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 RAVEAU MORALES

Diciembre 2024

SANTIAGO-CHILE

**AGRADECIMIENTO**

Dedico este trabajo a mi familia y a mis padres, quienes me han apoyado incondicionalmente a lo largo de este camino. Agradezco a Dios por la oportunidad de seguir aprendiendo.

Agradezco a los profesores que, durante estos dos años, me han guiado y enseñado en las distintas áreas del conocimiento que abarca este programa. Su dedicación y profesionalismo los convierten en un gran equipo.

Quiero expresar también mi agradecimiento a mi grupo de trabajo, conformado por Matías Mercandino S., Ricardo Chacón A. y Nicolás Lagos B., con quienes compartí una experiencia de trabajo serio y comprometido.

Tabla de contenido

[Resumen 1](#_Toc55504658)

[1. Introducción 2](#_Toc55504659)

[2. Trabajo Relacionado 4](#_Toc55504660)

[3. Hipótesis y Objetivos 6](#_Toc55504661)

[4. Datos y Metodología 8](#_Toc55504662)

[4.1. Datos 8](#_Toc55504663)

[4.2. Metodología 17](#_Toc55504664)

[5. Resultados Error! Bookmark not defined.](#_Toc55504665)

[6. Conclusión y Trabajo Futuro 55](#_Toc55504666)

[Bibliografía 57](#_Toc55504667)

# Resumen

Este estudio, desarrollado como parte del Proyecto Capstone del Magíster en Data Science de la Universidad del Desarrollo, tiene como objetivo identificar patrones comunes en los programas de gobierno de candidatos presidenciales a nivel mundial, con foco en el fenómeno de la Corrupción. La investigación agrupa países en comunidades con prioridades temáticas similares, utilizando técnicas de análisis de datos y clustering, y contrastando los hallazgos con el índice CPI (Corruption Perception Index) de Transparency International. Los resultados ofrecen una herramienta útil para el análisis comparado de políticas públicas, facilitando la identificación de referentes exitosos en la lucha contra la corrupción y permitiendo medir el progreso en países con desafíos similares. Asimismo, sienta las bases para futuros trabajos que combinen estos hallazgos con datos empíricos sobre la implementación y efectividad de políticas gubernamentales.

# 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.

“Se ha elegido el tema de la Corrupción por su relevancia para las democracias. Definimos la corrupción como el abuso del poder confiado para beneficio privado.

La corrupción erosiona la confianza, debilita la democracia, obstaculiza el desarrollo económico y exacerba aún más la desigualdad, la pobreza, la división social y la crisis ambiental. Exponer la corrupción y hacer que los corruptos rindan cuentas sólo puede suceder si entendemos la forma en que funciona la corrupción y los sistemas que la permiten.” 1

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 5

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.

# 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** Corrupción Política, Ley y Orden, Moralidad Tradicional, 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 (mapas de calor) 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 corresponden 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 Instrucciones de Codificación 5) 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 table with names of countries/regions

Description automatically generated

Tabla 1: Países en el estudio de Manifesto Project

**4.1.3 Categorías y Dominios**

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

Las variables se agrupan en 7 “dominios” que se detallan en las siguientes tablas:

Tabla 2: Variables en el dominio “Relaciones Exteriores”

|  |  |  |
| --- | --- | --- |
| Nombre-Dominio | Categoría/Variable | Descripción |

|  |  |  |
| --- | --- | --- |
| Relaciones Exteriores | per101 | Relaciones Exteriores Especiales: Favorable |
| per102 | Relaciones Exteriores Especiales: Desfavorable |
| per103 | Anti-Imperialismo |
| per104 | Militarización: Favorable |
| per105 | Militarización: Desfavorable |
| per106 | Paz |
| per107 | Internacionalismo: Favorable |
| per108 | Unión de la Comunidad Europea: Favorable |
| per109 | Internacionalismo: Desfavorable |
| per110 | Unión de la Comunidad Europea: Desfavorable |

Tabla 3: Variables en el dominio “Libertad y Democracia”

|  |  |  |
| --- | --- | --- |
| Nombre-Dominio | Categoría/Variable | Descripción |

|  |  |  |
| --- | --- | --- |
| Libertad y Democracia | per201 | Libertad y Derechos Humanos |
| per202 | Democracia |
| per203 | Constitucionalismo: Favorable |
| per204 | Constitucionalismo: Desfavorable |

Tabla 4: Variables en el dominio “Sistema Político”

|  |  |  |
| --- | --- | --- |
| Nombre-Dominio | Categoría/Variable | Descripción |

|  |  |  |
| --- | --- | --- |
| Sistema Político | per301 | Descentralización |
| per302 | Centralización |
| per303 | Eficiencia Gubernamental y Administrativa |
| per304 | Corrupción Política |
| per305 | Autoridad Política |

Tabla 5: Variables del dominio “Economía”

|  |  |  |
| --- | --- | --- |
| Nombre-Dominio | Categoría/Variable | Descripción |

|  |  |  |
| --- | --- | --- |
| Economía | per401 | Economía Libre de Mercado |
| per402 | Incentivos: Favorable |
| per403 | Regulación del Mercado |
| per404 | Planificación Económica |
| per405 | Corporativismo/ Economía Mixta |
| per406 | Proteccionismo: Favorable |
| per407 | Proteccionismo: Desfavorable |
| per408 | Objetivos Económico |
| per409 | Gestión de la Demanda: Keines |
| per410 | Crecimiento Económico: Favorable |
| per411 | Tecnología e Infraestructura: Favorable |
| per412 | Economía Centralizada |
| per413 | Nacionalismo |
| per414 | Ortodoxia Económica |
| per415 | Análisis Marxista |
| per416 | Economía Anti Crecimiento: Favorable |

Tabla 6: Variables del dominio “Bienestar y Calidad de Vida”

|  |  |  |
| --- | --- | --- |
| Nombre-Dominio | Categoría/Variable | Descripción |

|  |  |  |
| --- | --- | --- |
| Bienestar y Calidad de Vida | per501 | Protección Medioambiental |
| per502 | Cultura: Favorable |
| per503 | Igualdad: Favorable |
| per504 | Expansión del Estado de Bienestar |
| per505 | Limitación al Estado de Bienestar |
| per506 | Expansión de la Educación |
| per507 | Limitación a la Educación |

Tabla 7: Variables del dominio “Tejido de la Sociedad”

|  |  |  |
| --- | --- | --- |
| Nombre-Dominio | Categoría/Variable | Descripción |

|  |  |  |
| --- | --- | --- |
| Tejido Social | per601 | Modo de Vida Nacional: Favorable |
| per602 | Modo de Vida Nacional: Desfavorable |
| per603 | Tradición Moral: Favorable |
| per604 | Tradición Moral: Desfavorable |
| per605 | Ley y Orden: Favorable |
| per606 | Mentalidad Cívica: Favorable |
| per607 | Multiculturalismo: Favorable |
| per608 | Multiculturalismo: Desfavorable |

Tabla 8: Variables del dominio “Grupos Sociales”

|  |  |  |
| --- | --- | --- |
| Nombre-Dominio | Categoría/Variable | Descripción |

|  |  |  |
| --- | --- | --- |
| Grupos Sociales | per701 | Grupos Laborales: Favorable |
| per702 | Grupos Laborales: Desfavorable |
| per703 | Agricultura y Agricultores: Favorable |
| per704 | Grupos Profesionales y Clase Media |
| per705 | Grupos Minoritarios Desfavorecidos |
| per706 | Grupos Demográficos no Económicos |

**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:

Tabla 9: Columnas del dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Columna | Descripción | Tipo | Formato |
| country | Código del país | int |  |
| countryname | Nombre del país | string |  |
| oecdmember | Indica si el país es miembro de la OECD al momento de la elección | int | 0 ó 10 |
| eumember | Indica si el país es miembro de la Comunidad Europea o no al momento de la elección | int | 0, 10, 20 |
| edate | Fecha de la elección | string | AAAA-MM-DD |
| date | Año Mes de la elección | string | AAAAMM |
| party | Código del partido político | int |  |
| partyname | Nombre del partido político | string |  |
| partyabbrev | Abreviación del partido político | string |  |
| parfam | Tipo de partido (según agrupación de MP) | int |  |
| candidatename | Nombre del candidato | string |  |
| coderid | Identificación del Codificador (etiquetador) | int |  |
| manual | Versión de instrucciones de codificación | int |  |
| coderyear | Año en que se realizó la codificación | int |  |
| testresult | Resultado del test de evaluación del codificador a su ingreso/ Prueba de confiabilidad | int |  |
| testeditsim | Evaluación de confiabilidad del codificador | int |  |
| pervote | Porcentaje de votos ganado por el partido | float |  |
| voteest | Fuente del dato de pervote | int |  |
| presvote | Porcentaje de votos obtenidos en una elección presidencial, en la primera vuelta | float |  |
| absseat | Número absoluto de asientos de cada partido o coalición | float |  |
| totseats | Número total de asientos en el Congreso | int |  |
| progtype | Indica si el programa es de un partido único o si es de una coalición | int |  |
| datasetorigin | Fecha de publicación del dataset | int |  |
| corpusversion | Versión del dataset con que Manifesto Project publica analíticas | string |  |
| total | Número total de códigos asignados | int |  |
| peruncod | Porcentaje de frases en el programa que son etiquetadas con '000' en señal de que ninguna etiqueta se ajusta | float |  |
| per101 | Porcentaje de veces que el programa fue clasificado con la etiqueta 'per101' | float |  |
| per102 | Porcentaje de veces que el programa fue clasificado con la etiqueta 'per102'' | float |  |
| per103 | Porcentaje de veces que el programa fue clasificado con la etiqueta 'per103'' | float |  |
| . |  | float |  |
| . |  | float |  |
| . |  | float |  |
| per7062 | Porcentaje de veces que el programa fue clasificado con la etiqueta 'per7062' | float |  |
| rile | Ubicación del partido en una escala de "derecha-izquierda" según Michael Laver/Ian Budge (eds.): Party Policy and Government Coalitions, Houndmills, Basingstoke, Hampshire: The MacMillan Press 1992 | float |  |
| planeco | per403 + per404 + per412 | float |  |
| markeco | per401 + per414 | float |  |
| welfare | per503 + per504 | float |  |
| intpeace | per102 + per105 + per106 | float |  |
| datasetversion | Versión del dataset | string |  |
| id\_perm | String de caracteres aleatorio permanente que identifica mismas observaciones en diferentes versiones del dataset | int |  |

Para el objetivo de este trabajo las columnas que se utilizó fueron:

Tabla 10: Columnas utilizadas para la investigación:

|  |  |  |  |
| --- | --- | --- | --- |
| Columna | Descripción | Tipo | Formato |
| countryname | Nombre del país | string |  |
| edate | Fecha de la elección | string | AAAA-MM-DD |
| date | Año Mes de la elección | string | AAAAMM |
| pervote | Porcentaje de votos ganado por el partido | float |  |
| presvote | Porcentaje de votos obtenidos en una elección presidencial, en la primera vuelta | float |  |
| per101 | Porcentaje de veces que el programa fue clasificado con la etiqueta 'per101' | float |  |
| per102 | Porcentaje de veces que el programa fue clasificado con la etiqueta 'per102'' | float |  |
| per103 | Porcentaje de veces que el programa fue clasificado con la etiqueta 'per103'' | float |  |
| . |  | float |  |
| . |  | float |  |
| . |  | float |  |
| per7062 | Porcentaje de veces que el programa fue clasificado con la etiqueta 'per7062' | float |  |

**4.1.6 Valores ponderados**

Para el análisis nos interesa poder comparar programas de gobierno entre países en el tiempo. Dado que el dataset contiene información desagregada para cada elección según los candidatos o coaliciones que participaron, la forma de agrupar los programas para una elección en un año en particular en un país será ponderando las etiquetas de cada variable ‘per’ por el porcentaje de votos que obtuvo cada candidato o coalición. La columna del dataset que contiene este porcentaje es la columna ‘presvote’. De esta manera se obtiene un ‘programa de gobierno’ ponderado como reflejo de programa de gobierno país.

## 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
* Se pre procesará el dataset a fin de asegurar que haya un registro por país para cada período en análisis. Seleccionaremos el período 2019-2022.
* Se aplicará K-MEANS con las variables seleccionadas para identificar clusters. Con el método del codo identificamos el número óptimo de clusters con WCSS. Graficamos clusters.
* Se aplicará K-MEANS para todos los períodos de 2003 en adelante en períodos de 3 años. Se crea la Matriz de Coincidencias que asocia a los países la cantidad de veces que están en un mismo cluster con otros países. Se grafica el heatmap de coincidencias
* Con base en la Matriz de Coincidencias, se crea una red en la cual los nodos son países y los enlaces son coincidencias en un cluster. En esta red se identificarán comunidades.
* Se hará estadística básica en cada comunidad y compararemos los resultados.
* Contrastaremos lo obtenido con el índice de corrupción CPI (Corruption Perception Index), elaborado por Transparency International 8).

# Resultados

**5.1 Conocimiento del dataset**

Esta etapa ha sido relevante para poder validar el hecho de que es posible usar el dataset como un indicador de la importancia de un tema determinado en una realidad nacional.

**5.1.1: Orden y Corrupción**

Gráfico 2: Evolución de las variables que explican el aspecto Orden y Corrupción en Chile

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 (Ley y Orden: Favorable), 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**

Gráfico 3: Evolución de las variables que explican el aspecto Pueblos Originarios en Chile

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**

Gráfico 4: Evolución de las variables que explican el aspecto Inmigración en Chile

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 Venezuela. En 2021 se dicta la nueva ley de Migración y Extranjería.

**5.1.4: Derechos Humanos**

Gráfico 5: Evolución de las variables que explican el aspecto Derechos Humanos en Chile

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: Corrupción Política
* per603: Moralidad Tradicional: Favorable
* per604: Moralidad Tradicional: Negative
* per605: Ley y Orden: Favorable

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

|  |  |
| --- | --- |
| Variable | Descripción |
| per304 | **Corrupción Política**  Necesidad de eliminar la corrupción política y los abusos asociados de poder político y/o burocrático.  Necesidad de abolir estructuras y prácticas clientelistas. |

|  |  |
| --- | --- |
| Variable | Descripción |
| per603 | **Moralidad Tradicional: Favorable**  Menciones favorables a los valores morales tradicionales y/o religiosos. Puede incluir:   * Prohibición, censura y supresión de inmoralidad y comportamientos indebidos. * Mantenimiento y estabilidad de la familia tradicional como un valor. * Apoyo al rol de las instituciones religiosas en el estado y la sociedad. |

|  |  |
| --- | --- |
| Variable | Descripción |
| per604 | **Moralidad Tradicional: Desfavorable**  Oposición a los valores morales tradicionales y/o religiosos. Puede incluir:   * Apoyo al divorcio, aborto, etc. * Apoyo general a composiciones familiares modernas. * Llamados a la separación de la iglesia y el estado. |

|  |  |
| --- | --- |
| Variable | Descripción |
| per605 | **Ley y Orden: Favorable**  Menciones favorables a la aplicación estricta de la ley y acciones más severas contra el crimen doméstico. Solo se refiere a la aplicación del código legal del país del manifiesto. Puede incluir:   * Incremento del apoyo y recursos para la policía. * Actitudes más estrictas en los tribunales. * Importancia de la seguridad interna. |

**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.

Tabla 11: Estadísticas descriptivas considerando todos los períodos:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **count** | **promedio** | **desviación std** | **min** | **25%** | **50%** | **75%** | **max** |
| **per304** | 739 | 1,263001 | 2,042249 | 0 | 0,0642589 | 0,5606774 | 1,5836373 | 17,596681 |
| **per603** | 739 | 1,43189 | 1,716172 | 0 | 0,272377 | 0,9319665 | 1,9795939 | 16,868819 |
| **per604** | 739 | 0,187187 | 0,426697 | 0 | 0 | 0 | 0,1838392 | 4,9656146 |
| **per605** | 739 | 2,638929 | 2,532031 | 0 | 0,4852115 | 2,0109026 | 4,1792031 | 14,205347 |

Gráfico 6: Distribución de per304: Corrupción Política

A graph with a line and numbers

Description automatically generated with medium confidence

Gráfico 7: Distribución de per603: Moralidad Tradicional: Favorable

A graph with a yellow line

Description automatically generated

Gráfico 8: Distribución de per604: Moralidad Tradicional: Desfavorable

A graph with numbers and a number

Description automatically generated

Gráfico 9: Distribución de per605: Ley y Orden: Favorable

A graph with a number of percentages

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 2, cuya forma es la siguiente:

Gráfico 10: Distribución Beta

A graph with a blue line

Description automatically generated

**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.

Gráfico 11: Boxplot de per304: Corrupción Política

A graph with a line and numbers

Description automatically generated with medium confidence

Gráfico 11: Boxplot de per603: Moralidad Tradicional: Favorable

A graph with a line

Description automatically generated

Gráfico 11: Boxplot de per604: Moralidad Tradicional: Desfavorable

A graph with numbers and a line

Description automatically generated

Gráfico 12: Boxplot de per605: Ley y Orden: Favorable

A graph with a bar and a line

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:

Gráfico 13: Matriz de Correlaciones

A blue and red squares with white text

Description automatically generated

* La máxima correlación positiva es 0,21: esto significa 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 Preprocesamiento 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 subperíodos de 3 años a cada uno de los cuales se les aplicará K-MEANS.

La razón de elegir períodos de 3 años tiene que ver con que no todos los países tienen elecciones en un mismo año. Para asegurar que el análisis concentre una buena cantidad de países se eligió 3 años. También se podría haber elegido un período de 4 años o más, pero para cuando se quiera hacer análisis temporal se tendrían menos períodos para comparar.

K-MEANS es un algoritmo de clustering que agrupa los datos en un número predefinido 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’.

**5.5.1 Número de clusters**

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.

Gráfico 14: Número óptimo de clusters según el método del codo

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.

Con el 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 Score”. El Silhouette Score mide qué tan bien se agrupan las muestras dentro de un clúster y qué tan separados están los clústeres entre sí. Los valores varían entre -1 y 1.

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

Gráfico 15: Calidad de clusters según Silhouette Score

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.

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

Gráfico 16: Clusters identificados con K-MEANS (2D)

A graph with colored dots and numbers

Description automatically generated

Gráfico 17: Clusters identificados con K-MEANS (3D)

A graph of a graph with a grid and dots

Description automatically generated with medium confidence

* Países en Cluster 0: Armenia, Australia, Colombia, Dinamarca, Finlandia, Francia, Georgia, Grecia, Islandia, Montenegro, Nueva Zelanda, Macedonia del Norte, Polonia, Rusia, Eslovenia, Corea del Sur, Ucrania, Estados Unidos
* Países en Cluster 1: Argentina, Brasil, Costa Rica, Chipre, Letonia, Lituania, Serbia, Turquía, Ucrania
* Países en Cluster 2: Austria, Bosnia-Herzegovina, República Checa, Estonia, Irlanda, Italia, Holanda, Rumania, Eslovaquia, España, Suecia, Suiza, Canadá

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 en el Gráfico 16 que en el período de tiempo entre 2006 y 2008, los programas presidenciales se pueden visualizar en clusters que se distinguen bien el uno del otro. Dado que este es un período en particular, nada garantiza que en otro período el resultado sea tan nítido. Sin embargo, esta nitidez no es lo relevante como si lo es la cantidad de veces que dos países coinciden en un mismo cluster en el tiempo.

En este paso lo que haremos es que tomaremos un lapso 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 1, ya en el año 2000 hay más de 60 países informando, por lo cual elegimos 2003 a 2022.

**5.6.2 Matriz de Coincidencias**

La matriz de coincidencias nos permite ver con qué frecuencia los programas presidenciales de distintos países tiende a estar en un mismo cluster. Esta frecuencia es relevante para el análisis final puesto que aporta la dimensión temporal al análisis (recordar el objetivo general en 3.2), y es a partir de lo cual construiremos la red e identificaremos comunidades.

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.

Gráfico 18: Matriz de Coincidencias

A crossword puzzle with blue squares 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:

Gráfico 19: Grafo de Comunidades

A diagram of a network

Description automatically generated

Tabla 12: Países en cada Comunidad

|  |  |
| --- | --- |
| Comunidad | Países |
| 0 | Armenia, Australia, Dinamarca, Finlandia, Alemania, Grecia, Nueva Zelanda, Noruega, Rumania, Eslovenia, España, Estados Unidos |
| 1 | República Checa, Estonia, Holanda, Portugal, Eslovaquia |
| 2 | Austria, Brasil, Canadá, Chile, Panamá, Sudáfrica, Suecia, Reino Unido |
| 3 | Bosnia-Herzegovina, Chipre, Georgia, Islandia, Japón, Letonia, Montenegro, Serbia, Corea del Sur |

**5.6.4 Análisis de centralidades**

Como se puede observar en las centralidades, Australia (0.42) y Nueva Zelanda (0,36), son los principales nodos en cuanto a centralidad de Grado; es decir, están directamente relacionados con más conexiones directas con otros nodos en la red. Tratándose de programas de gobierno, esto podría indicar que comparten prioridades similares en torno a ciertos temas con otros nodos.

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

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

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

Tabla 13: Métricas de Centralidad

|  |  |  |
| --- | --- | --- |
| **País** | **Degree Centrality** | **Betweenness Centrality** |
| Australia | 0,424242424 | 0,060161135 |
| Nueva Zelanda | 0,363636364 | 0,400829726 |
| Alemania | 0,272727273 | 0,051978114 |
| Finlandia | 0,272727273 | 0,049084596 |
| Islandia | 0,272727273 | 0,153908129 |
| Armenia | 0,242424242 | 0,047316919 |
| Estados Unidos | 0,212121212 | 0,033306277 |
| España | 0,212121212 | 0,002840909 |
| Eslovenia | 0,212121212 | 0,002840909 |
| Letonia | 0,212121212 | 0,302299784 |
| Georgia | 0,212121212 | 0,157230038 |
| Rumania | 0,151515152 | 0 |
| Dinamarca | 0,151515152 | 0 |
| Estonia | 0,151515152 | 0,21969697 |
| Grecia | 0,151515152 | 0 |
| Portugal | 0,121212121 | 0 |
| Reino Unido | 0,121212121 | 0,353584957 |
| Suecia | 0,121212121 | 0,111904762 |
| Canadá | 0,121212121 | 0,062851732 |
| República Checa | 0,121212121 | 0 |
| Montenegro | 0,121212121 | 0,00094697 |
| Eslovaquia | 0,121212121 | 0 |
| Holanda | 0,121212121 | 0 |
| Brasil | 0,090909091 | 0,084496753 |
| Corea del Sur | 0,090909091 | 0 |
| Serbia | 0,090909091 | 0,060606061 |
| Sudáfrica | 0,060606061 | 0 |
| Chile | 0,060606061 | 0 |
| Austria | 0,060606061 | 0,037297078 |
| Japón | 0,060606061 | 0 |
| Noruega | 0,060606061 | 0 |
| Bosnia-Herzegovina | 0,03030303 | 0 |
| Chipre | 0,03030303 | 0 |
| Panamá | 0,03030303 | 0 |

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

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

Tabla 14: Promedios por variable y comunidad y del dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Promedios de las variables por comunidad y del dataset | | | | |
|  |  |  |  |  |
| **Comunidad** | **per603** | **per604** | **per605** | **per304** |
| 0 | 1,520047 | 0,133878 | 2,409427 | 0,612475 |
| 1 | 2,001001 | 0,260074 | 4,777077 | 0,84237 |
| 2 | 0,947457 | 0,129773 | 3,415475 | 1,13582 |
| 3 | 0,910009 | 0,041275 | 1,960428 | 2,650676 |
| Dataset | 1,532251 | 0,195463 | 2,742213 | 1,324202 |

Gráfico 20: Matriz de promedios por variable y comunidad

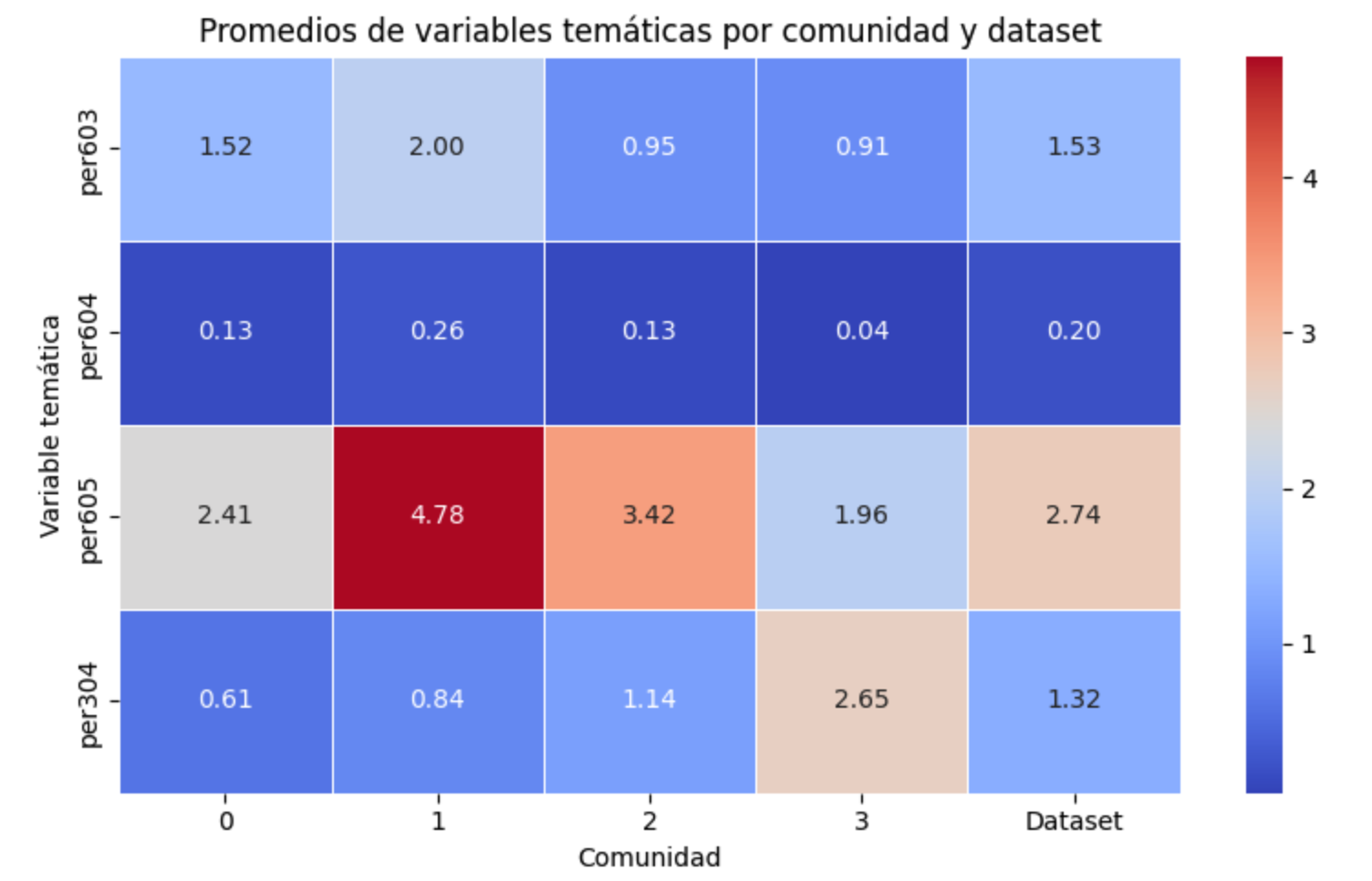


Gráfico 21: Histogramas del promedio de cada variable, por comunidad, y promedio del dataset en el período en análisis.

A group of different colored bars

Description automatically generated

Tabla 15 (repetición de Tabla 12)

|  |  |
| --- | --- |
| Comunidad | Países |
| 0 | Armenia, Australia, Dinamarca, Finlandia, Alemania, Grecia, Nueva Zelanda, Noruega, Rumania, Eslovenia, España, Estados Unidos |
| 1 | República Checa, Estonia, Holanda, Portugal, Eslovaquia |
| 2 | Austria, Brasil, Canadá, Chile, Panamá, Sudáfrica, Suecia, Reino Unido |
| 3 | Bosnia-Herzegovina, Chipre, Georgia, Islandia, Japón, Letonia, Montenegro, Serbia, Corea del Sur |

**5.7 Análisis de Comunidades**

En el siguiente análisis asignaremos una frase que describa a la comunidad.

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

* Países: Armenia, Australia, Dinamarca, Finlandia, Alemania, Grecia, Nueva Zelanda, Noruega, Rumania, Eslovenia, España, Estados Unidos

En esta comunidad la mención de per304 (Corrupción Política) es la menor entre todas las comunidades y sustancialmente menor al promedio global (0.61 y 1.32 respectivamente). Además, en per605 (Ley y Orden: 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 (Corrupción Política) 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, Ley y Orden, 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: República Checa, Estonia, Holanda, Portugal, Eslovaquia

Entre los países de esta comunidad están República Checa, Estonia y Eslovaquia 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 Corrupción Política, 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, Brasil, Canadá, Chile, Panamá, Sudáfrica, Suecia, Reino Unido

Esta comunidad tiene el segundo promedio más alto entre las comunidades en per605: Ley y Orden con 3.41, por sobre el promedio global de 2.74. En per304: Corrupción Política 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 Moralidad Tradicional, 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, Chipre, Georgia, Islandia, Japón, Letonia, Montenegro, Serbia, Corea del Sur.

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.

**5.7.5 Paralelo con datos de Transparency Internacional** 8

“Transparencia Internacional es un movimiento global que trabaja en más de 100 países para poner fin a la injusticia de la corrupción”.9

Su equipo de investigación realiza evaluaciones rigurosas e independientes de la corrupción en todo el mundo. A nivel global, producen datos comparativos que miden la prevalencia de la corrupción, las experiencias de los ciudadanos y las actitudes hacia ella. También exploran cómo la corrupción se propaga a través de las fronteras y en diferentes sectores de la economía. A nivel nacional, local e incluso sectorial investigan la manifestación específica de la corrupción, sus causas, sus consecuencias y qué funciona para reducirla.10

**5.7.5.1 Índice de la Percepción de la Corrupción (CPI)**

Uno de los principales productos de investigación de Transparency International es el “CPI” (Corruption Percepction Index), un índice de percepción de la Corrupción en el sector público. Este índice se calcula desde 2012 en adelante en base a varios datasets que son obtenidos de 13 entidades externas e independientes a Transparency International; ejemplos de estas entidades son el Banco Mundial (World Bank), el Foro Económico Mundial (World Economic Forum), empresas consultoras y clasificadoras de riesgo, y centros de investigación (Think Tanks).

Cada una de estas 13 fuentes califica a los países con un puntaje según una escala propia (por ejemplo, de 1 a 7, ó de 1 a 100). Como resultado la comparación directa entre países no es posible. Luego, cada uno de estos puntajes es transformado a una escala única estandarizada lo cual permite la comparación relativa entre países. Estos puntajes estandard son convertidos a una escala de 0 a 100, que es la que se usa para el CPI. El CPI final se obtiene de los promedios de los puntajes estandarizados para cada país.11

**5.7.5.2. Interpretación del CPI**

En un extremo de la escala de CPI (CPI = 0) están los países “Highly Corrupt” ó “altamente corruptos”, mientras que en el otro extremo (CPI = 100), están los países “Very Clean” ó “muy limpios”.

**5.7.5.3 CPI 2012-2023**

Utilizaremos el CPI para estimar el promedio del CPI para los países de cada comunidad.

Tabla 16: CPI de 2012 a 2023 para los países en cada comunidad

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **País** | **Comunidad** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** | **2023** |
| Armenia | 0 | 34 | 36 | 37 | 35 | 33 | 35 | 35 | 42 | 49 | 49 | 46 | 47 |
| Australia | 85 | 81 | 80 | 79 | 79 | 77 | 77 | 77 | 77 | 73 | 75 | 75 |
| Alemania | 79 | 78 | 79 | 81 | 81 | 81 | 80 | 80 | 80 | 80 | 79 | 78 |
| Dinamarca | 90 | 91 | 92 | 91 | 90 | 88 | 88 | 87 | 88 | 88 | 90 | 90 |
| España | 65 | 59 | 60 | 58 | 58 | 57 | 58 | 62 | 62 | 61 | 60 | 60 |
| Finlandia | 90 | 89 | 89 | 90 | 89 | 85 | 85 | 86 | 85 | 88 | 87 | 87 |
| Grecia | 36 | 40 | 43 | 46 | 44 | 48 | 45 | 48 | 50 | 49 | 52 | 49 |
| Noruega | 85 | 86 | 86 | 88 | 85 | 85 | 84 | 84 | 84 | 85 | 84 | 84 |
| Nueva Zelanda | 90 | 91 | 91 | 91 | 90 | 89 | 87 | 87 | 88 | 88 | 87 | 85 |
| Rumania | 44 | 43 | 43 | 46 | 48 | 48 | 47 | 44 | 44 | 45 | 46 | 46 |
| Eslovenia | 61 | 57 | 58 | 60 | 61 | 61 | 60 | 60 | 60 | 57 | 56 | 56 |
| Estados  Unidos | 73 | 73 | 74 | 76 | 74 | 75 | 71 | 69 | 67 | 67 | 69 | 69 |
| República Checa | 1 | 49 | 48 | 51 | 56 | 55 | 57 | 59 | 56 | 54 | 54 | 56 | 57 |
| Estonia | 64 | 68 | 69 | 70 | 70 | 71 | 73 | 74 | 75 | 74 | 74 | 76 |
| Holanda | 84 | 83 | 83 | 84 | 83 | 82 | 82 | 82 | 82 | 82 | 80 | 39 |
| Portugal | 63 | 62 | 63 | 64 | 62 | 63 | 64 | 62 | 61 | 62 | 62 | 61 |
| República Eslovaca | 46 | 47 | 50 | 51 | 51 | 50 | 50 | 50 | 49 | 52 | 53 | 54 |
| Austria | 2 | 69 | 69 | 72 | 76 | 75 | 75 | 76 | 77 | 76 | 74 | 71 | 71 |
| Brazil | 43 | 42 | 43 | 38 | 40 | 37 | 35 | 35 | 38 | 38 | 38 | 36 |
| Canadá | 84 | 81 | 81 | 83 | 82 | 82 | 81 | 77 | 77 | 74 | 74 | 76 |
| Chile | 72 | 71 | 73 | 70 | 66 | 67 | 67 | 67 | 67 | 67 | 67 | 66 |
| Reino Unido | 74 | 76 | 78 | 81 | 81 | 82 | 80 | 77 | 77 | 78 | 73 | 71 |
| Panamá | 38 | 35 | 37 | 39 | 38 | 37 | 37 | 36 | 35 | 36 | 36 | 35 |
| Suecia | 88 | 89 | 87 | 89 | 88 | 84 | 85 | 85 | 85 | 85 | 83 | 82 |
| Sudáfrica | 43 | 42 | 44 | 44 | 45 | 43 | 43 | 44 | 44 | 44 | 43 | 41 |
| Bosnia-Herzegovina | 3 | 42 | 42 | 39 | 38 | 39 | 38 | 38 | 36 | 35 | 35 | 34 | 35 |
| Chipre | 66 | 63 | 63 | 61 | 55 | 57 | 59 | 58 | 57 | 53 | 52 | 53 |
| Georgia | 52 | 49 | 52 | 52 | 57 | 56 | 58 | 56 | 56 | 55 | 56 | 53 |
| Islandia | 82 | 78 | 79 | 79 | 78 | 77 | 76 | 78 | 75 | 74 | 74 | 72 |
| Japón | 74 | 74 | 76 | 75 | 72 | 73 | 73 | 73 | 74 | 73 | 73 | 73 |
| Corea del Sur | 56 | 55 | 55 | 54 | 53 | 54 | 57 | 59 | 61 | 62 | 63 | 63 |
| Montenegro | 41 | 44 | 42 | 44 | 45 | 46 | 45 | 45 | 45 | 46 | 45 | 46 |
| Serbia | 39 | 42 | 41 | 40 | 42 | 41 | 39 | 39 | 38 | 38 | 36 | 36 |

Tabla 17: Promedio de 2012 a 2023 del índice CPI, por comunidad

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **País** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** | **2023** |
| Comunidad 0 | 69 | 69 | 69 | 70 | 69 | 69 | 68 | 69 | 70 | 69 | 69 | 69 |
| Comunidad 1 | 61 | 62 | 63 | 65 | 64 | 65 | 66 | 65 | 64 | 65 | 65 | 57 |
| Comunidad 2 | 64 | 63 | 64 | 65 | 64 | 63 | 63 | 62 | 62 | 62 | 61 | 60 |
| Comunidad 3 | 56 | 56 | 56 | 55 | 55 | 55 | 56 | 56 | 55 | 54 | 54 | 54 |

Gráfico 22: Promedio de 2012 a 2023 del índice CPI, por Comunidad

A graph with lines and numbers

Description automatically generated

Por último, para obtener un valor general para comparar usaremos el promedio ya que los valores son relativamente uniformes, no hay valores atípicos y no presentan grandes variaciones en el tiempo.

Tabla 18: Valores promedio de CPI 2012 a 2023, por comunidad:

|  |  |
| --- | --- |
| Comunidad | Promedio |
| 0 | 69,1 |
| 1 | 63,5 |
| 2 | 62,8 |
| 3 | 55,2 |

En resumen, tenemos la siguiente tabla de análisis de Comunidades:

Tabla 19: Resumen Análisis de Comunidades en cuanto a la Corrupción

|  |  |  |  |
| --- | --- | --- | --- |
| Comunidad | Países | Frase Descriptiva | Promedio CPI |
| 0 | Armenia, Australia, Dinamarca, Finlandia, Alemania, Grecia, Nueva Zelanda, Noruega, Rumania, Eslovenia, España, Estados Unidos | “Democracias con liderazgo en transparencia y compromiso con el orden legal y social” | 69,1 |
| 1 | República Checa, Estonia, Holanda, Portugal, Eslovaquia | “Conservadurismo Social y Fortalecimiento de la Ley y el Orden” | 63,5 |
| 2 | Austria, Brasil, Canadá, Chile, Panamá, Sudáfrica, Suecia, Reino Unido | “Heterogeneidad en Gobernanza con Enfoque en Seguridad y Estabilidad” | 62,8 |
| 3 | Bosnia-Herzegovina, Chipre, Georgia, Islandia, Japón, Letonia, Montenegro, Serbia, Corea del Sur | “Problemas de gobernanza equilibrando crecimiento económico y estabilidad social.” | 55,2 |

Los patrones identificados en cada comunidad permiten a los países compartir experiencias exitosas o enfrentar desafíos comunes, especialmente en temas como corrupción, gobernanza y estabilidad social.

# 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. Esto valida el método seguido como herramienta para analizar prioridades de los países y ver qué patrones surgen, resultando de utilidad para recomendaciones de políticas.

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.

Las comunidades individualizadas responden adecuadamente a la clasificación de un organismo como Transparency International cuyos índices son utilizados para evaluar a los países sobre la base de resultados concretos de medición de la Corrupción.

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 progreso.

**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

1.Transparency International/ About/ What is Corruption. <https://www.transparency.org/en/what-is-corruption>

2.Statistical Odds & Ends. Marginal distributions of the Dirichlet distribution and the aggregation property.<https://statisticaloddsandends.wordpress.com/2021/04/20/marginal-distributions-of-the-dirichlet-distribution-and-the-aggregation-property/>

3.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

4.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).

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

6.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/>

7.Transparency International. Corruption Perceptions Index 2023. CPI-2023-Report.pdf

8.Transparency International. Corruption Percepctions Index 2023. <https://www.transparency.org/en/cpi/2023>

9. Transparency International. About. <https://www.transparency.org/en/about>

10. Transparency International. Research. <https://www.transparency.org/en/research>

11.Transparency International. Research. <https://youtu.be/9JoNjIfbPV0>