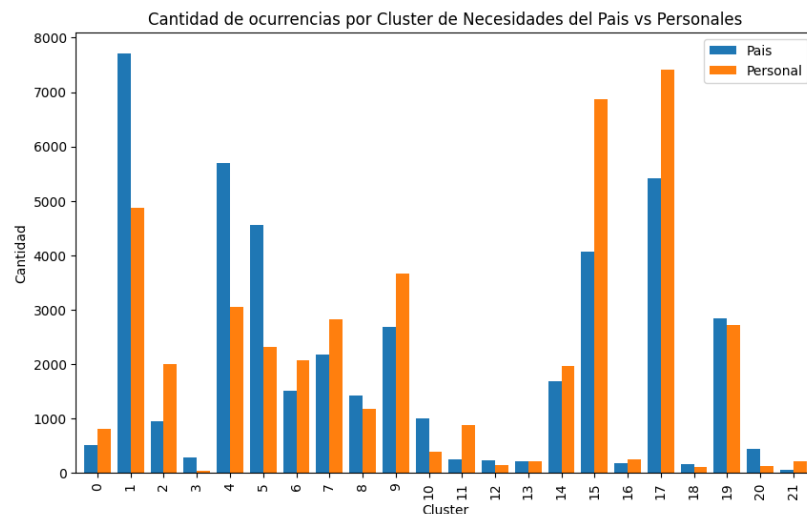
Name: P3\_A\_lemmatized\_tokens, dtype: int64

Cantidad de Ocurrencias Clúster 17:

['agua']	594
['ingreso']	273
['costo', 'vida']	149
['acceso', 'salud']	133
['educacion', 'calidad']	91
['calidad', 'vida']	82
['miedo']	70
['soledad']	54
['desigualdad']	54
['atencion', 'salud']	51

Name: P3\_A\_lemmatized\_tokens, dtype: int64

Esto puede deberse que necesidades personales debiese tener otro número de clústers o problemas con el preprocesamiento, ya sea vectorización o clustering con Kmeans.



No obstante, en el resto de clúster se captura la necesidad de manera más clara, vemos algunos casos:

- Para el clúster 1 Salud:

Se destaca un considerable factor de coocurrencia de 1718 menciones, reflejando la importancia de esta necesidad tanto a nivel personal como a nivel país, tal como se observa en el heatmap previamente mencionado (imagen 1.3). Al examinar algunas de las respuestas de los individuos dentro de este clúster, se puede obtener una comprensión más profunda de las preocupaciones específicas y las experiencias personales que subrayan la preeminencia de la salud en el discurso público y privado.

```

Necesidades Personales \
11 tengo que siempre estar esperando para que ...
15 -las horas de atención médica son pocas en el ...
21 mejorar atención y preferencia al adultos mayo...
31 mejorar atención y dar preferencia adulto mayor
62 tener un buen sistema de salud es básico para ...
65 familias enfermas.\n-demora en entrega de hora...
87 - familiares con sist. salud público, deben so...
93 sigue todo incluido\nbuen servicio la salud pu...
180 - abuela debe viajar a otra ciudad lejana a at...
186 auge: todo incluido. tuve cancer soy diagnosti...
196 - falta de especialista\n- visitas de médicos...
199 dar urgencia a la dotación de profesionales co...
253 vivo con adultos mayores se dificulta la atenc...
279 - la salud es muy cara, no todos tienen acceso...
330 falta de oportunidad a atención de calidad y p...
344 personal que tienen lugares cotizaciones previ...
349 -todos tengamos accesos al sistema de salud.\n...
356 necesidad de comprar medicamentos en caso de e...
402 -que disminuya el tiempo de espera en operacio...
428 difícil acceso costos esperas interminables\nm...

```

```

Necesidades País
11 equipar bien los hospitales\ny tener mas ...
15 largas listas de espera
21 mejorar el sistema de salud, por las ----- ...
31 mejor sistema de salud
62 una burla el funcionamiento de listas de espera
65 mala atención.\n-no se respeta horarios.\n- lo...
87 - largas esperas: para medicamentos; operacion...
93 mejorar \nequipamiento, dotación\n sistema d...
180 - hay pocos especialistas en el sistema públic...
186 mejorar infraestructura, equipamiento, dotació...
196 - debería ser prioridad y no hay mejoras en el...
199 ya que existe falta de especialistas con exper...
253 -mejorar el nivel de salud,igualitaria para todos
279 más médicos, incentivos para que se queden en ...
330 calidad\noportunidad\nges ampliadas\nlucro\nco...
344 falta merito profesionales ----- .....
349 mejor salud\nsin desigualdad
356 - mala atención en los hospitales, urgencias.\n...
402 -mas especialistas en la zona de arica y parin...
428 la salud publica necesita mejorar\ndebe ser de...

```

La presencia de términos como "listas de espera", "atención médica", "sistema de salud", y "adultos mayores" en ambas categorías indica que la mejora del sistema de salud es una prioridad tanto a nivel personal como en la agenda pública. Esta preocupación compartida podría reflejar una percepción de que el bienestar personal está directamente vinculado a la eficacia de los servicios y políticas de salud del país.

- Para el clúster 4 Educación:

Se agruparon 978 participantes que expresaron sus necesidades relacionadas con la educación, reflejando preocupaciones tanto a nivel individual como nacional. Esta categoría resalta la importancia atribuida a la educación por los encuestados, evidenciando un consenso sobre la necesidad de abordar y mejorar aspectos educativos que afectan personalmente a los individuos y, por extensión, al desarrollo y progreso del país en su conjunto.

Necesidades Personales

16 -me afecta el proceso actual de la tómbola, no...  
 24 NaN  
 36 - la mayoría de los ni siquiera son equitativo...  
 61 es la base para el funcionamiento de cualquier...  
 67 los profesores son de mala calidad al igual qu...  
 81 que sea más completa y no sigan eliminando asi...  
 179 - hermanos menores y primos deben viajar a otr...  
 201 mejorar la calidad de la educación pública en ...  
 249 inseguridad por los hijos, su futuro  
 280 - la educación es muy cara, y la educación sup...  
 347 - para un aporte al país\n- dignidad en su tra...  
 404 -que la beca indígena sea un derecho.\n-que la...  
 456 la calidad de educación comparada entre la pri...  
 487 gastos elevados por conceptos de mensualidades...  
 511 -mas bajos los pagos en las carreras universit...  
 536 sistema muy pobre para todo lo que tiene que a...  
 588 falta educación de calidad y gratuita  
 620 es difícil a que mi familia acceda a ésta ya q...  
 657 los profesores se pusieron flojos  
 690 sus nietos no pueden acceder a universidades p...

Necesidades País

16 malo el proceso de admisión a la educación bás...  
 24 que se fomente la educación valores, en la fam...  
 36 - una educación más responsable y equitativa p...  
 61 sin educación no hay nada que pueda funcionar  
 67 -no debe ser negocio.\n-gratis para todos en q...  
 81 que sea completa y de mejor calidad que favore...  
 179 - los cursos son de muchos alumnos, lo que dif...  
 201 mejorar la calidad de la educación pública, lo...  
 249 hay muchos profesionales jovenes endeudados co...  
 280 - fue al cae\n- mejorar el acceso  
 347 educación  
 404 -que la beca indígena sea un derecho para los ...  
 456 el cómo se asignan los beneficios comparado co...  
 487 educación de calidad.\neducación gratuita.\nau...  
 511 que no se pague tanto en las universidades  
 536 mejoras de infraestructura y profesores con vo...  
 588 falta de calidad en la educación y cantidad de...  
 620 ya que son muy altos costos para acceder  
 657 que existan colegios con profesores mejor prep...  
 690 poco acceso a educación de calidad

## 4.2.2. Análisis por Región

En el análisis regional, observamos tendencias distintivas en términos de las necesidades nacionales. Los temas recurrentes que emergen en las respuestas de los participantes son salud, educación y pensiones. Estos asuntos no solo son pilares fundamentales en los debates políticos actuales, sino que también reflejan preocupaciones persistentes y prioritarias entre los ciudadanos.

Respecto al análisis regional vemos una clara tendencia en lo que a necesidad país se refiere, los principales temas a las que las personas se refieren son a salud, educación y pensiones, temas que por lo general siempre están en los principales debates políticos.

Principales Necesidades País (Cantidad de Ocurrencias)

<p><b>Región Antofagasta:</b> salud, educación, pensión</p> <p><b>Región Arica y Parinacota:</b> salud, educación, pensión</p> <p><b>Región Atacama:</b> salud, educación, pensión</p> <p><b>Región Aysén del Gral. Carlos Ibáñez del Campo:</b> salud, educación, pensión</p> <p><b>Región Coquimbo:</b> salud, educación, pensión</p> <p><b>Región Los Ríos:</b> salud, educación, pensión</p> <p><b>Región Magallanes:</b> salud, educación, pensión</p> <p><b>Región Metropolitana:</b> salud, educación, pensión</p> <p><b>Región Tarapacá:</b> salud, educación, pensión</p> <p><b>Región Valparaíso:</b> salud, educación, pensión</p> <p><b>Región de la Araucanía:</b> salud, educación, pensión</p> <p><b>Región de los Lagos:</b> salud, educación, pensión</p> <p><b>Región del BioBío:</b> salud, educación, pensión</p> <p><b>Región del Libertador Gral. Bernardo O`higgins:</b> salud, educación, pensión</p> <p><b>Región del Maule:</b> salud, educación, pensión</p> <p><b>Región del Ñuble:</b> salud, educación, pensión</p>
---

Por otro lado, en cuanto a las necesidades personales o familiares, se observa también una marcada inclinación hacia la salud y la educación. Sin embargo, se nota una variación regional en cuanto a la vivienda, considerada una necesidad primordial en regiones como Antofagasta, Arica, Atacama, y la pensión, destacada en Maule, Los Lagos, Valparaíso. En la Región Metropolitana a diferencia de las demás se destaca el Sueldo como una necesidad particular que puede inducir a diferentes conclusiones como un alto nivel de precio de las viviendas, comida o arriendo.

#### Principales Necesidades Personales (Cantidad de Ocurrencias)

<b>Región Antofagasta:</b> salud, educación, vivienda
<b>Región Arica y Parinacota:</b> salud, educación, vivienda
<b>Región Atacama:</b> salud, educación, vivienda
<b>Región Aysén del Gral. Carlos Ibáñez del Campo:</b> salud, educación, vivienda
<b>Región Coquimbo:</b> salud, educación, trabajo
<b>Región Los Ríos:</b> salud, pensión, educación
<b>Región Magallanes:</b> salud, pensión, educación
<b>Región Metropolitana:</b> salud, educación, sueldo
<b>Región Tarapacá:</b> salud, educación, pensión
<b>Región Valparaíso:</b> salud, educación, pensión
<b>Región de la Araucanía:</b> salud, educación, trabajo
<b>Región de los Lagos:</b> salud, educación, vivienda
<b>Región del BioBío:</b> salud, educación, pensión
<b>Región del Libertador Gral. Bernardo O`higgins:</b> salud, educación, pensión
<b>Región del Maule:</b> salud, educación, pensión
<b>Región del Ñuble:</b> salud, educación, trabajo

## 5. Conclusiones

Este proyecto ha logrado avances significativos hacia el entendimiento de las necesidades nacionales y personales/familiares en Chile, así como su variación según región y demografía. Sin embargo, es importante reconocer que, aunque los resultados son satisfactorios, hay espacio para mejorar y profundizar en el análisis.

Identificación y Categorización de Necesidades: El análisis de los datos de "El Chile que Queremos" ha permitido categorizar eficazmente las principales necesidades expresadas por los ciudadanos. Temas como salud, educación, vivienda y seguridad se destacaron como preocupaciones clave. No obstante, la metodología utilizada para esta categorización podría beneficiarse de la aplicación de modelos de vectorización más robustos, lo cual podría proporcionar insights más detallados.

Correlaciones y Variaciones Regionales: El estudio ha identificado correlaciones relevantes entre las necesidades personales/familiares y nacionales, y ha destacado variaciones significativas entre diferentes regiones y demografías. Sin embargo, la exploración de estos aspectos podría enriquecerse con un análisis más detallado de la estructura lingüística de las respuestas, como el uso de sustantivos, adjetivos y otras categorías gramaticales para mejorar el proceso de clustering.

### Trabajo Futuro y Limitaciones

En futuros estudios, sería provechoso:

Analizar y Comparar Modelos de Vectorización: Explorar otros modelos de vectorización podría mejorar la precisión y la profundidad del análisis, permitiendo una mejor comprensión de las complejidades subyacentes en las necesidades y preocupaciones de los ciudadanos.

Experimentación con POS (Partes del Discurso): Manipular y jugar con diferentes categorías gramaticales durante el proceso de clustering podría proporcionar una perspectiva más matizada de cómo se expresan y priorizan las necesidades.

Análisis agregados: Explorar otras características demográficas o socioeconómicas para buscar patrones de respuesta más profundos.

## 6. Bibliografía

Gobierno de Chile. (2019, 22 de noviembre). Gobierno presenta “El Chile que Queremos”, iniciativa de diálogos y escucha social para avanzar en nuevas propuestas para el país. Gob.cl. Recuperado de

<https://www.gob.cl/noticias/gobierno-presenta-el-chile-que-queremos-iniciativa-de-dialogos-y-escucha-social-para-avanzar-en-nuevas-propuestas-para-el-pais/>

Explosion AI. (2023). Spanish · spaCy Models Documentation. Recuperado de

<https://spacy.io/models/es>

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of Tricks for Efficient Text Classification. arXiv preprint arXiv:1607.01759. Disponible en:

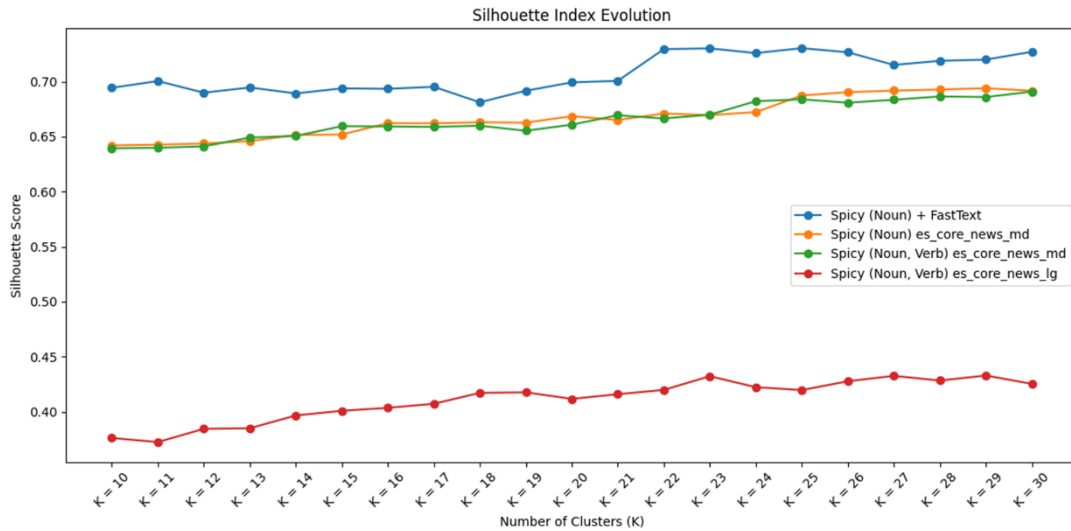
<https://arxiv.org/abs/1607.01759>

Introducción al Análisis Cluster. Consideraciones generales. Universidad de Granada.

<https://www.ugr.es/~gallardo/pdf/cluster-g.pdf>

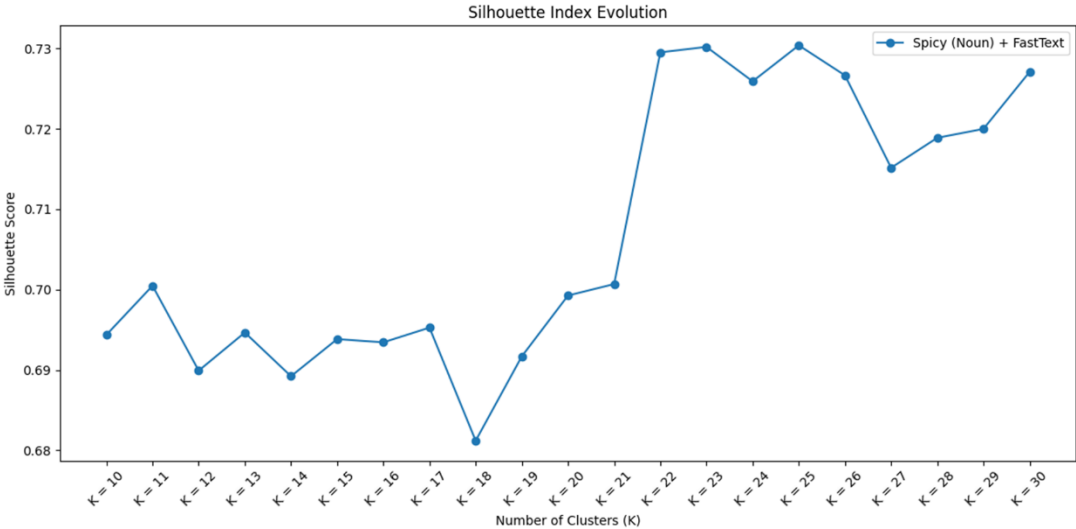
## 7. Anexos

Imagen 3.1.- Grafico y Tabla de Silhouette Score para pruebas de diferentes combinaciones de lematización y vectorización.

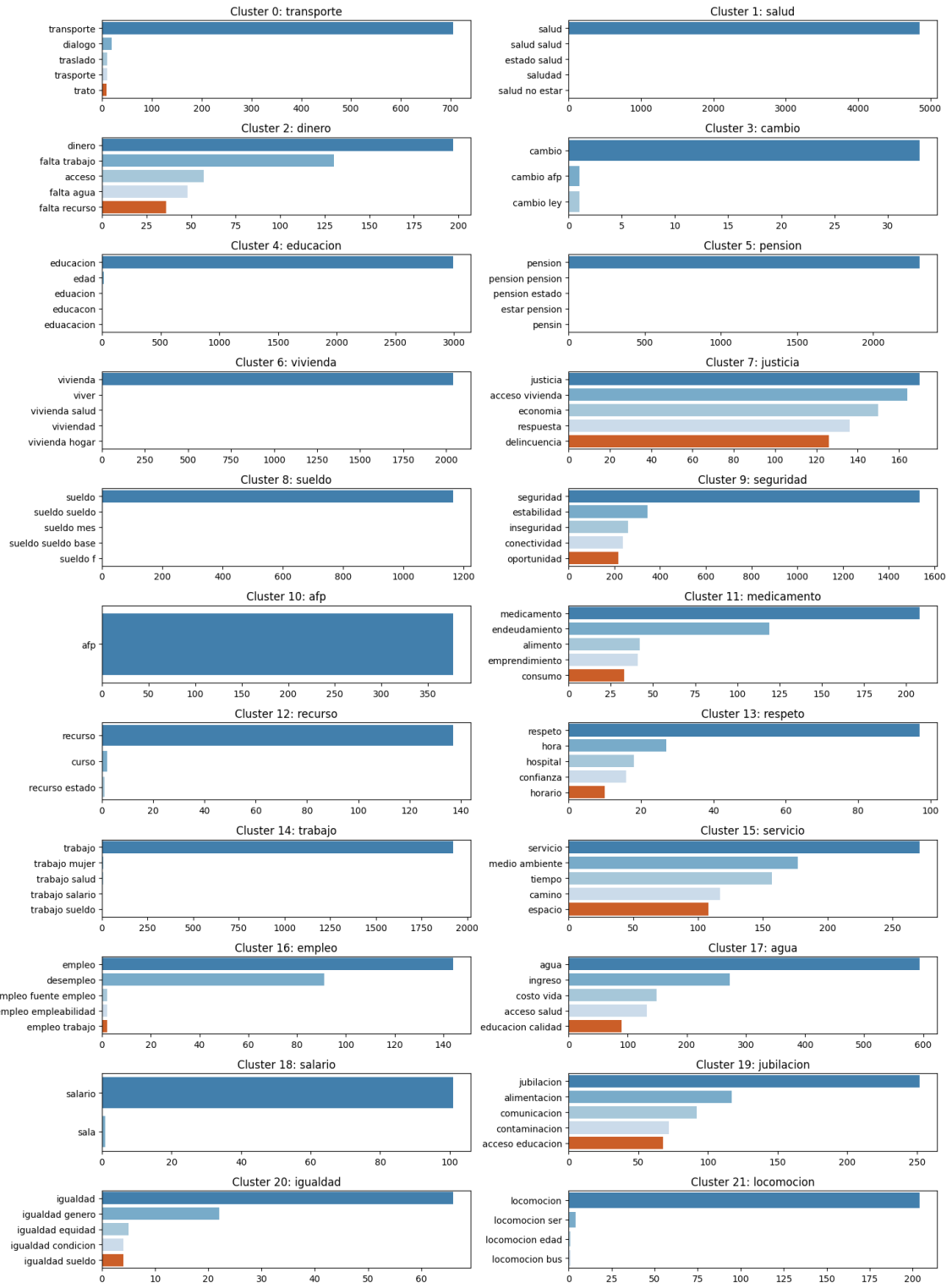


Lemma	Spicy	Spicy	Spicy	Spicy
POS	(NOUN, SYM)	(NOUN)	(NOUN, VERB)	(NOUN, VERB)
Word Embedding	FastText	Spicy	Spicy	Spicy
pipeline		es_core_news_md		es_core_news_lg
Cluster	Score	Score	Score	Score
K = 10	0,694	0,642	0,639	0,376
K = 11	0,700	0,643	0,640	0,372
K = 12	0,690	0,644	0,641	0,384
K = 13	0,695	0,646	0,649	0,385
K = 14	0,689	0,651	0,651	0,397
K = 15	0,694	0,652	0,659	0,401
K = 16	0,693	0,662	0,659	0,403
K = 17	0,695	0,662	0,659	0,407
K = 18	0,681	0,663	0,660	0,417
K = 19	0,692	0,663	0,655	0,418
K = 20	0,699	0,669	0,661	0,412
K = 21	0,701	0,665	0,669	0,416
K = 22	<b>0,730</b>	0,671	0,666	0,420
K = 23	0,730	0,670	0,670	0,432
K = 24	0,726	0,672	0,682	0,422
K = 25	0,730	0,687	0,684	0,420
K = 26	0,727	0,690	0,681	0,428
K = 27	0,715	0,692	0,683	0,432
K = 28	0,719	0,693	0,686	0,428
K = 29	0,720	0,694	0,686	0,433
K = 30	0,727	0,692	0,691	0,425

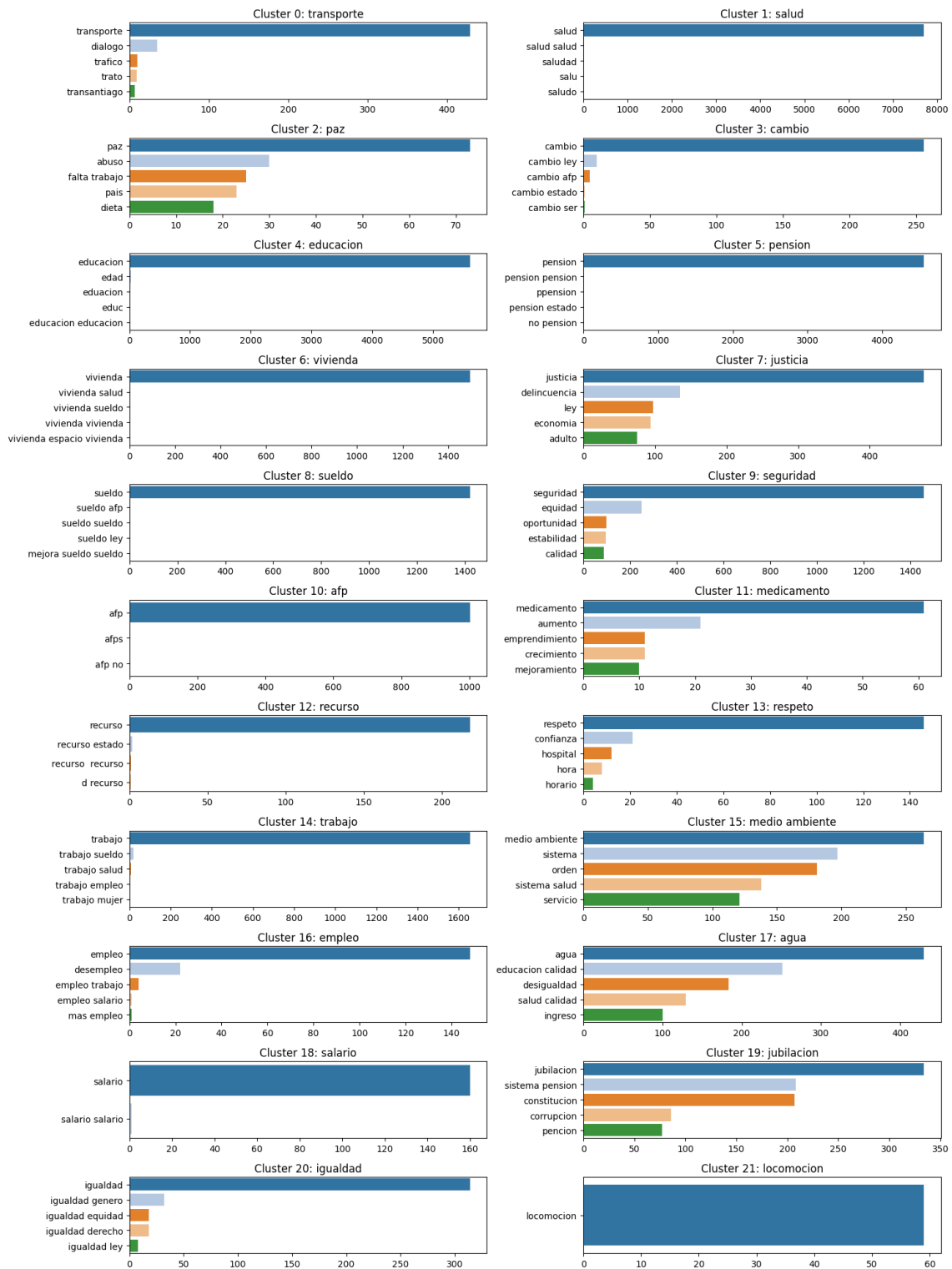
Imagen 3.2.- Silhouette Score para la determinación de los Clústers seleccionados.



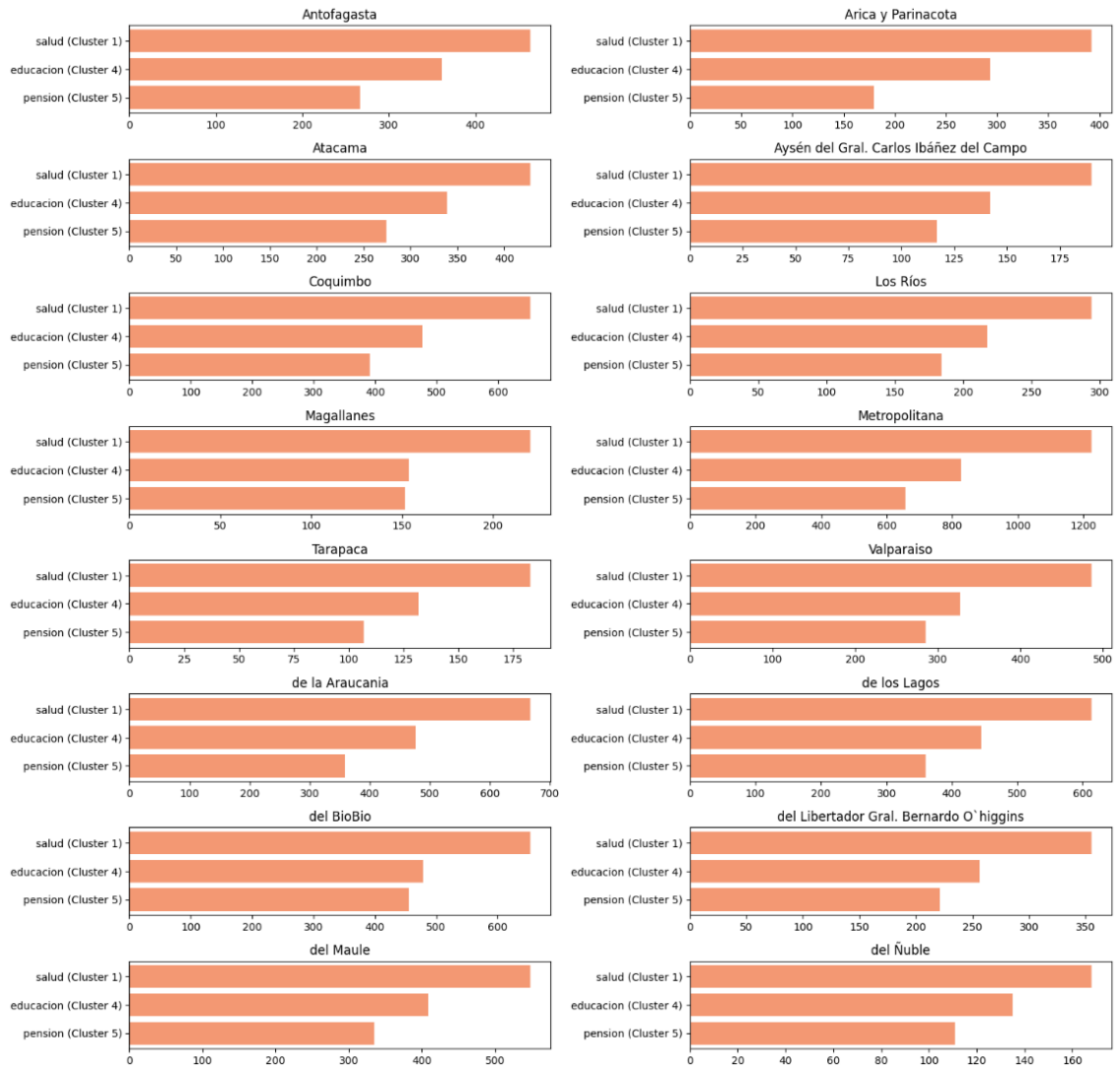
## Grafico de Clusters Necesidades Personales



## Grafico de Clusters Necesidades Pais



## Grafico de Cluster por Region de necesidades Pais



## Grafico de Cluster por Region de necesidades Personales

