Logo

Description automatically generated

[PROYECTO CAPSTONE]

[Uso de Manifesto Project para identificar patrones comunes en Programas de Gobierno para enfrentar la Corrupción]

POR: J. MATÍAS BUNSTER RABY

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. MARÍA PAZ REVEAU MORALES

[Diciembre 2024]

[SANTIAGO-CHILE]

AGRADECIMIENTO

[Escriba aquí la dedicatoria para su trabajo, recuerde que ésta es opcional]

[Escriba aquí los agradecimientos a las personas que colaboraron con usted en la realización de este trabajo]

Tabla de contenido

[Resumen 1](#_Toc55504658)

[1. Introducción 2](#_Toc55504659)

[2. Trabajo Relacionado 4](#_Toc55504660)

[3. Hipótesis y Objetivos 5](#_Toc55504661)

[4. Datos y Metodología 6](#_Toc55504662)

[4.1. Datos 6](#_Toc55504663)

[4.2. Metodología 6](#_Toc55504664)

[5. Resultados 7](#_Toc55504665)

[6. Conclusión y Trabajo Futuro 8](#_Toc55504666)

[Bibliografía 9](#_Toc55504667)

# Resumen

Escriba un resumen del documento

# Introducción

Este informe responde a los requerimientos de la Universidad del Desarrollo para obtener el grado de Magíster en Data Science a través de un “Proyecto Capstone”.  
El objetivo principal del estudio es identificar patrones comunes entre países tomando como base los programas de gobierno de candidatos presidenciales o equivalentes, ya sea que representen a un partido político o a una coalición, que permita clasificarlos frente a una problemática común como la Corrupción.

Los programas de gobierno son, en esencia, una declaración de intenciones. Una vez que un candidato triunfa en una elección, la implementación del programa enfrenta diversos factores, como la realidad económica y política del país, y la necesidad de negociación legislativa. En este estudio, el análisis se centra exclusivamente en los programas de gobierno, dejando como posible extensión futura el análisis de su implementación.

La Corrupción es un fenómeno global presente en las sociedades a lo largo de la historia. En las últimas décadas, se ha señalado un aumento en su penetración en diversas instituciones, lo que ha llevado a los gobiernos a implementar políticas específicas para combatirla; sin embargo, no se sabe si las prioridades políticas en torno a ella permiten agrupar países en patrones comunes. Considerando esta problemática, este trabajo analiza los programas de gobierno de un conjunto amplio de países en un período de tiempo determinado, en torno al fenómeno de la Corrupción, con el objetivo de identificar patrones comunes y agrupar a los países en comunidades con características similares, tomando como base la estructura inherente de los datos.

Este es un primer paso para generar conocimiento estructurado sobre este y otros fenómenos. En futuros trabajos, la combinación de estos resultados con datos empíricos sobre la implementación y resultados de políticas contra la Corrupción podría servir como una base para el diseño de políticas públicas comparadas.

Los datos utilizados en este estudio provienen de la plataforma Manifesto Project, que recopila programas de gobierno desde 1920 hasta 2023 de la mayoría de los países, facilitando análisis de preferencias y tendencias políticas. Se aplicó un enfoque de clustering y análisis de comunidades.

# Trabajo Relacionado

# Manifesto Project

(\*) Extraído de la página web de Manifesto Project (<https://manifesto-project.wzb.eu/information/documents/information>)

El ‘Manifesto Project’ (Proyecto Manifesto), analiza los manifiestos electorales de los partidos políticos para estudiar sus preferencias políticas. Desde octubre de 2009, el Proyecto Manifiesto ha sido financiado con una subvención a largo plazo de la Fundación Alemana para la Ciencia (DFG) como MARPOR (Manifiesto de Investigación sobre Representación Política). MARPOR continúa el trabajo del Grupo de Investigación Manifiesto (MRG 1979-1989) y el Proyecto Manifiestos Comparados (CMP 1989-2009). En 2003, el proyecto recibió el premio de la Asociación Estadounidense de Ciencias Políticas (APSA) al mejor conjunto de datos en política comparada.

MARPOR aborda la recopilación y el análisis comparativo de contenido de los manifiestos de los partidos con el apoyo de codificadores de más de 50 países diferentes. El equipo del proyecto coordina la generación y el análisis de los datos. El conjunto de datos del Proyecto Manifiesto proporcionado para el análisis de las preferencias políticas cubre más de 1.000 partidos desde 1945 hasta la actualidad en más de 50 países en los cinco continentes. El conjunto de datos principal del proyecto se actualiza dos veces al año. Proporciona acceso instantáneo a textos de manifiestos y datos analíticos de contenido y ofrece formas accesibles de explorar y visualizar fácilmente los datos y el corpus de texto.

El proyecto MARPOR tiene como objetivo analizar sustancialmente el papel de los partidos en diferentes etapas del proceso político y examina específicamente la calidad de la representación programática. Estudia la oferta programática de los partidos, la relación entre partidos y votantes, el papel de los partidos en el parlamento y la traducción de los programas de los partidos en resultados políticos.

Referencia: <https://manifesto-project.wzb.eu/information/documents/information>

# Hipótesis y Objetivos

3.1 Hipótesis

Los programas de gobierno reflejan patrones comunes que permiten agrupar países en comunidades según sus prioridades temáticas, mostrando diferencias consistentes entre comunidades en términos de énfasis político.

3.2 Objetivo General

Identificar patrones comunes en los programas de gobierno relacionados con la corrupción que permitan agrupar países en comunidades, analizando las diferencias y similitudes temáticas entre las comunidades a lo largo del tiempo.

**3.3 Objetivos Específicos**

3.3.1Analizar las variables del Manifesto Project relacionadas con la corrupción (Political Corruption, Law and Order, Traditional Morality) y cómo estas se distribuyen en los programas de gobierno de los países.

3.3.2 Detectar comunidades de países: Aplicar técnicas de clustering para identificar comunidades de países en base a sus programas de gobierno y analizar su composición en diferentes períodos de tiempo.

3.3.3 Calcular y visualizar las diferencias en las variables relacionadas con la corrupción y las áreas temáticas principales entre las comunidades detectadas.

3.3.4 Visualizar coincidencias entre países: Construir heatmaps y redes de coincidencias para entender cuántas veces los países comparten un mismo clúster en diferentes períodos

3.3.5 Caracterizar las comunidades en base a sus prioridades temáticas: Explorar si las comunidades detectadas reflejan diferencias estructurales relacionadas con características políticas, económicas o sociales.

# Datos y Metodología

## Datos

Los datos utilizados corresponde a los del dataset de Manifesto Project disponible a Marzo 2024. El dataset es el resultado de un trabajo de etiquetado de los programas de gobierno de partidos o coaliciones que postulan al cargo político principal de los países o a la cámara baja. En el caso de elecciones presidenciales, el criterio ha sido etiquetar aquellos programas que han obtenido el 5% de los votos o más en la primera vuelta. En el caso de postulaciones a la cámara baja, se han etiquetado aquellos programas que han conseguido 2 o más asientos en el caso de países de Europa central y del este, y América del Sur, y uno o más asientos en el caso de Australia, Japón, Nueva Zelanda, Norte América, Corea del Sur y Europa del oeste). También incluye los programas de partidos que fueron actores relevantes en el pasado pero que ya no cumplen con el criterio de mínimo de votos.

En términos generales el dataset reúne información de:

* 67 países (principalmente desde 1945 en adelante ó desde su primera elección democrática),
* 1.387 partidos políticos (según el criterio indicado)
* 856 procesos eleccionarios
* 5.151 programas de gobierno obtenidos de fuentes públicas.

4.1.1 Etiquetado

Cada programa de gobierno ha sido procesado por un equipo entrenado de codificadores (‘etiquetadores’ entrenados en un conjunto de reglas de etiquetado), de manera de asegurar la compatibilidad y la comparabilidad entre programas.

Cada programa es separado en las frases (“quasi sentences” según las Coding Instructions) que lo componen, y cada frase es etiquetada con una y sólo una etiqueta.

La pregunta central en el proceso de etiquetado es cuál es el mensaje que el partido ó el candidato están tratando de transmitir a los votantes, y cuáles son los problemas que considera importantes.

Las etiquetas posibles son respecto de 56 categorías que a su vez se organizan en 7 “Dominios”. Existen sub categorías para análisis especiales pero éstas son agregadas en las categorías principales.

Una vez que un programa ha sido etiquetado, se cuentan las etiquetas que recibió cada categoría y se divide por el número total de etiquetas. Este es el valor que queda registrado para cada programa, en cada categoría. Dada esta construcción, la suma de los valores de cada categoría de un programa cualquiera es 100% .

4.1.2 Países

A group of countries/regions with names

Description automatically generated

4.1.3 Categorías y Dominios

Nota: en lo sucesivo, cuando se mencione a una “variable” se debe entender lo mismo que una “categoría”.

A table with text and numbers

Description automatically generated

A blue and white box with black text

Description automatically generated

A blue and white box with black text

Description automatically generated

A screenshot of a table

Description automatically generated

A screenshot of a table

Description automatically generated

A screenshot of a table

Description automatically generated

A blue and white text with black text

Description automatically generated

4.1.4 Temporalidad

El registro más antiguo del dataset corresponde a Estados Unidos en 1920, siendo el único para el cual hubo datos ese año. El registro más reciente corresponde a Montenegro en Junio 2023.

Los países se fueron incorporando paulatinamente. De manera acumulada, en 1960 hay información de 25 países, en 1980 de 28, en 1990 de 45 y en 2003 de 67 países, según se puede apreciar en este gráfico:

Gráfico 1: Número de países por año presentes en la dataset

A graph with blue lines and numbers

Description automatically generated

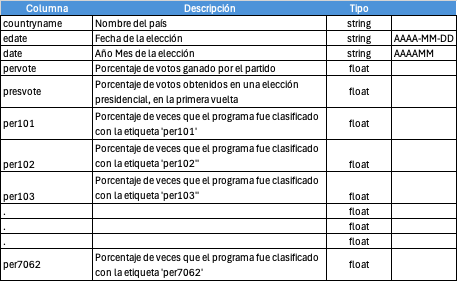
4.1.5 Columnas del dataset de Manifesto Project

El dataset de Manifesto Project contiene 5.151 registros con información de programas de gobierno según las siguientes columnas:

A document with text on it

Description automatically generated

Para el objetivo de este trabajo la columnas que se utilizó fue:



## Metodología

El método utilizado fue el siguiente:

* Conocer el dataset y validar su calidad respecto de una realidad conocida: la de Chile
* Selección de las variables a analizar para describir el fenómeno de la Corrupción
* Hacer análisis estadístico sobre dichas variables
* Preprocesamos el dataset a fin de asegurar que haya un registro por país para cada período en análisis. Seleccionaremos el período 2019-2022.
* Aplicaremos K-MEANS con las variables seleccionadas para identificar clusters. Con el método del codo identificamos el número óptimo de clusters con WCSS. Graficamos clusters.
* Aplicaremos K-MEANS para todos los períodos de 2003 en adelante en períodos de 3 años. Creamos la Matriz de Coincidencias que asocia a los países la cantidad de veces que están en un mismo cluster con otros países. Graficamos el heatmap de coincidencias
* Basados en la Matriz de Coincidencias, creamos una red en la cual los nodos son países y los enlaces son coincidencias en un cluster. En esta red identificaremos comunidades.
* Haremos estadística básica en cada comunidad y compararemos los resultados.

# Resultados

5.1 Conocimiento del dataset

Esta etapa ha sido relevante para poder validar los datos. La forma elegida para validar fue analizar el comportamiento de algunas variables del dataset para un período de tiempo reciente en torno a un aspecto relevante y contrastar con la realidad conocida de Chile.

5.1.1: Orden y Corrupción

A graph with lines and numbers

Description automatically generated

Comentario: Se puede apreciar claramente la importancia que ha ido adquiriendo este aspecto en Chile, variando desde 1,1 en 1989 a 5,2 en 2021 (4.8 veces). Esto ha sido impulsado por un fuerte aumento de las menciones en los programas presidenciales a dotar de mayores recursos a las policías y mayor rigurosidad en los tribunales. Debe mencionarse el caso MOP-GATE (financiamiento ilegal de la política) en 2003 lo que es consistente con el aumento de la variable per304. En forma similar, la variable per605\_1 (Law and Orde: Positive), toma relevancia a partir de 2009 con aumentos en 2014 y 2015, coincidentemente con los escándalos de financiamiento irregular de la política (casos Penta y SQM), y “PacoGate” en 2017.

5.1.2: Pueblos Originarios

A graph with green line

Description automatically generated

Comentario: per607\_3 (menciones a Multiculturalismo: Derechos Indígenas: Positivo) creció en menciones desde 1993 en adelante. Esto coincide con la promulgación de Ley Indígena y la creación de la CONADI (1993), y con el conflicto por la construcción de la central Ralco a fines de la década de 1990 e inicios de los 2000). Llama la atención la disminución en las menciones a partir de 2009 lo que se debe a otros aspectos que generaron más menciones.

5.1.3: Inmigración

A graph with lines and numbers

Description automatically generated

Comentario: Este aspecto está definido por menciones en favor de la inmigración y el multiculturalismo asociado a los inmigrantes. Se observa una tendencia creciente desde 1990 con un aumento fuerte a partir de 2009. Esto coincide con la llegada gradual de migrantes desde Perú, Bolivia y Colombia (1990), y luego la oleada migratoria desde 2015 en adelante desde Haití y Venzuela. En 2021 se dicta la nueva ley de Migración y Extranjería.

5.1.4: Derechos Humanos

A graph with different colored lines

Description automatically generated

Comentario: Se observa como la variable per201\_2 (Human Rights), crece en forma sostenida desde 1993 a 2013. En 1991 fue la publicación del informe Rettig y en 2012 se publicó la Ley Antidiscriminación (Ley Zamudio).

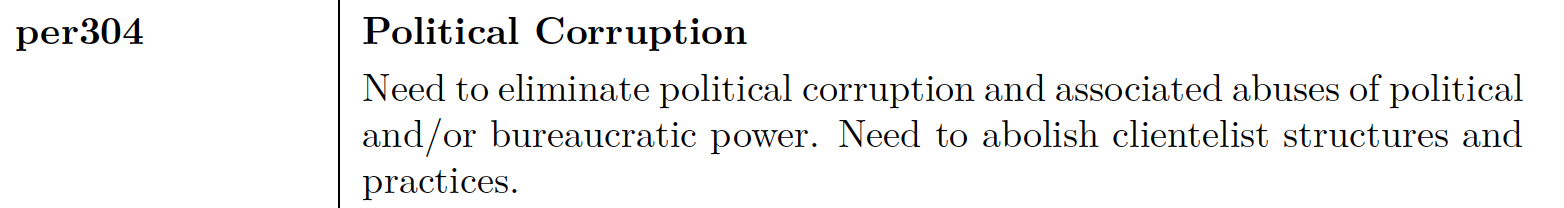
La conclusión de la validación del dataset como herramienta para el análisis es que si es válida, en el sentido de que lo que va reflejando algunas variables en el tiempo son consistentes con lo que ha pasado en la realidad en Chile.

5.2 Identificación de las variables para el análisis de programas con el foco en la Corrupción

Para analizar el fenómeno de la Corrupción, se seleccionó las siguientes variables:

* per304: Political Corruption
* per603: Traditional Morality: Positive
* per604: Traditional Morality: Negative
* per605: Law and Order: Positive

La descripción de cada una de estas variables según el Manifesto Project es la siguiente:

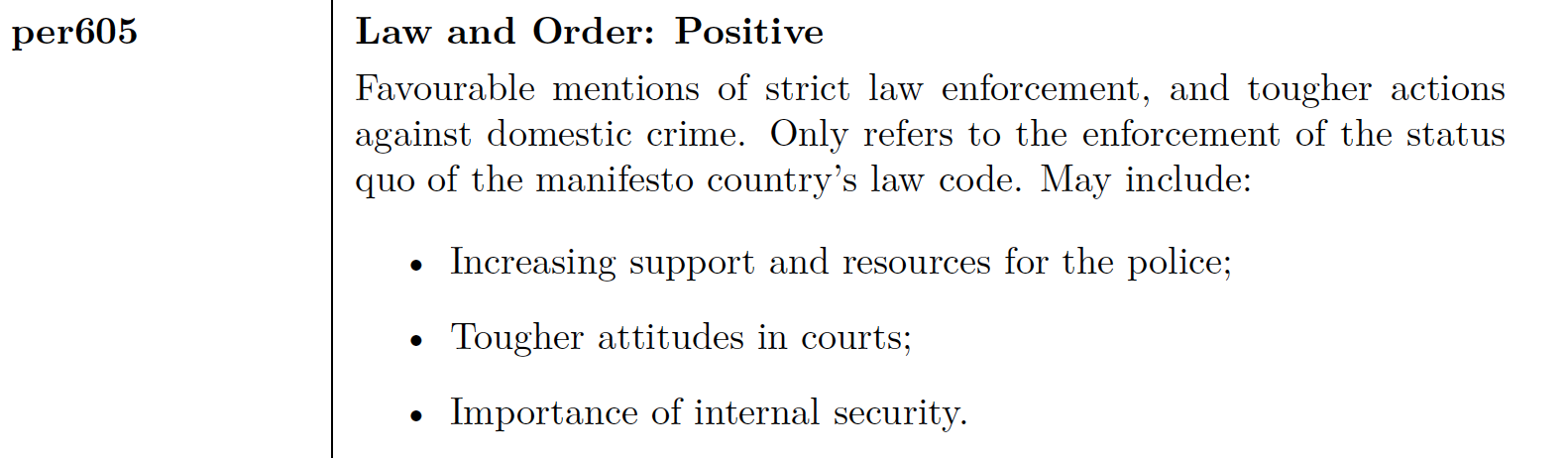


A close-up of a paper

Description automatically generated

A black text on a white background

Description automatically generated



5.3 Análisis exploratorio del dataset de las variables seleccionadas

El dataset del Manifesto Project es limpio y no contiene valores del tipo NaN, lo que reduce la necesidad de preprocesamiento para su análisis. La ausencia de problemas como datos mal codificados o valores inconsistentes permite trabajar de manera eficiente y garantiza la integridad de los resultados obtenidos. Esta calidad facilita la replicabilidad del estudio y refuerza la confianza en las conclusiones derivadas del análisis.

5.3.1 Distribución de valores en el dataset completo

Ahora se analizará la distribución de los valores de cada de las variables con las cuales analizaremos el fenómeno de la Corrupción. Se busca entender la distribución, valores característicos e identificar la existencia de outliers.

A screenshot of a graph

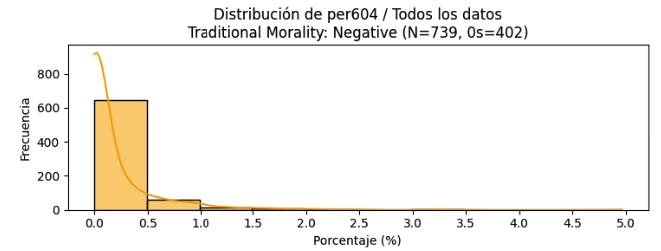
Description automatically generated

A graph with numbers and lines

Description automatically generated

A graph with numbers and a line

Description automatically generated



A graph with numbers and a line

Description automatically generated

Como se puede observar, los gráficos están cargados a la izquierda, en torno al cero.

Esta distribución toma esta forma principalmente por el hecho de que la suma de los valores en cada fila -por diseño del dataset-, es 100%; a esto se le denomina “datos composicionales” y su distribución se llama distribución Dirichlet. En una distribución Dirichlet el análisis de la contribución marginal de una de las variables sigue una distribución “Beta”, cuya forma es la siguiente:

A graph with a blue line

Description automatically generated

<https://statisticaloddsandends.wordpress.com/2021/04/20/marginal-distributions-of-the-dirichlet-distribution-and-the-aggregation-property/>

**5.3.2 Outliers**

Como se observa en los gráficos box-plot más abajo, da la impresión de que hay muchos outliers. Sin embargo, se debe considerar que dado que una función Beta está acotada en el rango [0, 1], la definición de tradicional de “outlier” en el sentido de valores que salen de un rango “normal”, se ve modificada en una distribución acotada. Lo que en una distribución no acotada interpretamos como un valor “extremo”, en la Beta puede ser simplemente la cola natural de la distribución.

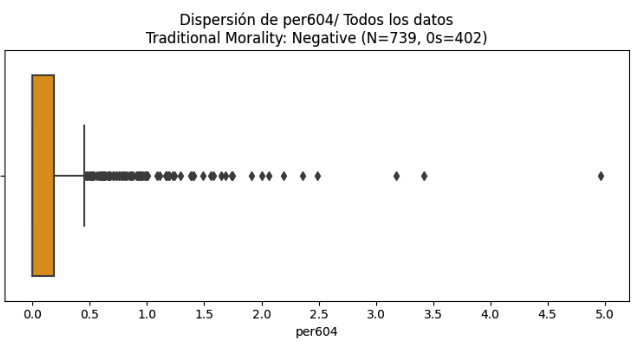
En conclusión, es correcto decir que muchos valores que parecen atípicos no son realmente “outliers” en el sentido usual.

A graph of a political corruption

Description automatically generated

A graph with numbers and dots

Description automatically generated



A graph with a bar graph and numbers

Description automatically generated with medium confidence

**5.3.3 Matriz de correlaciones**

Con el fin de identificar la existencia de correlaciones entre las variables seleccionadas se construyó la matriz respectiva:

A blue and red squares with white text

Description automatically generated

* La máxima correlación positiva es 0,21: esto signica que las variables tienen alguna relación en ciertos contextos, pero no son interdependientes de manera significativa en la mayoría de los casos.
* La máxima correlación negativa es -0,13: esto significa que hay competencia leve entre algunas variables, pero no es significativa. Esto puede deberse a que las variables representan temas conceptualmente distintos.
* La falta de correlaciones fuertes sugiere que las variables están actuando de manera relativamente independiente, lo cual es valioso para el análisis de clusters.

**5.4 Pre procesamiento de datos**

Las variables en análisis son numéricas y todas están en la misma escala.

Teniendo presente que utilizaremos el algoritmo de clasificación K-MEANS se asegurará que todas las filas sumen 1.

Además, cuando se procesen períodos de tiempo de más de 1 año, y ocurra que en dicho período un país tuvo más de una elección, se reducirá a un registro por año por país en el período, promediando los valores de las variables seleccionadas.

**5.5 Aplicación de K-MEANS para un período de 3 años**

Nota: Para el análisis, elegiremos el período de tiempo de 2006 a 2008. Posteriormente, dividiremos un período mayor de tiempo en sub períodos de 3 años a cada uno de los cuales se les aplicará K-MEANS.

K-MEANS es un algoritmo de clustering que agrupa los datos en un número pre definido de clusters, con el objetivo de minimizar la suma de las distancias al cuadrado (SSE: Sum of Squared Errors) entre cada punto de datos y el centroide de su cluster. De esta manera, K-MEANS requiere el input del número de clusters ‘k’.

Por otro lado, el Método del Codo es una técnica que ayuda a determinar el número óptimo de clusters (k) para K-MEANS. Para esto evalúa el valor de la suma de las distancias al cuadrado (SSE) a medida que aumenta k.

Para el análisis en el período de 2006 a 2008 el método del codo nos indica que k\_optimo=2 como se indica en el siguiente gráfico.

A graph with a blue line

Description automatically generated

Sin embargo, dado que la Corrupción es un fenómeno multi dimensional, aplicaremos K-MEANS con k=3 clusters.

El resultado de K-MEANS en el período 2006-2008 es el siguiente:

A graph with colored dots and numbers

Description automatically generated

A graph of a graph with a grid and dots

Description automatically generated with medium confidence

* Países en Cluster 0: Armenia, Australia, Colombia, Denmark, Finland, France, Georgia, Greece, Iceland, Montenegro, New Zealand, North Macedonia, Poland, Russia, Slovenia, South Korea, Ukraine, United States
* Países en Cluster 1: Argentina, Brazil, Costa Rica, Cyprus, Latvia, Lithuania, Serbia, Serbia, Turkey, Ukraine
* Países en Cluster 2: Austria, Austria, Bosnia-Herzegovina, Czech Republic, Estonia, Ireland, Italy, Netherlands, Romania, Slovakia, Spain, Sweden, Switzerland, Canada

Como se puede observar, con el algoritmo de K-MEANS con k=3 clusters en el período 2006 a 2008, se distinguen 3 clusters bien definidos.

**5.6 Aplicación de K-MEANS para un período largo de tiempo**

Acabamos de ver que en un período de tiempo los programas presidenciales de los países se pueden visualizar en clusters. En este paso lo que haremos es que tomaremos un lapso de tiempo de 20 años (2003 a 2022) y lo dividiremos en períodos de 3 años. Para cada período aplicaremos K-MEANS con k=3 clusters, y registraremos los países en cada cluster.

La elección del período entre 2003 y 2022 obedece a que elegir un período mayor (por ejemplo 1940 a 2022), va a sobre representar a los países que llevan más años en el Manifesto Project. Como se puede ver en el gráfico en 5.1.4, ya en el año 2000 hay más de 60 países informando, por lo cual elegimos 2003 a 2022.

**5.6.1 Un comentario sobre k=3 clusters.**

Como se mencionó, el fenómeno de la Corrupción es multidimensional lo cual sugiere que son más de dos (k\_optimo=2) grupos en los cuales se pueden clasificar los programas presidenciales de los 56 países. Es por esta razón que procesaremos K-MEANS con k=3. Por otro lado, a fin de evaluar la posible pérdida de calidad de los clusters, evaluamos una métrica de calidad para 1, 2, 3 y 4 clusters, conocida como “Silhoulette”. Esta métrica hace: xxx

El resultado se puede apreciar en el siguiente gráfico (mientras mayor es la métrica, mejor es la calidad):

A graph with different colored lines

Description automatically generated

Como se puede observar, el único período en que k=2 es evidentemente mejor que k=3 es en el período 2020 a 2022. En los demás, k=3 es similar a k=2; con esto concluimos que no hay una pérdida de calidad relevante si es que usamos k=3 clusters para lo que sigue de nuestro análisis.

**5.6.2 Matriz de Coincidencias**

Nota:

* el valor dentro de cada celda del heatmap indica cuántas veces esos dos países estuvieron en el mismo cluster.
* Sólo se muestran países que coinciden 3 o más veces en un mismo cluster

Por ejemplo, en el período de 2003 a 2022, Chile y Canadá estuvieron en un mismo cluster 4 veces.

A crossword puzzle with blue squares and red and white text

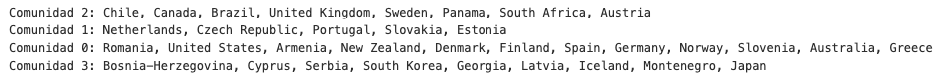
Description automatically generated

**5.6.3 Red de coincidencias**

Creamos una red de coincidencias con el paquete Networkx. Luego aplicamos el método de Louvain para identificar comunidades en la red. El resultado son 4 comunidades:

A diagram of a network

Description automatically generated



-

**5.6.4 Análisis de centralidades**

Como se puede observar en las centralidades, Australia (0.42) y New Zealand (0,36), son los principales nodos en cuanto a centralidad de Grado; es decir, están directamente relacionados con más conexiones directas con otro nodos en la red. Tratándose de programas de gobierno, esto podría indicar que comparten más similitudes temáticas con otros nodos.

En cuanto a Betweenness, New Zealand (0,40) y United Kingdom (0,35) son las principales centralidades de intermediación. Esto significa que facilitan la conexión entre nodos que no están directamente conectados. New Zealand es un nodo con más visibilidad.

Tratándose de similitudes temáticas, Australia y New Zealand podrían ser referentes globales o puntos de convergencia para ciertos temas o políticas que otros países adoptan o consideran relevantes.

United Kingdom, aunque menos directamente conectado, podría tener un rol histórico, cultural o político que lo posiciona como un enlace clave entre diferentes bloques de países.

A table of numbers with black text

Description automatically generated

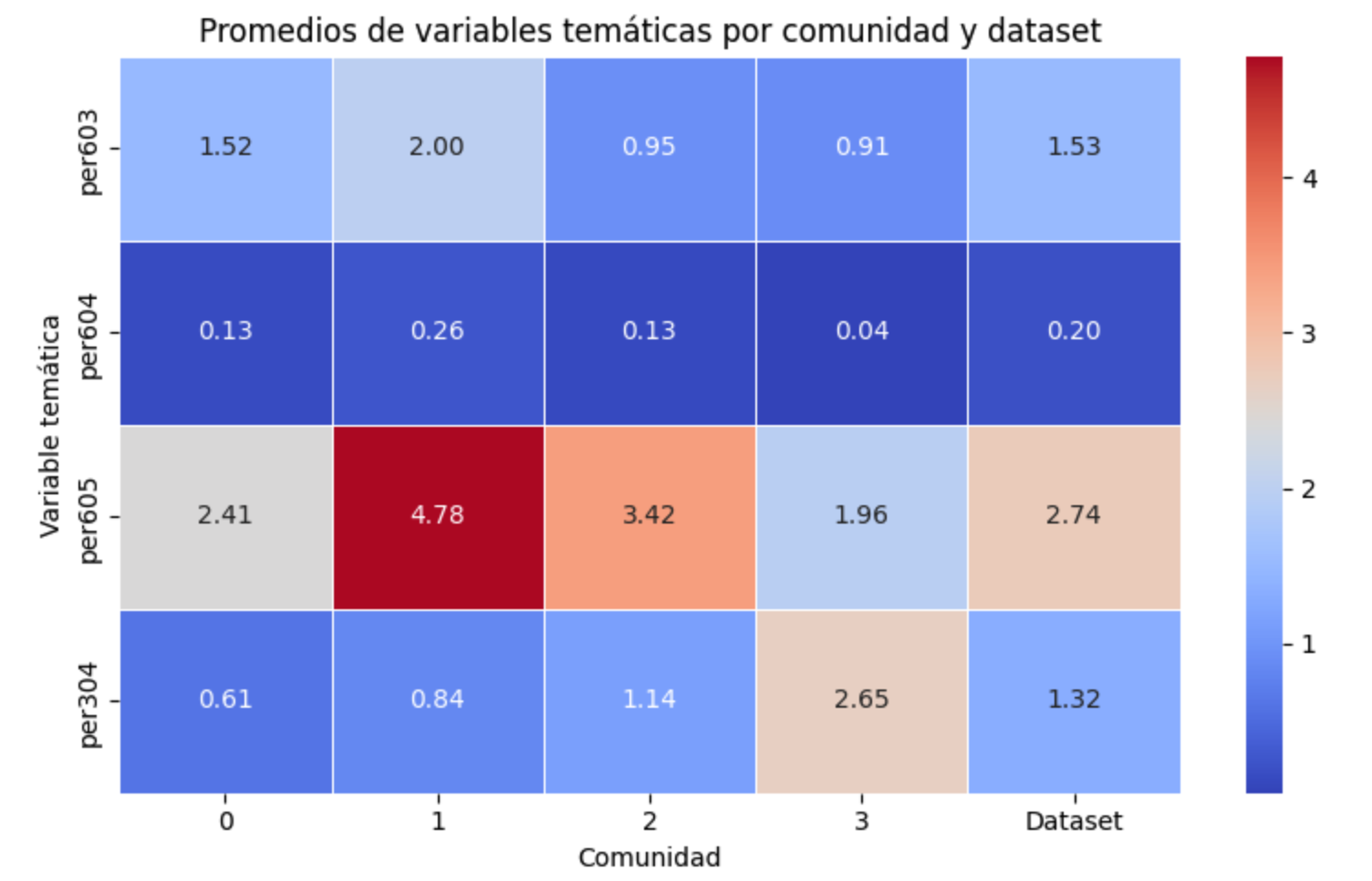
**5.6.5 Estadística básica en las Comunidades (promedios)**

Analicemos los promedios de las variables por comunidad de diferentes maneras para poder comparar.

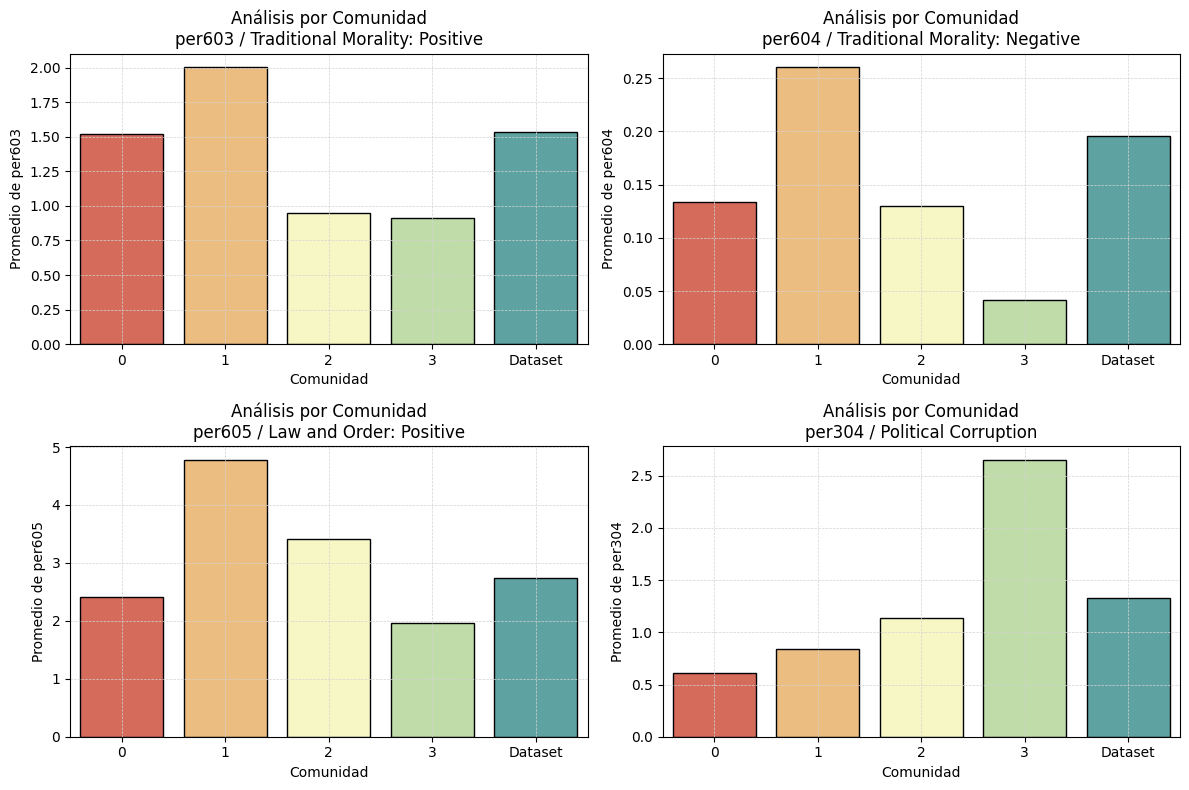
* Tabla de promedios por variable y comunidad

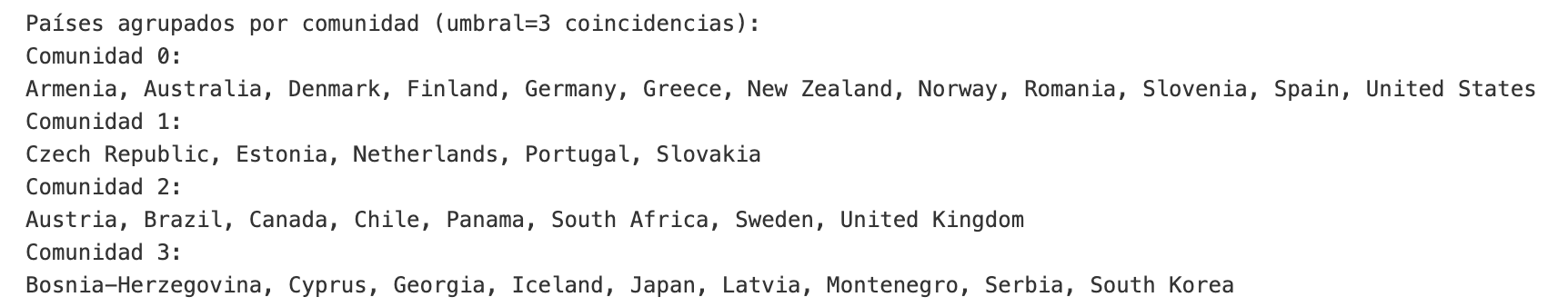


* Matriz de promedios por variable y comunidad



* Histogramas del promedio de cada variable, por comunidad, y promedio del dataset en el período en análisis.





**5.7 Análisis de Comunidades**

**5.7.1 Análisis Comunidad 0: “Democracias con liderazgo en transparencia y compromiso con el orden legal y social”**

* Países: Armenia, Australia, Denmark, Finland, Germany, Greece, New Zealand, Norway, Romania, Slovenia, Spain, United States

En esta comunidad la mención de per304 (Political Corruption) es la menor entre todas las comunidades y sustancialmente menor al promedio global (0.61 y 1.32 respectivamente). Además, en per605 (Law and Order: Positive) el promedio de 2.41 es de nivel medio entre las comunidades e inferior al promedio global de 2.74. Estas dos variables en niveles bajos indican que no son tan importantes en estos países como en las demás comunidades, en comparación con otras. Podemos desprender que en estos países el nivel de corrupción política es bajo y que son países donde el respeto a las leyes y el orden son fundamentales.

En per304 (Political Corruption) tiene el promedio más bajo de todos los cluster (0.61), muy por debajo del promedio del dataset (1.32). En per603 y per604 tiene promedios 1.52 y 0.13 respectivamente (siendo el segundo más alto en per603 entre comunidades), en comparación con el promedio global de 1.53 y 0.20 respectivamente. En per605, Law and Order, tiene un promedio de 2.41 por debajo del promedio global de 2.74.

Esta comunidad agrupa a países con democracias maduras que destacan por sus sistemas de gobierno estables y bajos niveles de corrupción.

En cuanto a las variables vinculadas a la moralidad (per603 y per604), sus promedios son igual al promedio globales o inferiores, no siendo países progresistas del todo.

**5.7.2 Análisis Comunidad 1: “Conservadurismo Social y Fortalecimiento de la Ley y el Orden”**

Países: Czech Republic, Estonia, Netherlands, Portugal, Slovakia

Entre los países de esta comunidad están República Checa, Estonia y Slovakia los cuales transitaron recientemente de gobiernos autoritarios a democracias participativas.

Esta comunidad tiene los promedios más altos en las variables per603, per604 y per605 (Moralidad Tradicional y Ley y Orden), muy por sobre las otras comunidades y por sobre el promedio del dataset, lo cual refleja prioridad en garantizar la seguridad pública, la aplicación estricta de la ley, y el orden social como pilares centrales en los programas. En cuanto a per304 Political Corruption, tiene un promedio de 0.84, el segundo más bajo entre comunidades y por debajo del promedio global de 1.32, lo cual significa que la corrupción política no es percibida como un problema central o que la atención se ha desplazado hacia otros temas.

**5.7.3 Análisis Comunidad 2: Heterogeneidad en Gobernanza con Enfoque en Seguridad y Estabilidad**

Países: Austria, Brazil, Canada, Chile, Panama, South Africa, Sweden, United Kingdom

Esta comunidad tiene el segundo promedio más alto entre las comunidades en per605: Law and Order con 3.41, por sobre el promedio global de 2.74. En per304: Political Corruption tiene un promedio de 1.14, el segundo más alto entre las comunidades y similar al promedio global de 1.32. En las variables per603 y per604 asociadas a Traditional Morality, tiene promedios de 0.95 y 0.13 respectivamente, por debajo del promedio global.

**5.7.4 Análisis Comunidad 3**: **“Problemas de gobernanza equilibrando crecimiento económico y estabilidad social.”**

Países: Bosnia-Herzegovina, Cyprus, Georgia, Iceland, Japan, Latvia, Montenegro, Serbia, South Korea.

En esta comunidad, Bosnia-Herzegovina, Georgia, Montenegro y Serbia han experimentado procesos de transición democrática o consolidación institucional tras conflictos o regímenes autoritarios. Esto puede influir en un enfoque más visible hacia la corrupción política (per304) y los esfuerzos por establecer estructuras de gobernanza más transparentes, lo que se ve bien explicado por el promedio de 2.65, el mayor entre todas las comunidades y muy superior al promedio global (1.32).

Además es la que tiene en promedio menos menciones en las variables per603 (0.91) y per604 (0.04), ambas asociadas a la Moralidad Tradicional, también por debajo del promedio global (1.53 y 0.20 respectivamente). Algunos de sus países son avanzados tecnológicamente y con alta urbanización, que se alejan de la moralidad tradicional En cuanto a per605 (Law and Order), tiene un promedio de 1.96, el más bajo entre las comunidades y muy por debajo del promedio global (2.74) reflejando una menor prioridad en seguridad.

# Conclusiones

El análisis de las comunidades permite identificar patrones comunes entre los programas de gobierno y agrupar a los países en comunidades con características similares, tomando como base la estructura inherente de los datos.

Durante el período 2003-2022, con un enfoque en el fenómeno de la Corrupción, logramos clasificar a los países en comunidades bien definidas:

* **Países con democracias con liderazgo en transparencia y compromiso con el orden legal y social:** Caracterizados por bajos niveles de menciones a corrupción política y un enfoque en el cumplimiento del orden legal.
* **Países con conservadurismo social y fortalecimiento de la Ley y el Orden:** Enfocados en valores tradicionales y en la seguridad pública como eje principal de sus programas.
* **Países heterogéneos en gobernanza con enfoque en seguridad y estabilidad:** Muestran un énfasis en la seguridad pública, aunque con enfoques variados en otros aspectos.
* **Países con problemas de gobernanza equilibrando crecimiento económico y estabilidad social:** Presentan desafíos en la gobernanza, pero también intentos por abordar múltiples prioridades.

Los programas de gobierno reflejan patrones comunes que permiten agrupar países en comunidades según sus prioridades temáticas, mostrando diferencias consistentes entre comunidades en términos de énfasis político. Estos resultados ofrecen un marco para entender las prioridades políticas globales y podrían ser utilizados como base para diseñar estrategias de gobierno y colaboración internacional. También es útil para hacer comparaciones entre resultados con países del mismo grupo ya que ayuda a medir progresos y evitar enfoques redundantes

# Trabajo Futuro, Mejoras y Limitaciones

7.1 Trabajo Futuro y Mejoras

* Explorar correlaciones entre las comunidades identificadas y otros factores externos como el sistema político, el crecimiento económico o los índices de democracia

7.2 Limitaciones

* Este análisis está basado en los datos disponibles en el Manifesto Project, lo que implica que las conclusiones están sujetas a la calidad y representatividad de esos datos.

# Bibliografía

Manifesto Project Dataset/ Lehmann, Pola / Franzmann, Simon / Al-Gaddooa, Denise / Burst, Tobias / Ivanusch, Christoph / Regel, Sven / Riethmüller, Felicia / Volkens, Andrea / Weßels, Bernhard / Zehnter, Lisa (2024): The Manifesto Data Collection. Manifesto Project (MRG / CMP/ MARPOR). Version 2024a. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB) / Göttingen: Institut für Demokratieforschung (IfDem). <https://doi.org/10.25522/> manifesto.mpds.2024a

Manifesto Project Dataset - Codebook/ Lehmann, Pola / Franzmann, Simon / Al-Gaddooa, Denise / Burst, Tobias / Ivanusch, Christoph / Regel, Sven / Riethmüller, Felicia / Volkens, Andrea /Weßels, Bernhard / Zehnter, Lisa (2024): The Manifesto Project Dataset - Codebook. Manifesto Project (MRG / CMP / MARPOR). Version 2024a. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB) / Göttingen: Institut für Demokratieforschung (IfDem).

Manifesto Project Dataset – Coding Instructions (5th re-revised edition)/ [manifesto-communication@wzb.eu/](mailto:manifesto-communication@wzb.eu/) website: <https://manifesto-project.wzb.eu/>

Manifesto-Project Dataset/ List of Political Parties/ Version 2024a from March 2024/ Lehmann, Pola / Franzmann, Simon / Al-Gaddooa, Denise / Burst, Tobias / Ivanusch, Christoph / Regel, Sven / Riethmüller, Felicia / Volkens, Andrea / Weßels, Bernhard / Zehnter, Lisa (2024): The Manifesto Project Party List - Codebook. Version 2024a. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB) / Göttingen: Institut für Demokratieforschung (IfDem). [manifesto-communication@wzb.eu/](mailto:manifesto-communication@wzb.eu/) website: <https://manifesto-project.wzb.eu/>

Statistical Odds &. Ends

<https://statisticaloddsandends.wordpress.com/2021/04/20/marginal-distributions-of-the-dirichlet-distribution-and-the-aggregation-property/>