



**Estudio de la relación entre los Valores Sociales y la Aceptación de Sobornos como conducta corrupta: Un Estudio con Modelos SEM y datos de la Encuesta Mundial de Valores**

**Study of the Relationship Between Social Values and the Acceptance of Bribery as Corrupt Conduct: A Study Using SEM Models and Data from the World Values Survey**

Giovanny Hernando Gómez Convers

Proyecto de Grado

**Director**

Sergio A. Castrillón-Orrego. Ph.D.

**Codirectora**

Paula María Almonacid Hurtado. Ph.D.

UNIVERSIDAD EAFIT  
ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA  
MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA  
MEDELLÍN  
2024

## CONTENIDO

1. PLANTEAMIENTO DEL PROBLEMA.....	8
2. JUSTIFICACIÓN.....	11
3. OBJETIVOS.....	13
3.1. GENERAL .....	13
3.2. ESPECÍFICOS .....	13
4. MARCO TEÓRICO .....	14
5. METODOLOGÍA .....	21
5.1. ENTENDIMIENTO DEL NEGOCIO .....	21
5.2. ENTENDIMIENTO DE LOS DATOS.....	21
5.3. PREPARACION DE LOS DATOS .....	22
5.4. MODELADO .....	22
5.5. EVALUACION .....	23
5.6. DESPLIEGUE.....	24
6. PRESENTACION DE RESULTADOS .....	25
6.1. EXPLICACIÓN DEL PROBLEMA E HIPÓTESIS .....	25
6.2. ANÁLISIS EXPLORATORIO DE DATOS .....	26
6.3. PROCESAMIENTO DE DATOS.....	30
6.3.1. Análisis de datos faltantes e imputación.....	30
6.3.2. Reducción de dimensionalidad.....	34
6.3.3. Estandarizar o no estandarizar.....	34
6.3.4. Análisis de la composición de los clústeres y su relación con el marco teórico	37
6.4. MODELACION SEM.....	40

6.4.1.	Especificación del modelo .....	42
6.4.2.	Identificación del modelo .....	44
6.4.3.	Estimación y evaluación del modelo.....	45
6.4.4.	Revisión y mejoras del modelo.....	47
6.5.	DESPLIEGUE.....	49
7.	CONCLUSIONES .....	51
8.	REFERENCIAS .....	55
9.	ANEXOS.....	61
9.1.	Formulario maestro de la Encuesta Mundial de Valores, Ola 7, 2017-2020 61	
9.2.	Proporción de respuestas con valor negativo por bloque temático .....	61
9.3.	Matrices de correlación sin datos faltantes por bloque temático .....	66
9.4.	Matrices de correlación final por bloque temático.....	68

## LISTA DE FIGURAS

<i>Ilustración 1. Comparación p-value bloque temático capital social</i> .....	32
<i>Ilustración 2. Comparación p-value bloque temático seguridad</i> .....	33
<i>Ilustración 3. Comparación p-value bloque temático demografía</i> .....	33
<i>Ilustración 4. UMAP con datos sin estandarizar</i> .....	35
<i>Ilustración 5. UMAP con datos estandarizados</i> .....	35
<i>Ilustración 6. Técnica del codo para encontrar número óptimo de clústeres con datos sin estandarizar</i> .....	36
<i>Ilustración 7. Técnica del codo para encontrar número óptimo de clústeres con datos estandarizados</i> .....	36
<i>Ilustración 8. Técnica de la silueta para encontrar número óptimo de clústeres con datos sin estandarizar</i> ...	36
<i>Ilustración 9. Técnica de la silueta para encontrar número óptimo de clústeres con datos estandarizados</i> ....	36
<i>Ilustración 10. UMAP con datos sin estandarizar con 3 clústeres</i> .....	37
<i>Ilustración 11. UMAP con datos estandarizados con 3 clústeres</i> .....	37
<i>Ilustración 12. Preguntas con mayor prevalencia por clúster</i> .....	38
<i>Ilustración 13. Composición de Clústeres por bloque temático</i> .....	39
<i>Ilustración 14. Enunciado pregunta Q118</i> .....	42
<i>Ilustración 15. Modelo SEM propuesto</i> .....	45
<i>Ilustración 16. Cargas factoriales para preguntas seleccionadas pilar Regulatorio</i> .....	47
<i>Ilustración 17. Cargas factoriales para preguntas seleccionadas pilar Normativo</i> .....	47
<i>Ilustración 18. Cargas factoriales para preguntas seleccionadas pilar Cognitivo</i> .....	48
<i>Ilustración 19. Proporción de preguntas seleccionadas en el modelo SEM para el Pilar Regulatorio, distribuidas en clústeres generados</i> .....	48
<i>Ilustración 20. Proporción de preguntas seleccionadas en el modelo SEM para el Pilar Normativo, distribuidas en clústeres generados</i> .....	49
<i>Ilustración 21. Proporción de preguntas seleccionadas en el modelo SEM para el Pilar Cognitivo, distribuidas en clústeres generados</i> .....	49

## LISTA DE TABLAS

<i>Tabla 1. Olas de la WVS.....</i>	<i>26</i>
<i>Tabla 2. Características generales del conjunto de datos .....</i>	<i>27</i>
<i>Tabla 3. Evaluación de variables candidatas.....</i>	<i>28</i>
<i>Tabla 4. Preguntas con proporción de datos faltantes superior al 10%.....</i>	<i>29</i>
<i>Tabla 5. Evaluación de preguntas candidatas.....</i>	<i>30</i>
<i>Tabla 6. p-values menores a 0.05 en imputación de datos .....</i>	<i>32</i>
<i>Tabla 7. Cálculo del coeficiente de silueta.....</i>	<i>37</i>
<i>Tabla 8. Preguntas con mayor prevalencia clúster 0.....</i>	<i>39</i>
<i>Tabla 9. Preguntas con mayor prevalencia clúster 1.....</i>	<i>40</i>
<i>Tabla 10. Preguntas con mayor prevalencia clúster 2.....</i>	<i>40</i>
<i>Tabla 11. Preguntas seleccionadas para el pilar Regulatorio .....</i>	<i>43</i>
<i>Tabla 12. Preguntas seleccionadas para el pilar Normativo .....</i>	<i>43</i>
<i>Tabla 13. Preguntas seleccionadas para el pilar Cognitivo .....</i>	<i>44</i>
<i>Tabla 14. Resultados estadísticos del modelo SEM.....</i>	<i>46</i>

## RESUMEN

En un contexto global de rápidos cambios sociales, investigar la relación entre valores sociales y corrupción es cada vez más urgente y trascendente. ¿Cuáles comportamientos son deseables? ¿Cuáles manifestamos cotidianamente? La Encuesta Mundial de Valores (*World Values Survey*, WVS) constituye una fuente de datos primordial para comprender los valores sociales en diversos contextos. Sin embargo, aún no se ha explorado suficientemente cómo estos valores influyen en la aceptación de sobornos y, por ende, en la corrupción. Este estudio examina los patrones subyacentes en las agrupaciones de respuestas y los analiza sistemáticamente aprovechando las posibilidades holísticas que ofrece el marco teórico del institucionalismo. El objetivo ha consistido en identificar las causalidades y magnitudes de las influencias más significativas en la relación valores sociales y corrupción. Mediante un robusto análisis de datos, técnicas de imputación, reducción de dimensionalidad, análisis de *clustering* y modelación SEM, identificamos los principales factores con mayor incidencia en la aceptación de sobornos. Los resultados demuestran que los tres pilares del institucionalismo proporcionan un enfoque valioso para entender la corrupción, mediante la consideración simultánea de variables y componentes clave. Los valores sociales, al ser internalizados, facilitan la aceptación de sobornos en ciertos contextos, destacando la influencia de la dimensión cognitiva. Aunque los marcos legales pueden mejorar la transparencia, el entorno cultural y las costumbres ejercen una influencia determinante en la aceptación de prácticas corruptas. Estos hallazgos subrayan la necesidad de fomentar una cultura ética sólida e implementar programas educativos que promuevan integridad y transparencia para mitigar eficazmente la corrupción.

**Palabras clave:** Corrupción, Valores Sociales, Institucionalismo, Aprendizaje de máquina, *World Values Survey*, UMAP, SEM, *K-Means*, *Clustering*

## ABSTRACT

In a global context of rapid social change, investigating the relationship between social values and corruption has become increasingly urgent and significant. Which behaviors are desirable? Which do we manifest in daily life? The World Values Survey (WVS) serves as a crucial data source for understanding social values in various contexts. However, how these values influence the acceptance of bribery, and thus corruption, has not been sufficiently explored. This study examines the underlying patterns in response clusters and systematically analyzes them using the holistic possibilities offered by the institutionalism theoretical framework. The objective is to identify the most significant causalities and influences in the relationship between social values and corruption. Through robust data analysis, imputation techniques, dimensionality reduction, clustering analysis, and SEM modeling, we identify the main factors impacting the acceptance of bribery. The results demonstrate that the three pillars of institutionalism provide a valuable approach to understanding corruption by simultaneously considering key variables and components. When internalized, social values facilitate the acceptance of bribery in certain contexts, highlighting the influence of the cognitive dimension. Although legal frameworks can enhance transparency, cultural environment and customs have a more determining influence on the acceptance of corrupt practices. These findings underscore the need to foster a strong ethical culture and implement educational programs that promote integrity and transparency to effectively mitigate corruption.

**Keywords:** Corruption, Social Values, Institutionalism, Machine Learning, *World Values Survey*, UMAP, SEM, *K-Means*, *Clustering*.

## 1. PLANTEAMIENTO DEL PROBLEMA

La corrupción es un problema extendido y grave que afecta a todas las sociedades en mayor o menor medida. En particular, la aceptación de sobornos representa una forma específica de corrupción que disminuye la confianza en las instituciones, distorsiona los procesos económicos, políticos y perpetúa la desigualdad social (Pope, 2000). A pesar de los esfuerzos globales por combatir la corrupción, los enfoques tradicionales han demostrado ser insuficientes debido a su enfoque principalmente institucional y legal (Johnston, 1986). Entre las diversas formas de corrupción, la aceptación de sobornos se destaca como una de las más prevalentes y perjudiciales. Esta conducta no solo implica una transacción ilegal entre el sobornador y el sobornado, sino que refleja una serie de valores sociales subyacentes que permiten, habilitan y validan dichas prácticas.

A pesar de los numerosos esfuerzos y políticas implementadas para combatir la corrupción, la aceptación de sobornos sigue siendo una práctica común en muchos países. Esto apunta a que las raíces del problema son profundas y están entrelazadas con los valores sociales subyacentes que influyen en las actitudes y comportamientos de los individuos. La pregunta clave que surge es: ¿cómo y en qué medida los valores sociales influyen en la predisposición de los individuos a aceptar sobornos? Para abordar esta pregunta, es crucial entender primero qué se entiende por valores sociales y cómo estos se manifiestan en diferentes contextos culturales.

Los valores sociales reflejan creencias compartidas sobre lo que es importante y deseable en una sociedad. Estos valores guían las acciones y decisiones de los individuos y están influenciados por factores históricos, culturales, económicos y políticos (Hofstede, 2001). En el contexto de la corrupción, los valores sociales pueden influir en la percepción de lo que constituye un comportamiento aceptable o inaceptable y, por ende, en la predisposición de los individuos a participar en actos corruptos como la aceptación de sobornos.

La Encuesta Mundial de Valores (WVS World Values Survey, por sus siglas en inglés) ofrece una visión integral de los valores, actitudes y creencias de los individuos en diversas sociedades (WVS, 2024). Sin embargo, la gran cantidad de variables y la diversidad de los contextos culturales representados en la WVS plantean un desafío significativo para los investigadores (Haerpfer et al., 2022). La multi-dimensionalidad y complejidad de los resultados requieren un análisis cuidadoso para identificar los factores más relevantes que influyen en la aceptación de sobornos. La aceptación de sobornos como conducta corrupta está influenciada por una serie de factores interrelacionados. Entre estos, los valores sociales juegan un papel fundamental. En sociedades donde la corrupción es vista como un mal menor o incluso como una necesidad para navegar las complejidades burocráticas, los individuos pueden estar más inclinados a aceptar sobornos. Por otro lado, en sociedades donde los valores de integridad, transparencia y responsabilidad son

predominantes, la aceptación de sobornos puede ser menos común. Este problema se complica aún más cuando se considera la variabilidad de los valores sociales entre diferentes países y regiones.

En algunos contextos, la aceptación de sobornos puede estar normalizada y culturalmente aceptada, mientras que en otros puede ser fuertemente condenada y sancionada. Esta variabilidad cultural implica que las políticas y estrategias anticorrupción deben ser adaptadas a los contextos específicos de cada sociedad, teniendo en cuenta los valores y creencias subyacentes que influyen el comportamiento de los individuos. Según el Índice de Percepción de la Corrupción (IPC) 2023 de Transparency International, más de 120 de los 180 países evaluados obtuvieron una puntuación por debajo de 50 sobre 100, indicando problemas significativos de corrupción en el sector público. La puntuación media global fue de 43/100. (2023 Corruption Perceptions Index: Explore The... - Transparency.Org, n.d.).

Además, la relación entre los valores sociales y la aceptación de sobornos no es directa ni simple. Está mediada por una serie de factores adicionales, como las estructuras institucionales, el nivel de desarrollo económico, la desigualdad social y la percepción de la efectividad del sistema judicial. Por ejemplo, en países donde las instituciones son débiles y la justicia es percibida como ineficaz o corrupta, los individuos pueden sentirse más justificados para participar en actos corruptos, incluyendo la aceptación de sobornos (Li, 2019). Otra dimensión del problema es el papel de las normas sociales y las redes de relaciones personales. En muchas sociedades, las relaciones personales y las redes de influencia juegan un papel crucial en la vida cotidiana y en la toma de decisiones. En tales contextos, la aceptación de sobornos puede ser vista como una forma de reciprocidad o de mantenimiento de las relaciones sociales, lo que complica aún más los esfuerzos para combatir esta forma de corrupción.

Con respecto a los datos, la multi dimensionalidad de la Encuesta Mundial de Valores (WVS) presenta un desafío adicional significativo. Con más de 300 variables disponibles para estudiar, la tarea de identificar cuáles son las más relevantes para el estudio específico de la aceptación de sobornos es compleja. Esta alta dimensionalidad no solo complica el análisis, sino que también plantea problemas específicos relacionados con los datos faltantes, la tipificación o normalización de datos, la identificación de grupos de individuos con características similares y la validación de los modelos seleccionados. Los datos faltantes son un problema común en encuestas a gran escala como la WVS. La imputación de datos es necesaria para manejar estos vacíos y asegurar que los análisis sean robustos y representativos. Sin embargo, elegir el método de imputación adecuado es crucial, ya que diferentes técnicas pueden llevar a resultados significativamente distintos. La imputación inadecuada puede introducir sesgos y afectar la validez de los hallazgos. La tipificación o normalización de datos es esencial para comparar variables que pueden estar en diferentes escalas o unidades. Sin una normalización adecuada, las variables con valores más altos pueden dominar los análisis y

distorsionar los resultados. Esto es particularmente importante cuando se utilizan técnicas de reducción de dimensionalidad y selección de grupos de individuos con características similares, comúnmente llamado, clusterización, ya que estas dependen en gran medida de la escala de los datos.

Para la Organización para la Cooperación y el Desarrollo Económico (OCDE) se estima que la corrupción cuesta a la economía global más de USD 2.6 billones al año, lo que equivale al 5% del PIB mundial (“Consequences of Corruption at the Sector Level and Implications for Economic Growth and Development,” 2015). La aceptación de sobornos como una forma de conducta corrupta está profundamente arraigada en los valores sociales y culturales de las sociedades. Solo mediante una comprensión profunda y matizada de estos factores, a través de un enfoque basado en la ciencia de datos y la analítica, será posible diseñar estrategias y políticas eficaces para combatir la corrupción y promover la integridad y la transparencia en las organizaciones y sociedades.

## 2. JUSTIFICACIÓN

Abordar un estudio sobre la corrupción es relevante desde las diferentes dimensiones de los valores sociales y no solo desde los aspectos legislativos o sancionatorios. Pasarse un semáforo en rojo, no obedecer el turno en una fila, pagar una “mordida” para hacer pasar la revisión técnico-mecánica y normalizar el “usted no sabe quién soy yo” son acciones que deben estudiarse con muchísimos más elementos para lograr entender el comportamiento de cada individuo que se enmarca en una sociedad u organización. Aspectos tales como el biológico, neurológico, emocional son dimensiones que influyen en el comportamiento, a veces, mucho más que los aspectos meramente normativos (Salcedo Albarán, 2007).

La importancia de esta investigación radica en la necesidad de comprender cómo los valores sociales influyen en la predisposición de los individuos a aceptar sobornos y participar en actos de corrupción. Tradicionalmente, los enfoques para combatir la corrupción se han centrado en reformas legales e institucionales, las cuales, aunque necesarias, han demostrado ser insuficientes para erradicar el problema de manera efectiva (Johnston, 1986). La corrupción es un fenómeno complejo y multidimensional que requiere un análisis más profundo que considere también los factores sociales, morales y psicológicos que subyacen a los comportamientos corruptos. La teoría institucional planteada por Richard W. Scott (Scott, 1995) proporciona un marco teórico robusto para analizar estos fenómenos, destacando tres pilares esenciales: el regulativo, el normativo y el cultural-cognitivo. Estos pilares permiten entender cómo las normas, reglas y esquemas mentales compartidos guían las acciones dentro de una sociedad y cómo estos elementos pueden ser aprovechados para abordar problemas complejos como la corrupción.

Utilizando modelos de ecuaciones estructurales (SEM) y técnicas de reducción de dimensionalidad como UMAP o PCA, con este trabajo se espera analizar los datos de la Encuesta Mundial de Valores (WVS) para identificar patrones y factores subyacentes. Estos métodos proporcionarán conocimientos más profundos y precisos que los enfoques tradicionales, permitiendo una mejor comprensión de los valores sociales que predisponen a los individuos a aceptar sobornos. La capacidad de manejar grandes volúmenes de datos y de identificar relaciones complejas brindan la posibilidad de validar hipótesis con mayores características a las usadas con métodos tradicionales.

Además, la capacidad de aplicar modelos de Machine Learning no supervisados permitirá identificar y ofrecer explicaciones más amplias y profundas, además de plantear rutas metodológicas y escenarios replicables, de tal forma que puedan auscultar los patrones vistos en los datos. Esto es importante en la WVS, que incluye más de 300 variables. Las técnicas avanzadas de imputación de datos y normalización aseguran que los análisis sean robustos y representativos, garantizando la validez y confiabilidad de los hallazgos.

Al identificar los valores sociales que predisponen a los individuos a aceptar sobornos, se puede diseñar políticas que aporten en el ámbito social y cultural de manera más efectiva. Esto es importante, ya que las políticas anticorrupción que no consideran las diferencias culturales pueden resultar ineficaces o incluso contraproducentes. Además, comprender la relación entre los valores sociales y la corrupción puede contribuir a nutrir diálogos sociales y eventualmente a promover un cambio cultural hacia la integridad y la transparencia, fortaleciendo así los cuestionamientos respecto a la necesidad de cambiar las normas culturales que toleran o facilitan la corrupción.

### **3. OBJETIVOS**

#### **3.1. GENERAL**

Desarrollar un modelo de aprendizaje de máquina no supervisado utilizando modelos de ecuaciones estructurales (SEM) y técnicas de reducción de dimensionalidad para entender la relación entre los valores sociales y la propensión a la aceptación de sobornos como una forma de corrupción, basándose en los datos de la Ola 7 (2017-2022) de la Encuesta Mundial de Valores (WVS).

#### **3.2. ESPECÍFICOS**

- Explicar el contexto del problema de la corrupción, mediante la realización de un análisis descriptivo inicial y Exploratory Data Analysis (EDA) para familiarizarse con los datos de la WVS, preparar los datos mediante técnicas de limpieza, transformación, reducción de dimensionalidad UMAP e identificar los patrones subyacentes identificados en los datos mediante el uso de clustering.
- Implementar modelos de ecuaciones estructurales (SEM), y evaluarlos para encontrar los resultados más precisos que validen las hipótesis planteadas, y seleccionar el modelo más adecuado para el análisis final.
- Interpretar los resultados obtenidos de los modelos en el contexto de la corrupción y los valores sociales, proponer recomendaciones basadas en los hallazgos del estudio, e identificar áreas para futuros trabajos y mejoras en la metodología utilizada.

#### 4. MARCO TEÓRICO

La corrupción ha sido un fenómeno ampliamente estudiado y evaluado desde diversas perspectivas. Su compleja comprensión y definición se refleja en la variedad de definiciones teóricas y conceptuales que existen. Como señala Rose-Ackerman (Rose-Ackerman & Palifka, 2016) , "la corrupción es un fenómeno multifacético que implica el abuso de poder para obtener beneficios privados ilegítimos". Esta definición sugiere que la corrupción no se limita a contextos específicos, sino que permea todos los niveles de la sociedad y las instituciones. Para Nye, J. S, (Nye, 1967), se destaca la naturaleza subjetiva de la corrupción al afirmar que "la corrupción es un comportamiento que desvía de los deberes formales de un rol, porque el ofensor tiene poder o conocimiento". Aquí, Nye, J. S (Nye, 1967) enfatiza la relación intrínseca entre la corrupción y el ejercicio indebido de la autoridad. Esta perspectiva destaca que la corrupción no es simplemente un acto aislado, sino un patrón de comportamiento que desafía las normas y responsabilidades institucionales y sociales. En un enfoque más reciente, Mungiu-Pippidi, (Mungiu-Pippidi, 2015) ofrece una perspectiva integral al definir la corrupción como "el abuso de recursos públicos para beneficio privado". Esta definición resalta la dimensión económica de la corrupción, subrayando cómo afecta los recursos que deberían destinarse al bienestar general, adicionalmente, destaca la importancia de considerar la corrupción como un fenómeno arraigado en las estructuras sociales y económicas.

La definición de corrupción va más allá de actos individuales de soborno; involucra la erosión de la confianza social y la desviación de los propósitos institucionales. Como señala Klitgaard, (Klitgaard, 1988) "la corrupción es como un impuesto invisible que afecta negativamente la calidad de vida de la sociedad". Aquí, se plantea la idea de que la corrupción no solo tiene costos económicos tangibles, sino que también socava la cohesión social y la percepción de justicia. Desde una perspectiva histórica, hablar sobre corrupción ha sido relevante. En la antigüedad, Aristóteles proporcionó una de las primeras reflexiones en su obra "Política". La corrupción estaba intrínsecamente vinculada a la decadencia de las formas de gobierno, y la búsqueda desenfrenada del interés propio en detrimento del bien común era el núcleo de la corrupción política según, (Medina y Navarro, 1873).

Esta perspectiva antigua estableció las bases para entender la corrupción como un fenómeno que amenaza la estabilidad y la integridad de las instituciones de gobierno. Durante la Edad Media, la corrupción adquirió dimensiones morales y religiosas. Tomás de Aquino, (Tomás de Aquino, 2005) abordó la corrupción desde una perspectiva teológica, identificándola como un pecado que socava la moralidad de la sociedad. La corrupción se asoció con la desviación de los principios éticos y la traición de los valores religiosos, consolidando la visión de la corrupción como un mal moral y espiritual.

La Ilustración del siglo XVIII introdujo nuevas perspectivas sobre la corrupción, vinculándola a conceptos de iluminación, razón y progreso. Montesquieu exploró la relación entre la corrupción y la concentración excesiva de poder, abogando por la separación de poderes como un medio para prevenir la corrupción (Montesquieu, 1748). Este enfoque ilustrado sentó las bases para entender la corrupción en términos de estructuras políticas y mecanismos de control. En el siglo XX, la teoría crítica y el surgimiento de la sociología como disciplina académica ampliaron las dimensiones del concepto de corrupción. La Escuela de Fráncfort, con pensadores como Horkheimer, examinó la corrupción en el contexto de la alienación social y la manipulación de la verdad (Horkheimer et al., n.d.).

La corrupción no solo se percibía como un fenómeno político, sino como una manifestación de las disfunciones sociales más amplias. En la actualidad, la evolución del concepto de corrupción ha llevado a enfoques más holísticos. La obra de Paolo Mauro destaca la importancia de considerar la corrupción como un fenómeno económico, vinculado a la eficiencia y la distribución de recursos (Mauro, 1995). Esta perspectiva contemporánea reconoce la interconexión de factores económicos, políticos y sociales en la comprensión de la corrupción. El estudio de la corrupción evoluciona más allá de las explicaciones puramente estructurales o institucionales, adentrándose en la psicología para comprender las motivaciones individuales que se encuentran para comportamientos corruptos.

Los enfoques psicológicos en la corrupción exploran las dimensiones internas de los actores involucrados, desentrañando las complejidades de las decisiones éticas y las presiones psicológicas que pueden contribuir a la corrupción. Uno de los principales contribuyentes a esta línea de investigación es Hume (Hume, 1777), quien postuló que los juicios morales no son el resultado exclusivo de la razón, como sostenían muchos filósofos de su tiempo, sino que están intrínsecamente ligados a los sentimientos humanos. Hume argumentó que los juicios morales se derivan de las emociones y los sentimientos, y no de la razón pura. Sostuvo que las percepciones morales no se pueden deducir a partir de premisas puramente racionales, sino que están arraigadas en las experiencias emocionales y en la simpatía hacia los demás.

Desde un enfoque más contemporáneo, la teoría de la disonancia cognitiva, desarrollada por Leon Festinger, se ha aplicado para entender la corrupción. La teoría postula que las personas buscan la coherencia entre sus creencias y sus acciones, y cuando hay discrepancias, experimentan malestar psicológico (Festinger, 1957). En el contexto de la corrupción, este malestar podría llevar a individuos involucrados en actos corruptos a racionalizar o justificar sus acciones para reducir la disonancia cognitiva. Otro aspecto relevante es la influencia de los factores emocionales en la corrupción. En el campo de la toma de decisiones se destaca la conexión entre las emociones y la toma de decisiones éticas en situaciones de corrupción (Tenbrunsel & Smith-Crowe, 2008). La vergüenza, la culpa y la falta de empatía pueden desempeñar un papel crucial en cómo los individuos perciben y responden a sus propios comportamientos corruptos. Este

enfoque subraya la importancia de comprender la complejidad emocional detrás de las acciones corruptas. Además, la teoría de la psicología moral, desarrollada por Haidt (Haidt, 2001), ofrece perspectivas valiosas sobre cómo las personas evalúan la moralidad de sus acciones. La psicología moral sugiere que las personas toman decisiones éticas basadas en intuiciones emocionales, más que en un razonamiento lógico. Aplicado a la corrupción, este enfoque destaca la influencia de las emociones morales en la percepción y la justificación de comportamientos corruptos.

Algunos estudios han aportado evidencia empírica sobre la importancia de los valores sociales en la percepción y aceptación de la corrupción, apoyando la hipótesis de que la corrupción no es solo una cuestión legal o económica, sino también cultural y social. En el estudio “Cultural Dimension of Corruption: A Cross-Country Survey”, se observa que aproximadamente el 49% de la variación en los niveles de corrupción entre países puede explicarse por las dimensiones culturales, especialmente la distancia al poder, el individualismo-colectivismo y la orientación a largo plazo, (Achim, 2016). Para Adriana Ana Maria Davidescu y otros, se evidencia que la religiosidad y los valores sociales pueden influir en comportamientos relacionados con la corrupción, como la evasión fiscal, (Davidescu et al., 2022). Para Marcio Salles Melo Lima y Dursun Delen, el uso de técnicas avanzadas de la ciencia de los datos como el Machine Learning permiten identificar los principales predictores de la percepción de la corrupción en más de 132 países utilizando modelos de clasificación multiclase que descubren los factores relevantes relacionados con la corrupción, (Lima & Delen, 2020).

Ahora bien, la Encuesta Mundial de Valores es un proyecto de investigación global que explora los valores y creencias de las personas en diversas sociedades alrededor del mundo. Su objetivo es comprender cómo cambian los valores y cómo estos cambios afectan la vida social, política y económica de las personas. Esta encuesta se ha llevado a cabo desde 1981 en varias olas, hasta la ola 7, que recopiló datos desde 2017 hasta 2022. La mayoría de estos datos fueron recolectados antes de la pandemia por SARS-CoV-2 (COVID-19), hasta 2020, y posteriormente en otros países hasta 2021, alcanzando un total de 64 países con más de 90,000 encuestados. Cuenta con 14 áreas o temas de investigación tales como: Creencias religiosas, actitudes hacia la democracia e instituciones políticas, valores económicos y laborales, percepciones sobre la calidad de vida y el bienestar, entre otros. Usa cuestionarios estandarizados aplicados a muestras representativas de la población adulta en cada país participante. Para realizar este estudio y tener un entendimiento sobre los datos que recopila la Encuesta Mundial de Valores sobre “valores sociales”, es necesario contar con una definición que permita comprender las diversas fuerzas que influyen en el comportamiento humano y así lograr identificar las más preponderantes, que generen una inclinación en la balanza para aceptar o no sobornos y cometer un acto de corrupción.

El punto de referencia que se abordará en este estudio es desde el institucionalismo, planteado por Scott (Scott, 1995), el cual brinda un enfoque teórico que se centra en las estructuras, normas y procesos sociales que configuran y regulan el

comportamiento humano en organizaciones y sociedades. Para Scott, las instituciones se definen como sistemas multifacéticos que integran estructuras cognitivas, normativas y regulatorias, aportando estabilidad y significado al comportamiento social. Con este enfoque es posible comprender la estabilidad social, guiar el comportamiento, estructurar la realidad social y facilitar la coordinación y cooperación, proporcionando un marco para analizar la influencia, el mantenimiento y el cambio de las instituciones a lo largo del tiempo.

El análisis de resultados se realizará mediante el uso de los “tres pilares de las instituciones”, el pilar regulatorio, el pilar normativo y el pilar cognitivo, los cuales permiten una comprensión, profunda y multifacética de cómo las instituciones influyen en las acciones y pensamientos de los individuos y grupos dentro de una sociedad. El pilar regulatorio enfatiza los procesos explícitos de regulación que constriñen y regularizan el comportamiento humano. Según Scott (Scott, 1995), "las instituciones constriñen y regularizan el comportamiento". Los académicos que apoyan este pilar destacan la importancia de los procesos reguladores explícitos, como el establecimiento de reglas, el monitoreo de la conformidad y la imposición de sanciones. Para Douglass North (North, 1990), por ejemplo, define las instituciones como "análogas a las reglas del juego en un deporte de equipo competitivo. Es decir, consisten en reglas formales escritas, así como en códigos de conducta no escritos que subyacen y complementan las reglas formales".

Este enfoque puede observarse en economías de mercado y otras situaciones competitivas, donde los intereses divergentes son comunes y, por lo tanto, las reglas explícitas y los árbitros son necesarios para mantener un orden estable. Así, las instituciones se ven como estructuras que moldean el comportamiento al proporcionar incentivos y restricciones claras. El pilar normativo se centra en las reglas normativas que introducen una dimensión prescriptiva, evaluativa y obligatoria en la vida social. Este pilar incluye valores y normas que definen cómo deben hacerse las cosas y especifican medios legítimos para perseguir fines valorados. En el libro “Rediscovering Institutions: The Organizational Basis of Politics”, se argumenta que "las organizaciones siguen reglas, y gran parte del comportamiento en una organización está especificado por procedimientos operativos estándar", (Hood, 1990).

Los sistemas normativos no solo constriñen, sino que también facultan la acción social al conferir derechos y responsabilidades. Para James Jones, (Jones, 1959) se señala cómo el poder y el misticismo asociados con roles profesionales provienen de la licencia otorgada para participar en actividades "prohibidas" o fatales. En esta perspectiva, el comportamiento humano está dirigido no solo por intereses individuales, sino también por obligaciones sociales y expectativas de rol, que son internalizadas por los actores y reforzadas por las presiones externas.

El pilar cognitivo resalta los elementos cognitivos que constituyen la naturaleza de la realidad y los marcos a través de los cuales se crea el significado. Este pilar incluye construcciones simbólicas y procesos reguladores que configuran el

comportamiento social. "Un enfoque en las dimensiones cognitivas de las instituciones es la característica distintiva del nuevo institucionalismo dentro de la sociología" señala James Jones, (Jones, 1959). Las instituciones son vistas como sistemas que incorporan construcciones cognitivas y reglas normativas, que son internalizadas por los individuos y actúan como guías para el comportamiento. Para la centralidad de los roles en las instituciones: "Toda conducta institucionalizada implica roles" según (Berger & Luckmann, 2011).

Los roles, como expectativas normativas, no solo son anticipaciones de lo que los actores pueden hacer, sino prescripciones de lo que deben hacer. Existen los denominados portadores de estos pilares institucionales, los cuales varían, pero son igualmente importantes para entender cómo las instituciones influyen en el comportamiento humano. Las culturas actúan como portadores interpretativos, proporcionando esquemas de significado codificados y sistemas de reglas que informan y constriñen el comportamiento continuo. Las estructuras sociales, como los sistemas de roles y autoridad, incorporan elementos institucionales y crean marcos normativos y coercitivos que guían la acción. Las rutinas, como comportamientos habituales y programas de desempeño, encapsulan el conocimiento tácito de los actores y facilitan la estabilidad y la previsibilidad en las organizaciones.

En términos de portadores culturales, Scott (Scott, 1995) argumenta que las culturas enfatizan categorías, distinciones y tipificaciones en el caso del pilar cognitivo, y valores y expectativas compartidas en el caso del pilar normativo. Estas culturas pueden ser específicas de una organización, formando lo que se conoce como cultura corporativa, o pueden ser más generales, operando a nivel de campo organizacional o societal. Se introduce el concepto de hábitos para describir cómo las reglas culturales se internalizan, funcionando como una matriz de percepciones, apreciaciones y acciones que estructuran el comportamiento en diferentes situaciones, según (Bourdieu & Nice, 1977).

Las estructuras sociales, como portadores de las instituciones, se basan en expectativas normativas y roles conectados a redes de posiciones sociales. Los teóricos cognitivos enfatizan la isomorfa estructural, donde las tipificaciones cognitivas se codifican en las estructuras organizacionales como departamentos diferenciados y roles. Por otro lado, los teóricos normativos y regulatorios ven las estructuras como sistemas de gobernanza que crean y refuerzan normas y reglas, monitoreando y sancionando las actividades de los participantes. Para Oliver Williamson las estructuras de gobernanza son cruciales para ejercer control y asegurar la conformidad con las normas institucionales (Li, 2019).

Las rutinas, como portadores de las instituciones, se manifiestan en comportamientos estructurados y actividades rutinarias que reflejan el conocimiento tácito de los actores, son como "genes" de las organizaciones, que varían desde actividades codificadas en tecnologías hasta procedimientos operativos estándar. Estas rutinas son esenciales para la estabilidad y la predictibilidad de las

organizaciones, ya que encapsulan el conocimiento y las creencias profundamente arraigadas de los actores. Al considerar estos tres pilares y sus portadores, se puede observar cómo las instituciones influyen profundamente en el comportamiento humano y social. Los individuos y las organizaciones no solo operan dentro de un marco institucional, sino que también están constituidos por él. Las instituciones proporcionan los esquemas cognitivos, normativos y regulatorios que dan forma a las acciones y decisiones, creando un entorno social que es a la vez constrictivo y facultativo. Este marco teórico subraya la importancia de considerar las múltiples dimensiones y niveles en los que operan las instituciones para entender completamente su impacto en el comportamiento humano y social.

Ahora, dado que el objetivo de esta investigación es el de desarrollar un modelo analítico que permita explicar las diferentes relaciones entre valores sociales y corrupción, es importante establecer las definiciones fundamentales que orientarán su construcción y la evaluación de resultados. En primer lugar, el reconocido estadístico Hastie (Hastie et al., 2009) proporciona una definición importante para el aprendizaje estadístico, en la cual establece que es "un conjunto de herramientas estadísticas que utiliza modelos para descubrir patrones en datos y hacer predicciones".

Este enfoque se sitúa en la intersección de la estadística y el aprendizaje automático, donde los algoritmos buscan identificar regularidades en conjuntos de datos para realizar inferencias y generar conocimiento. El contexto del aprendizaje estadístico se inscribe en la creciente necesidad de extraer información significativa a partir de datos complejos y grandes volúmenes de información. El aprendizaje estadístico se ha convertido en una herramienta fundamental en diversas disciplinas, desde la bioinformática hasta la economía, transformando la forma en que abordamos problemas complejos mediante la modelización y la interpretación de datos, según (James et al., 2021). La estimación de una función  $f$  puede estar impulsada por dos objetivos principales: la predicción o la inferencia, según (Hastie et al., 2009).

El objetivo principal de predecir  $f$  es generar pronósticos o estimaciones precisas de la variable de salida basándose en los datos de entrada. En el modelado predictivo, se enfatiza en construir un modelo que pueda generalizar de manera efectiva a datos no vistos. El foco principal es minimizar el error de predicción, asegurando que el modelo pueda realizar predicciones precisas y confiables para nuevas observaciones. En cambio, el objetivo principal de estimar  $f$  para la inferencia es entender e interpretar las relaciones entre las variables de entrada y salida. El enfoque se centra en descubrir la estructura subyacente de los datos, identificar predictores significativos y obtener conclusiones sobre cómo los cambios en las variables de entrada se relacionan con los cambios en la variable de salida. La inferencia permite una comprensión más profunda de los mecanismos subyacentes.

En el ámbito del aprendizaje estadístico orientado a la inferencia, el uso de modelos estadísticos tradicionales como la regresión lineal, propuesto inicialmente por Galton en el siglo XIX, establece una relación lineal entre las variables de entrada y salida. A lo largo del tiempo, se han desarrollado generalizaciones para abordar la complejidad de las relaciones, como la regresión lineal múltiple propuesta por (Fisher, 1922). Estos modelos permiten estimar los coeficientes que cuantifican la fuerza y la dirección de las relaciones entre las variables, brindando interpretaciones sustanciales. El enfoque bayesiano, con raíces de Thomas Bayes en el siglo XVIII, ha ganado terreno en la inferencia estadística actual. Los modelos bayesianos, permiten la incorporación de información a priori en el análisis, mejorando la capacidad de inferir relaciones en conjuntos de datos limitados, como los descritos en (Gelman et al., 2014). La estimación bayesiana ofrece ventajas en términos de manejo de la incertidumbre y proporciona distribuciones posteriores que facilitan la interpretación de los resultados. La evaluación de relaciones en el aprendizaje estadístico para inferencia implica el uso de diversas herramientas y métodos.

Las pruebas de hipótesis, como las propuestas por Pearson en (NEYMAN & PEARSON, 1928), siguen siendo fundamentales para establecer la significancia estadística de las relaciones inferidas. El análisis de sensibilidad en variables ofrece una perspectiva valiosa al identificar la contribución relativa de cada variable a la variabilidad en la salida del modelo, según introducción dada por (Sobol & Kucherenko, 2005). Además, las técnicas de visualización, como las recomendadas por Turkey (Tukey, 1977), tales como, diagrama de tallo y hojas, gráficos de caja (box plots), y el análisis de residuo, permiten una comprensión intuitiva de las relaciones complejas.

## **5. METODOLOGÍA**

La minería de datos y el análisis de datos se han vuelto fundamentales en la investigación académica, especialmente en disciplinas que implican grandes conjuntos de datos. Para esta investigación, la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) (IBM, 2021) será la metodología para seguir al conjunto de datos expuesto por la Encuesta Mundial de Valores (WVS). CRISP-DM consta de seis fases interconectadas: Entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelado, evaluación y despliegue.

### **5.1. ENTENDIMIENTO DEL NEGOCIO**

En esta primera etapa se pretende comprender el contexto en el que se encuentra ubicado el problema a resolver, el cual radica en estudiar los resultados de la Encuesta Mundial de Valores y su relación o influencia en la aceptación de sobornos como una forma de conducta corrupta. Se tomará como sustento los tres pilares de institucionalidad establecidos por Scott (Scott, 1995) y elaborados por Hoffman (Hoffman, 2001): el pilar regulativo, que abarca las leyes y regulaciones formales; el pilar normativo, que incluye los valores y normas sociales que guían el comportamiento; y el pilar cognitivo, que se refiere a las creencias y suposiciones compartidas que configuran cómo las personas perciben y reaccionan ante la corrupción. Este marco teórico será un referente y permitirá evaluar los resultados y su relación con la aceptación de sobornos. Además, es fundamental comprender las limitaciones que se puedan presentar durante el análisis, considerando la disponibilidad y calidad de los datos de la Encuesta Mundial de Valores, así como los posibles sesgos que puedan influir en los resultados. Este análisis integral brindará una base sólida para la formulación de hipótesis específicas y guía la selección adecuada de variables en las etapas posteriores del proceso de ciencia de datos y analítica.

### **5.2. ENTENDIMIENTO DE LOS DATOS**

El enfoque en esta etapa se dirige hacia la exploración y comprensión detallada de la estructura y características específicas de los datos recopilados por la Encuesta Mundial de Valores. El entendimiento de los datos inicia con la identificación y recolección requerida, seguido de una evaluación de su calidad. Esto implica revisar la integridad de los datos, detectar y manejar los valores faltantes, y asegurar la consistencia y precisión de la información. Además, se realizarán análisis descriptivos para comprender las distribuciones de todas las variables y detectar posibles outliers o anomalías que puedan afectar el análisis y selección de variables. Se podrán evaluar técnicas de reducción de dimensionalidad, como PCA (Análisis de Componentes Principales) o UMAP (Proyección Uniforme y Aproximada de

Manifold), para manejar la alta dimensionalidad de los datos, identificar los factores principales y formar clústeres que serán sujetos de análisis detallado.

Durante esta fase se realizan las siguientes actividades:

- i. **Revisión de datos:** Se exploran las diferentes olas (waves) existentes de la Encuesta Mundial de Valores desde 1981 a la fecha. Puntualmente, se busca comprender la estructura de cuestionarios Core y el ámbito de respuesta para la ola 7.
- ii. **Exploración de datos:** Se seleccionan las principales preguntas en cada tema de la encuesta, es importante determinar los temas afines con la definición del problema.
- iii. **Análisis exploratorio de datos (EDA):** Se busca conocer la distribución de los datos, determinar las características estadísticas más relevantes, patrones, correlaciones, valores atípicos y datos faltantes.
- iv. **Definición de variables:** Identificar las características, en este caso, las preguntas relevantes con respecto a la definición dada en el marco teórico y su representatividad con respecto a la influencia que puedan tener sobre la aceptación de sobornos como una forma de corrupción.

### 5.3. PREPARACION DE LOS DATOS

En esta etapa, el enfoque se centra en transformar y limpiar los datos para garantizar que estén listos para el análisis. Este proceso implica abordar posibles problemas identificados durante la fase de anterior. Se realizan tareas como el manejo de valores faltantes mediante imputación o eliminación, la codificación adecuada de variables categóricas y la normalización de datos si es necesario. Además, se pueden crear nuevas variables o características que consoliden información importante para el análisis posterior. Esta preparación de datos podría implicar la creación de índices que capturen la intensidad de ciertos valores o la propensión hacia ciertos comportamientos sujetos de estudio. Se aplicarán técnicas de reducción de dimensionalidad debido a la gran cantidad de variables candidatas al estudio y se espera que las resultantes también requieran aplicar esta técnica. Se espera evaluar técnicas como PCA (Análisis de Componentes Principales) o UMAP (Proyección Uniforme y Aproximada de Manifold) y resaltar las estructuras más importantes. Además, se espera realizar análisis de clustering para identificar grupos y patrones dentro de los datos, permitiendo descubrir relaciones significativas sin la necesidad de etiquetas predefinidas.

### 5.4. MODELADO

Se usarán modelos de aprendizaje no supervisado para explorar y entender la aceptación de sobornos a partir de los valores sociales recopilados en la Encuesta Mundial de Valores. Primero, se implementarán Modelos de Ecuaciones Estructurales (SEM, Structured Equations Model) junto con técnicas de reducción

de dimensionalidad como UMAP. Los modelos SEM son particularmente adecuados para este estudio porque permiten analizar relaciones complejas entre múltiples variables observadas y latentes. Con SEM, podemos construir y validar modelos teóricos que expliquen cómo diferentes valores sociales y factores contextuales influyen en la aceptación de sobornos. Estos modelos nos permiten evaluar simultáneamente relaciones directas e indirectas, proporcionando una visión más completa y matizada de los procesos subyacentes.

El modelo construido se sustenta dado el marco teórico basado en los tres pilares de institucionalidad de Scott (Scott, 1995): el regulativo, el normativo y el cognitivo. Este modelo incluirá variables latentes que representan constructos teóricos clave, así como variables observadas derivadas de la Encuesta Mundial de Valores. Una vez definido el modelo teórico, se realizará la estimación y validación del modelo utilizando técnicas SEM. Esto incluye la evaluación de la bondad de ajuste del modelo, asegurando que los datos empíricos apoyen la estructura teórica propuesta.

Se usan las medidas como el Chi-cuadrado, CFI (Índice de Ajuste Comparativo) y RMSEA (Error de Aproximación de Raíz Cuadrada) para evaluar el ajuste del modelo. Se combinarán los hallazgos de los modelos SEM y las técnicas de reducción de dimensionalidad para obtener una visión integral de cómo los valores sociales influyen en la aceptación de sobornos. Esta combinación permitirá validar las hipótesis teóricas, descubrir nuevos patrones y relaciones que podrían no ser evidentes a partir de un único enfoque metodológico. Estos análisis proporcionarán evidencia desde los datos para comprender y apoyar el marco teórico dado por Scott (Scott, 1995) y brindarán herramientas valiosas para futuros trabajos de investigación en el ámbito del estudio de la corrupción y la institucionalidad.

Se usarán modelos de aprendizaje no supervisado para explorar y entender la aceptación de sobornos a partir de los valores sociales recopilados en la Encuesta Mundial de Valores. Primero, se implementarán Modelos de Ecuaciones Estructurales (SEM, Structured Equations Model) junto con técnicas de modelado más adecuadas para los objetivos del estudio. Dado que el interés es descubrir patrones no evidentes sin depender de etiquetas predefinidas, se usarán métodos como el clustering y UMAP (Proyección Uniforme y Aproximada de Manifold).

## **5.5. EVALUACION**

Esta fase es importante para asegurar que los modelos desarrollados en la fase de Modelado cumplen con los objetivos del estudio y proporcionan resultados confiables y útiles. En este estudio, la evaluación se centrará en verificar la validez y la robustez de los modelos de Ecuaciones Estructurales (SEM) y de las técnicas de reducción de dimensionalidad aplicadas. Se esperan resultados coherentes y para esto se utilizarán métricas y coherencia visual en los gráficos de UMAP para asegurar que las estructuras identificadas sean representativas y útiles para el análisis. Se compararán los resultados obtenidos con el marco teórico de Scott

(Scott, 1995) para evaluar la coherencia teórica. Se revisará si los patrones y relaciones descubiertos apoyan las hipótesis teóricas y cómo contribuyen a una comprensión más profunda de los pilares regulativo, normativo y cognitivo en la aceptación de sobornos.

Se analizarán los coeficientes de las rutas para determinar la significancia y la dirección de las relaciones entre las variables latentes y observadas.

Para las técnicas de reducción de dimensionalidad, como PCA (Análisis de Componentes Principales) y UMAP (Proyección Uniforme y Aproximada de Manifold), se evaluará cómo simplifican la complejidad de los datos y resaltado las estructuras subyacentes importantes. Las evaluaciones incluirán:

- **Proporción de varianza explicada:** En PCA, se verificará cuánta de la varianza total de los datos es explicada por los primeros componentes principales. Un valor alto sugiere que la técnica captura la mayoría de la información relevante.
- **Visualización de estructuras:** En UMAP, se examinarán las visualizaciones para identificar patrones significativos y agrupaciones en los datos que reflejen relaciones coherentes y significativas.
- **Coherencia de clustering:** Al usar clustering sobre los resultados de reducción de dimensionalidad, se evalúa la coherencia de los clústeres formados mediante métricas como la silueta, que mide la separación y compactación de los clústeres.
- **Estabilidad y reproducibilidad:** Se comprobará si los resultados obtenidos son consistentes cuando se aplican en diferentes subconjuntos de datos o con distintos parámetros, asegurando así la robustez de los hallazgos.

## 5.6. DESPLIEGUE

En esta fase los resultados y hallazgos del estudio se integran y se presentan de manera que puedan ser utilizados para la toma de decisiones y futuras investigaciones. En el contexto de este estudio sobre la relación entre los valores sociales y la aceptación de sobornos, el despliegue se centrará en comunicar claramente los resultados y proporcionar herramientas que faciliten su aplicación y análisis continuo. Se presenta un análisis que incluye una descripción completa de la metodología implementada, los métodos utilizados y los hallazgos principales. Una descripción detallada de los métodos de modelado y reducción de dimensionalidad empleados, así como de las métricas de evaluación utilizadas. Finalmente, una interpretación de los resultados en el contexto del marco teórico de Scott (Scott, 1995) y su implicación para la comprensión de la corrupción.

## 6. PRESENTACION DE RESULTADOS

### 6.1. EXPLICACIÓN DEL PROBLEMA E HIPÓTESIS

La corrupción, y específicamente la aceptación de sobornos, representa uno de los mayores desafíos para el desarrollo económico y social de cualquier país. Este estudio se enfoca en entender cómo los valores sociales influyen en la aceptación de sobornos como una conducta corrupta. Utilizando datos de la Encuesta Mundial de Valores (WVS), se busca identificar las relaciones entre diferentes valores sociales y la disposición de los individuos a aceptar sobornos. Tomando como referencia la teoría institucional contemporánea, este estudio se fundamenta en los tres pilares principales del institucionalismo establecidos por Scott (Scott, 1995) y elaborados por Hoffman (Hoffman, 2001):

- **Pilar Regulatorio:** Este pilar abarca las leyes y regulaciones formales que buscan controlar el comportamiento de los individuos en la sociedad. Por ejemplo, la existencia de leyes anticorrupción y organismos de supervisión como comisiones de control pueden disuadir a las personas de aceptar sobornos. La eficacia de este pilar depende de la estricta implementación y aplicación de estas regulaciones. Países con sistemas judiciales fuertes y transparentes tienden a tener menores niveles de corrupción debido a la firmeza con que se aplican las sanciones a los infractores.
- **Pilar Normativo:** Incluye los valores y normas sociales que guían el comportamiento de los individuos, estas normas no están escritas. Por ejemplo, en culturas donde la honestidad y la integridad son altamente valoradas, es menos probable que los individuos acepten sobornos. Las normas sociales también se refuerzan a través de la educación y la socialización, donde desde temprana edad se enseña a las personas sobre la importancia de actuar con ética. Las sociedades donde prevalecen valores como la lealtad y la reciprocidad también pueden influir en la aceptación de sobornos si estos valores se orientan más hacia el beneficio personal que hacia el bien común.
- **Pilar Cognitivo:** Se refiere a las creencias y suposiciones compartidas que configuran cómo las personas perciben y reaccionan ante la corrupción. Por ejemplo, en algunos contextos, la aceptación de sobornos puede ser vista como una práctica común o necesaria para avanzar en la vida profesional, lo que normaliza esta conducta. La percepción de que "todos lo hacen" puede llevar a la aceptación generalizada de la corrupción. Este pilar también incluye el conocimiento y la concienciación sobre los efectos negativos de la corrupción, lo cual puede variar significativamente entre diferentes culturas y contextos socioeconómicos.

La hipótesis del estudio afirma que los tres pilares del institucionalismo (regulatorio, normativo y cognitivo) influyen significativamente en la aceptación de sobornos como conducta corrupta. Específicamente, se enuncia que:

- Las regulaciones formales (pilar regulativo) tienen un efecto disuasorio en la aceptación de sobornos.
- Los valores y normas sociales (pilar normativo) que condenan la corrupción reducen la probabilidad de que los individuos acepten sobornos.
- Las creencias y suposiciones compartidas (pilar cognitivo) sobre la corrupción influyen en la percepción y disposición de los individuos a aceptar sobornos.

## 6.2. ANÁLISIS EXPLORATORIO DE DATOS

La Encuesta Mundial de Valores (WVS) proporciona una base de datos extensa sobre valores y creencias sociales a nivel global. La encuesta, que desde el año 1.981, ha recopilado datos de más de 80 países y territorios, ha permitido explorar los valores y creencias de las personas en diferentes contextos culturales y sociales, incluyendo actitudes hacia la corrupción, normas sociales, valores éticos y percepciones sobre las formas de gobierno y participación en el. A continuación, se presenta la tabla 1 con la relación de olas realizadas a la fecha de este estudio.

Ola	Periodo	Países encuestados
1	1981-1984	11
2	1990-1994	21
3	1995-1998	55
4	1999-2004	41
5	2005-2009	58
6	2010-2014	60
7	2017-2022	66

*Tabla 1. Olas de la WVS*

Para este estudio, se ha tomado datos de la de la Ola número 7, la cual en su mayoría de encuestas se completaron entre 2018 y 2020. Unos pocos países, aproximadamente una docena, llevaron a cabo su trabajo de campo durante la postpandemia de COVID-19, entre 2021 y 2022, La última encuesta de esta ola proviene de India y se completó en julio de 2023, los datos de esta ola son esenciales para comprender las dinámicas sociales, políticas y culturales a nivel global, especialmente en el contexto de eventos recientes como la pandemia de COVID-19.

Ola	7	
Versión	5.0	
Nombre de archivo	WVS Cross-National Wave 7 csv v5 0.zip	
Fuente	<a href="https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp">https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp</a>	
Data shape	94.278 filas x 606 columnas	
Variables totales	606 preguntas	
Variables candidatas a estudio	305 preguntas	
14 bloques temáticos	Valores sociales, actitudes y estereotipos	48 preguntas
	Bienestar social	11 preguntas
	Capital social, confianza y pertenencia a asociaciones	50 preguntas
	Valores económicos	6 preguntas
	Corrupción	9 preguntas
	Migración	10 preguntas
	Seguridad	21 preguntas
	Índice postmaterialismo	6 preguntas
	Ciencia y tecnología	6 preguntas
	Valores religiosos	12 preguntas
	Valores éticos y normas	23 preguntas
	Interés por la política y participación política	39 preguntas
	Cultura y sistemas políticos	25 preguntas
	Datos sociodemográficos	39 preguntas

Tabla 2. Características generales del conjunto de datos

Para las 305 preguntas candidatas del estudio presentadas en la tabla 2, la Encuesta Mundial de Valores determina su modalidad del tipo:

1. **Escala Likert:** Los encuestados indican su nivel de acuerdo o desacuerdo con una declaración en escalas de 1 a 10, de 1 a 4 o de 1 a 3. Por ejemplo: “Q50: ¿En qué medida está Ud. Satisfecho o insatisfecho con la situación económica de su hogar?” En una escala de 1 a 10, donde 1 significa completamente insatisfecho y 10 significa que completamente satisfecho.
2. **Preguntas Dicótomas:** Estas son preguntas de sí/no o verdadero/falso. Por ejemplo: “Q151: Todos esperamos que no haya otra guerra, pero si la hubiera, ¿estaría Ud. Dispuesto a luchar por su país?” 1.Sí 2.No
3. **Preguntas categóricas:** Son preguntas en las cuales se encuentra una opción predefinida. Por ejemplo: “Q290: Grupo étnico” 1. Afrocolombiano, 2. Gitano, 3. Indígena, 4. Ninguno, 5. Otro.

Por lo tanto, las respuestas en general son cuantitativas del tipo ordinal, algunas con escala de Likert o con escala de frecuencia. En ese sentido, se observan que

de las 305 preguntas candidatas del estudio algunas son del tipo categórico y para este estudio no se le dará alcance, es decir, no harán parte del análisis. Se presenta la tabla 3 con los resultados de las preguntas candidatas.

Valores sociales, actitudes y estereotipos	No se toman: Q33_3, Q34_3, Q35_3, las cuales son de escala 3	45 preguntas
Bienestar social	Se toman todas	11 preguntas
Capital social, confianza y pertenencia a asociaciones	No se toma: Q82, son organizaciones regionales	48 preguntas
Valores económicos	Se toman todas	6 preguntas
Corrupción	Se toman todas	9 preguntas
Migración	Se toman todas	10 preguntas
Seguridad	Se toman todas	21 preguntas
<u>Índice postmaterialismo</u>	Se toman todas	6 preguntas
Ciencia y tecnología	Se toman todas	6 preguntas
Valores religiosos	Se toman todas	12 preguntas
Valores éticos y normas	Se toman todas	23 preguntas
Interés por la política y participación política	No se toman: Q223, Q223_ABREV, Q223_LOCAL, son listados de partidos políticos, Q234A	35 preguntas
Cultura y sistemas políticos	Se toman todas	25 preguntas
Datos sociodemográficos	No se tomaron preguntas que son del tipo categórico.	17 preguntas
	<b>TOTAL</b>	<b>274 preguntas</b>

Tabla 3. Evaluación de variables candidatas

Adicionalmente se observa que el ámbito de respuesta no solo corresponde al establecido por cada pregunta, sino que también se brindan opciones genéricas con las siguientes opciones:

“La codificación general para los códigos faltantes es la siguiente:”

- -1 No sabe
- -2 No responde/rechaza
- -3 No aplicable (filtro)
- -5 Faltante; No aplicable por otras razones

Lo cual supone con certeza un tratamiento de datos sin respuesta que es necesario evaluar y determinar su proporción con respecto al total por cada pregunta. A continuación, se presentan las preguntas que tienen una proporción de datos faltantes superior al 10%. Estas preguntas no podrán ser parte del estudio debido a que a que pueden inducir un sesgo significativo en los análisis estadísticos según



<b>Bloques temáticos</b>	<b>Preguntas candidatas</b>	<b>Preguntas sin datos faltantes</b>	<b>Preguntas con correlaciones cercanas a cero</b>
Valores sociales, actitudes y estereotipos	45	44	27
Bienestar social	11	11	11
Capital social, confianza y pertenencia a asociaciones	48	38	37
Valores económicos	6	6	6
Corrupción	9	9	9
Migración	10	10	10
Seguridad	21	20	18
Índice postmaterialismo	6	6	6
Ciencia y tecnología	6	6	5
Valores religiosos	12	12	12
Valores éticos y normas	23	23	21
Interés por la política y participación política	35	25	24
Cultura y sistemas políticos	25	24	23
Datos sociodemográficos	17	12	11
<b>TOTAL</b>	<b>274</b>	<b>246</b>	<b>220</b>

*Tabla 5. Evaluación de preguntas candidatas*

Se ha pasado de iniciar con 274 preguntas candidatas a finalizar con 220 preguntas, con lo cual, se tiene el insumo necesario para iniciar el estudio estadístico de los datos. Los resultados gráficos detallados para cada bloque temático y para cada pregunta se pueden observar en la sección de presentación de resultados, numeral desplegable.

### **6.3. PROCESAMIENTO DE DATOS**

#### **6.3.1. Análisis de datos faltantes e imputación**

El análisis exploratorio de datos reveló ciertas preguntas con más del 10% de datos ausentes, las cuales se retiraron del estudio. No obstante, es imprescindible tratar de manera particular estas preguntas sin respuestas para asegurar que los modelos propuestos sean válidos e íntegros. La imputación de datos es por tanto un procedimiento esencial en la preparación de datos, ya que ayuda a abordar problemas de valores faltantes o imprecisos. Sin un manejo adecuado de los datos ausentes, el análisis podría resultar sesgado y llevar a interpretaciones incorrectas. Como señalan (Donald B. Rubin & Roderick J. A. Little, 2014), la imputación contribuye a mantener la fuerza estadística y reduce el sesgo generado al descartar casos con información incompleta.

Las respuestas de tipo ordinal seleccionadas para el estudio hacen que la moda sea una medida de tendencia central fiable, muy conveniente para la imputación de datos faltantes o atípicos. Según (R. J. A. Little & Rubin, 2002), la imputación simple implica la sustitución de los datos ausentes en un conjunto con estimaciones individuales, como la media, mediana o moda. Esta técnica es útil en investigaciones donde los datos faltantes son escasos y en el caso de variables continuas o discretas sencillas, aunque, puede introducir sesgos si los datos no están distribuidos normalmente.

Para este estudio, la suposición de normalidad no es apropiada pues los datos ordinales no siguen una distribución normal y por lo tanto las pruebas de normalidad no son aplicables. Por lo tanto, se procede a realizar la imputación por la media y se realiza una prueba estadística no paramétrica de Mann-Whitney U la cual no asume normalidad en los datos.

La prueba de Mann-Whitney U es una herramienta estadística no paramétrica utilizada para comparar dos muestras independientes y determinar si una de ellas tiende a tener valores mayores que la otra. Para (McKnight & Najab, 2010), el proceso implica la combinación de datos de ambos grupos (A y B), asignación de rangos a todos los valores, cálculo de la suma de rangos para cada grupo y, finalmente, la utilización de estas sumas para obtener la estadística U y se compara con el p-valor para determinar la significancia estadística. Tras la imputación usando la moda, se mantiene la naturaleza ordinal de los datos, lo que es importante para interpretarlos correctamente y prevenir sesgos derivados del uso de métodos más elaborados que presuponen ciertas distribuciones específicas.

Para el análisis del procesamiento de datos por imputación de la moda en respuestas con valores negativos, se definen las siguientes hipótesis:

### **Hipótesis Nula ( $H_0$ )**

La hipótesis nula establece que las distribuciones de los dos grupos A y B son iguales. Es decir, no hay diferencia significativa en las distribuciones de las dos poblaciones, lo que significa que las distribuciones de los dos grupos son iguales.

### **Hipótesis Alternativa ( $H_1$ )**

Las distribuciones de los dos grupos no son iguales.

Se establece un nivel de significancia de **p-value 0.05**. Si p-value es alto (mayor a 0.05) no se rechaza la hipótesis nula, es decir, no hay evidencia suficiente para concluir que hay una diferencia significativa entre las distribuciones de los dos grupos. Si p-value es bajo (menor o igual a 0.05) se rechaza la hipótesis nula, es decir, hay evidencia suficiente para concluir que existe una diferencia significativa entre las distribuciones de los dos grupos. A continuación, se presentan las preguntas para las cuales el p-value es bajo, tabla 6, y las ilustraciones 1, 2 y 3, que permiten visualizar su proporción.

Pregunta	U original	p-value original	U imputado	p-value imputado
Q50	23633083,5	0,0348	23633083,5	0,0348
Q69	780784276	0,0070	828688306	0,0074
Q76	641459615	0,0176	710664319	0,0317
Q149	1057533930	0,0104	1102180529	0,0131
Q285	1061200316	0,0011	1122120326	0,0021

Tabla 6. p-values menores a 0.05 en imputación de datos

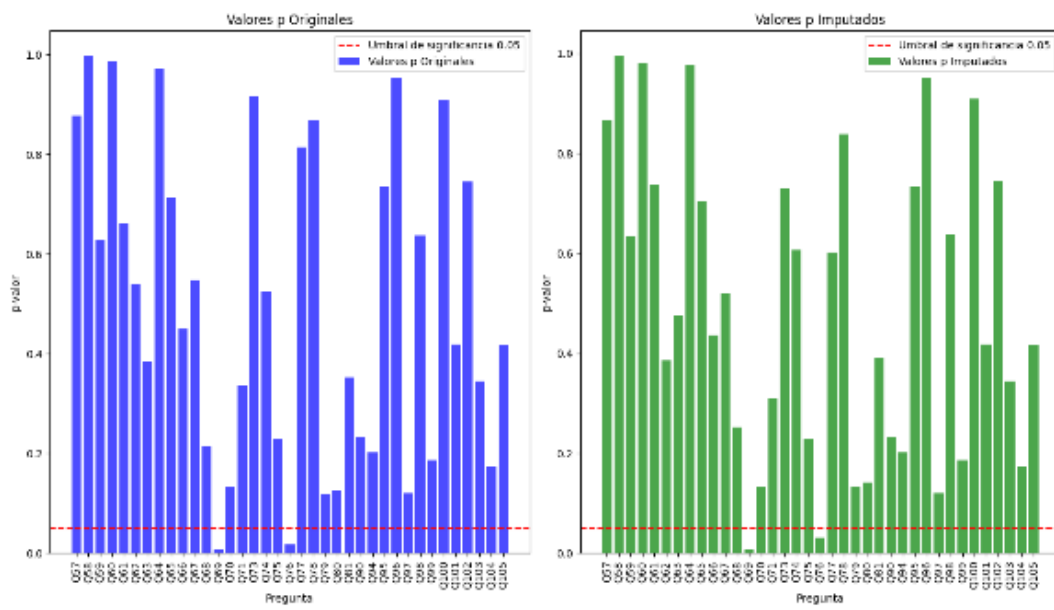


Ilustración 1. Comparación p-value bloque temático capital social

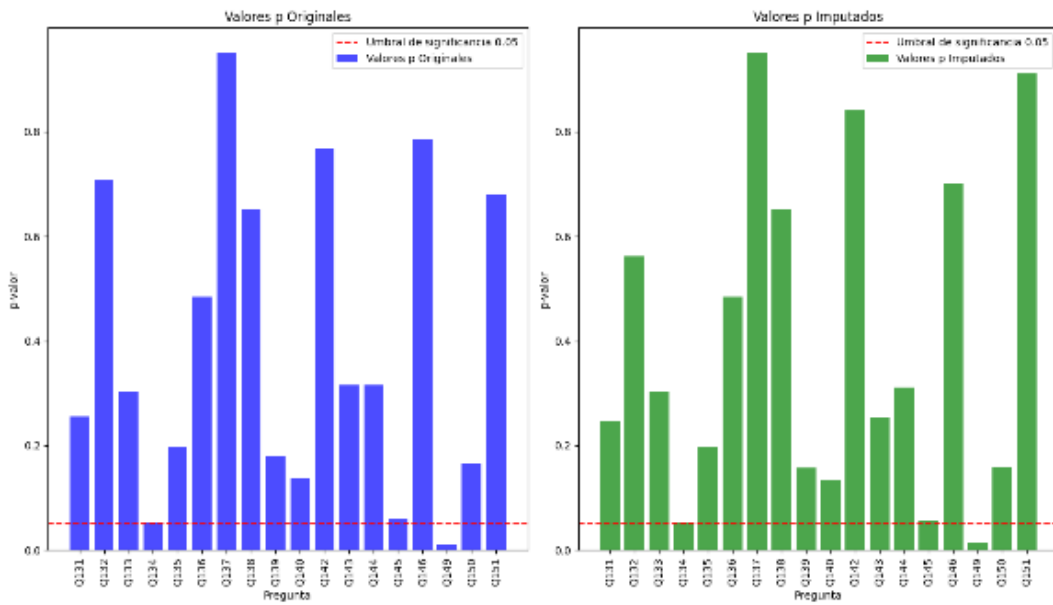


Ilustración 2. Comparación p-value bloque temático seguridad

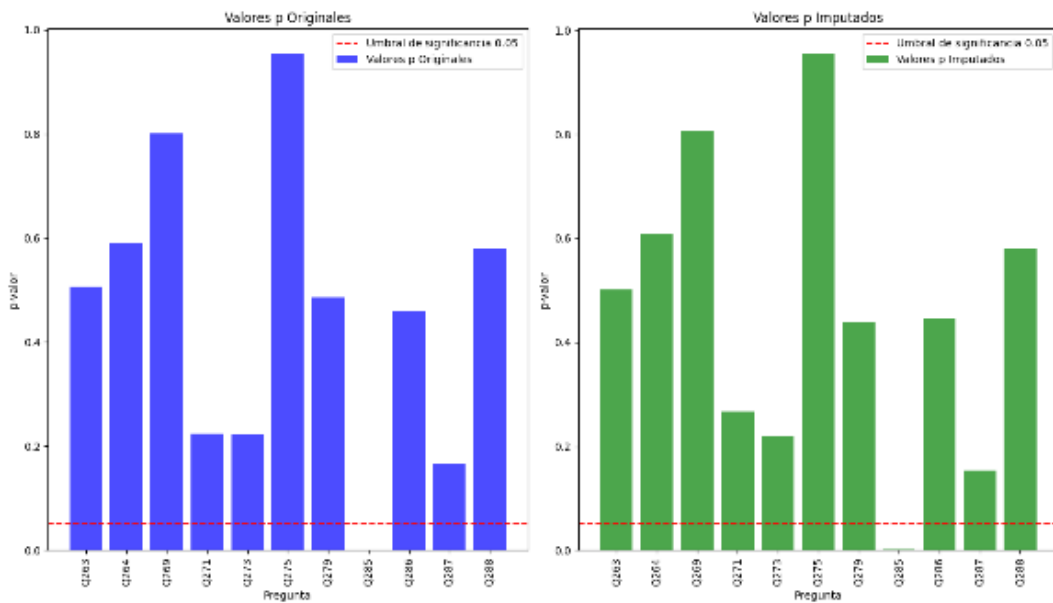


Ilustración 3. Comparación p-value bloque temático demografía

Las preguntas Q50, Q69, Q76, Q149 y Q285 fueron imputadas pero los resultados de la prueba Mann-Whitney U nos indican que pueden verse afectados por sesgos significativos que advierten sobre la validez de los resultados que los modelos propuestos puedan generar. Por lo tanto, se eliminan estas preguntas del estudio y se puede establecer con certeza que los resultados de las preguntas con valores

negativos imputadas por la moda son válidos y consistentes para continuar con el análisis y modelado. Ahora se cuenta con 215 preguntas para iniciar el estudio.

### **6.3.2. Reducción de dimensionalidad**

Continuando con la preparación de datos, se ha optado por emplear la técnica UMAP (Uniform Manifold Approximation and Projection) (McInnes et al., 2020) para la reducción de la dimensionalidad, debido a su eficacia superior en comparación con otras técnicas como el PCA (Principal Component Analysis) y t-SNE (t-distributed Stochastic Neighbor Embedding), sobre todo al trabajar con datos ordinales, caso particular del presente estudio. Los datos de la Encuesta Mundial de Valores (WVS) son complejos y tienen una alta dimensionalidad, 215 preguntas para estudiar, con los que se recopilan las percepciones y actitudes sociales mediante respuestas ordinales. Gracias a que UMAP conserva las estructuras globales y locales, resulta especialmente conveniente para estos tipos de datos, ya que permite una visualización y un análisis más efectivos.

### **6.3.3. Estandarizar o no estandarizar**

UMAP es eficaz con datos ordinales debido a su capacidad para preservar tanto la estructura global como la local de los datos. Según la documentación oficial de UMAP (*UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction — Umap 0.5 Documentation*, n.d.), aunque UMAP puede manejar datos ordinales de manera efectiva, también se recomienda la estandarización para asegurar que todas las variables tengan el mismo peso en el análisis. Por lo tanto, se evalúan ambos casos, con y sin estandarización de los datos, para determinar en cuál caso UMAP realiza una reducción de dimensionalidad evitando la menor pérdida de relación entre los datos. En las ilustraciones 4 y 5 se presenta el resultado de la aplicación de UMAP con datos sin estandarizar y con datos estandarizados en donde es posible se encuentra que sin estandarizar los puntos están más concentrados y compactos, pero con menos separación clara, con estandarización hay más variación en la dispersión, lo que podría indicar una mejor separación de las estructuras en los datos.

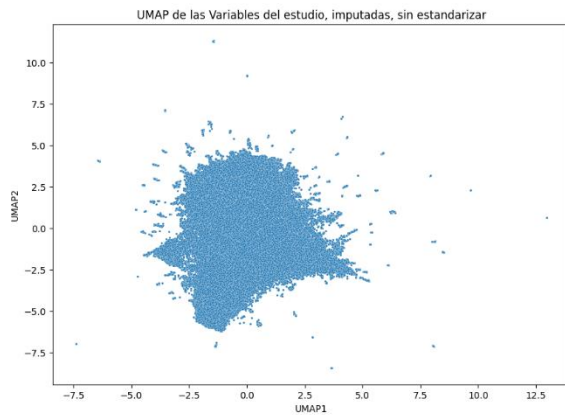


Ilustración 4. UMAP con datos sin estandarizar

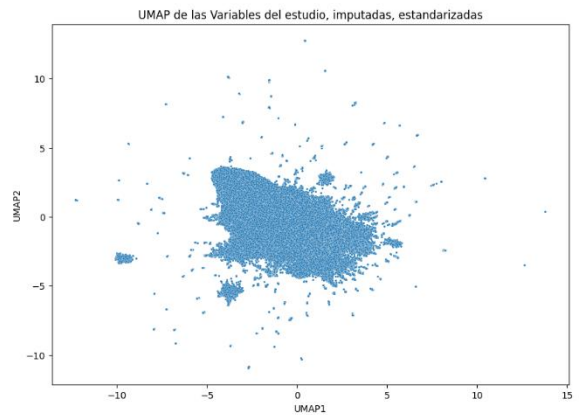


Ilustración 5. UMAP con datos estandarizados

Sin embargo, no es suficiente con la observación visual, es necesario aplicar métricas cuantitativas que brinden un análisis más preciso. Para esto, se procede a implementar un algoritmo de clustering como el K-Means con el cual se puedan generar etiquetas de clústeres para cada punto de los embeddings generados por UMAP, estas etiquetas son necesarias para calcular el coeficiente de silueta, métrica usada para identificar cuál de las dos reducciones de dimensionalidad producen clústeres más definidos y separados entre ellos. Sin embargo, elegir el número adecuado de clústeres para el algoritmo de K-Means es necesario y para esto se utilizan dos técnicas especiales, la técnica del codo y la de la silueta, las cuales son presentadas en las ilustraciones 6, 7, 8 y 9 a continuación:

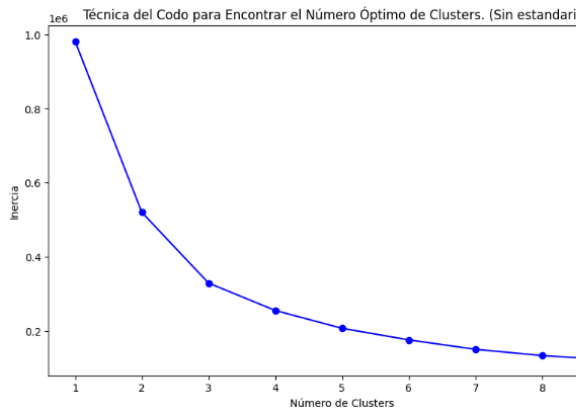


Ilustración 6. Técnica del codo para encontrar número óptimo de clústeres con datos sin estandarizar

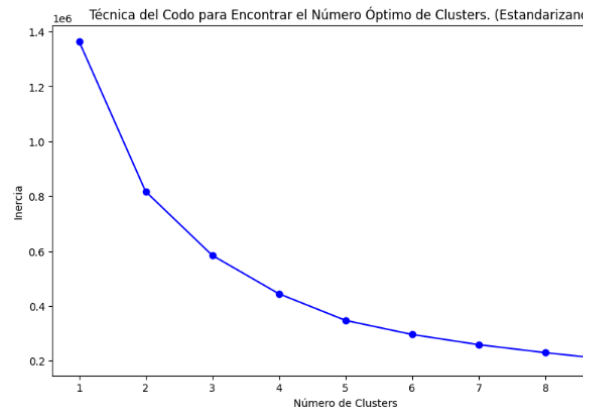


Ilustración 7. Técnica del codo para encontrar número óptimo de clústeres con datos estandarizados

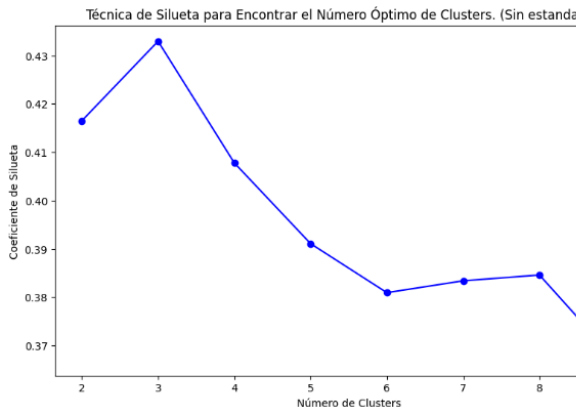


Ilustración 8. Técnica de la silueta para encontrar número óptimo de clústeres con datos sin estandarizar

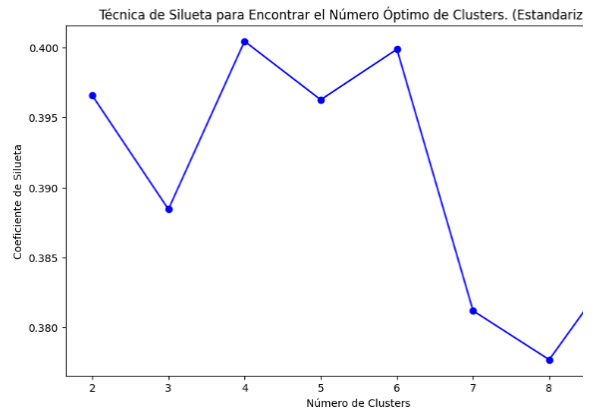


Ilustración 9. Técnica de la silueta para encontrar número óptimo de clústeres con datos estandarizados

Se puede observar que el número óptimo de clústeres es de tres, en especial para los datos sin estandarizar. Pero, aún es necesario revisar en detalle los valores del análisis del coeficiente de silueta el cual, proporciona una visión clara sobre la eficacia de la estandarización en la formación de clústeres para el presente estudio. El coeficiente de silueta para los datos sin estandarización presentado en la tabla 7 es de 0.4318, mientras que para los datos estandarizados es de 0.4042. Estos resultados indican que los datos sin estandarización presentan una mejor cohesión y separación de los clústeres, reflejando una mayor claridad y robustez en la estructura de estos, según (Rousseeuw, 1987).

La estandarización, aunque generalmente recomendada para asegurar que todas las variables tengan el mismo peso en el análisis, en este caso, no proporcionó una mejora en la definición de los clústeres. Por lo tanto, tomando como referencia el coeficiente de silueta, se concluye que **es más adecuado utilizar los datos sin estandarización para el análisis de clustering en este estudio**, alineándose con

la capacidad de UMAP para manejar datos ordinales de manera efectiva vistos en (McInnes et al., 2020).

Para datos sin estandarización	0.4318
Para datos estandarizados	0.4042

Tabla 7. Cálculo del coeficiente de silueta

En las ilustraciones 10 y 11 se pueden observar los tres clústeres generados para datos no estandarizados y estandarizados.

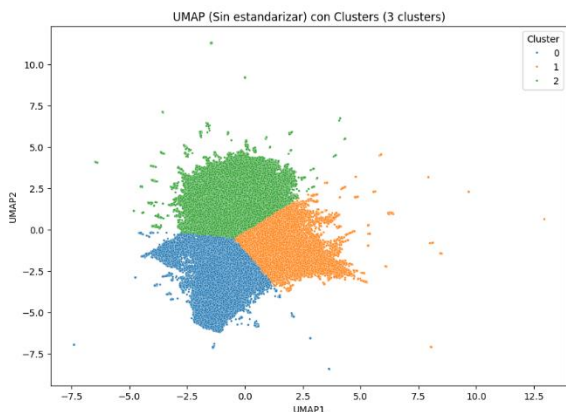


Ilustración 10. UMAP con datos sin estandarizar con 3 clústeres

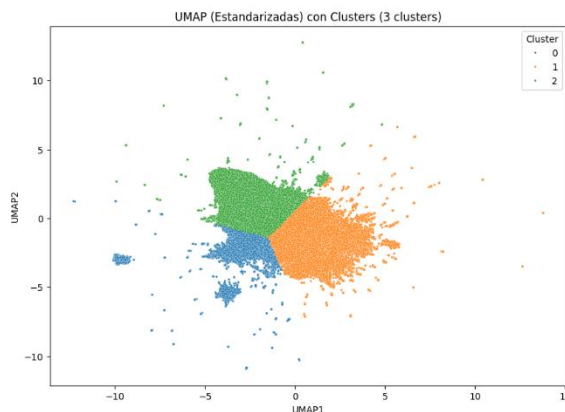


Ilustración 11. UMAP con datos estandarizados con 3 clústeres

Finalmente, el número óptimo de clústeres encontrados tienen mucho sentido con relación al marco teórico expuesto en donde se establecen los tres pilares del institucionalismo, Regulatorio, Normativo y Cognitivo, permitiendo así iniciar un análisis de composición y los siguientes pasos para la modelación.

#### 6.3.4. Análisis de la composición de los clústeres y su relación con el marco teórico

El proceso de clusterización tiene como fin contrastar las agrupaciones de preguntas conformadas a partir de los patrones subyacentes encontrados en los datos contra la agrupación de las variables determinadas por el marco teórico planteado. Los resultados de la composición de los clústeres se observan en la ilustración 12 en donde, con 88 preguntas, el clúster 0 tiene la mayor prevalencia en el mayor número de preguntas, seguido del clúster 1 con 79 preguntas y finalmente el clúster 2 con 48 preguntas.

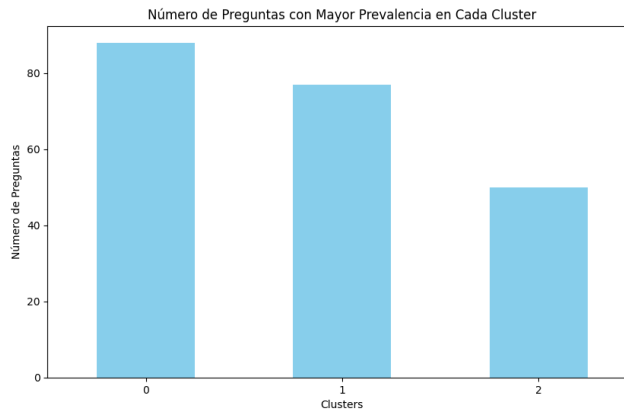
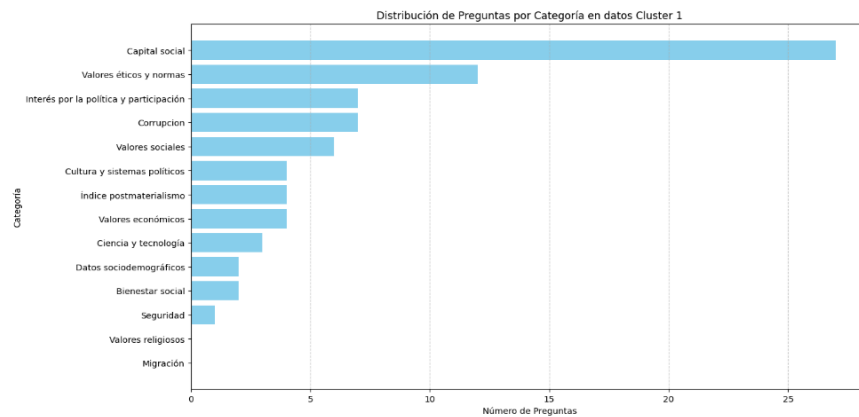
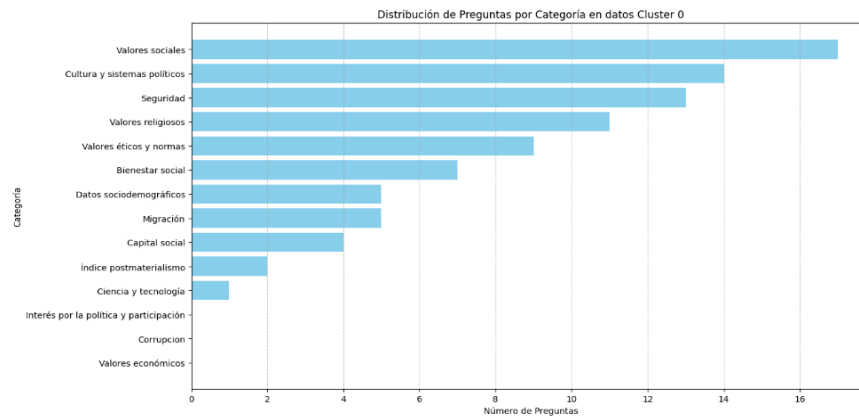


Ilustración 12. Preguntas con mayor prevalencia por clúster

Al analizar a qué bloque temático pertenece cada una de las preguntas que componen los Clústeres, se evidencia en la ilustración 13 una distribución muy definida para los clústeres 1 y 2, mientras que para el clúster 0, es más difusa debido a que se compone por varios bloques temáticos.



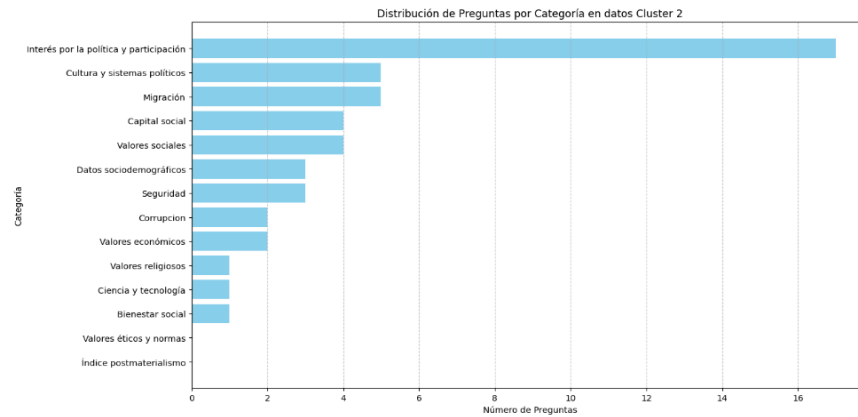


Ilustración 13. Composición de Clústeres por bloque temático

Al establecer una conexión entre los tres pilares descritos en el marco teórico y los hallazgos del clustering, es posible obtener ciertas aproximaciones útiles para la elección de preguntas durante el proceso de modelado. Los resultados indican que 17 preguntas correspondientes al bloque temático de valores sociales, seguido de 14 para cultura y sistemas políticos, seguido de 13 preguntas para seguridad, muestran una cierta inclinación a que la composición de este clúster sea más del tipo Regulatorio y Normativo. Con 22 preguntas correspondientes al bloque temático de capital social y 14 para valores éticos y normas, se observa una fuerte inclinación hacia la composición del pilar Normativo. Con 17 preguntas correspondientes al bloque temático de interés por la política y participación, 14 para cultura y sistemas políticos seguido por 13 correspondientes a migración, junto con la pregunta con mayor peso para el bloque correspondiente a valores religiosos, revela una mayor inclinación hacia el pilar Cognitivo.

Las preguntas que componen cada uno de los clústeres son un insumo valioso para contrastar la selección de preguntas que se realizarán en el modelado SEM y así permitir evaluar su influencia en los resultados de la modelación. En las tablas 8, 9 y 10, se relacionan la composición de preguntas para cada clúster.

Q5	Q6	Q17	Q22	Q25	Q27	Q28	Q29	Q34	Q35
Q37	Q38	Q39	Q40	Q41	Q42	Q45	Q46	Q49	Q51
Q52	Q53	Q54	Q55	Q64	Q66	Q67	Q78	Q121	Q122
Q123	Q125	Q127	Q132	Q133	Q134	Q136	Q137	Q138	Q140
Q142	Q143	Q144	Q145	Q146	Q151	Q152	Q154	Q163	Q165
Q166	Q167	Q168	Q169	Q170	Q171	Q172	Q173	Q174	Q175
Q182	Q183	Q184	Q185	Q186	Q188	Q195	Q197	Q198	Q235
Q236	Q237	Q239	Q243	Q246	Q249	Q250	Q251	Q252	Q254
Q255	Q256	Q258	Q263	Q264	Q269	Q275	Q288		

Tabla 8. Preguntas con mayor prevalencia clúster 0

Q1	Q2	Q12	Q18	Q32	Q44	Q47	Q56	Q58	Q59
Q60	Q65	Q68	Q70	Q71	Q73	Q74	Q75	Q77	Q79
Q80	Q81	Q90	Q94	Q95	Q96	Q97	Q98	Q99	Q100
Q101	Q102	Q103	Q104	Q105	Q108	Q109	Q110	Q111	Q113
Q114	Q115	Q116	Q117	Q118	Q119	Q131	Q153	Q155	Q156
Q157	Q160	Q161	Q162	Q176	Q177	Q178	Q179	Q180	Q181
Q187	Q189	Q190	Q191	Q192	Q196	Q202	Q221	Q222	Q224
Q232	Q233	Q234	Q238	Q245	Q253	Q257	Q271	Q273	

Tabla 9. Preguntas con mayor prevalencia clúster 1

Q3	Q4	Q16	Q43	Q48	Q57	Q61	Q62	Q63	Q106
Q107	Q112	Q120	Q124	Q126	Q128	Q129	Q130	Q135	Q139
Q150	Q158	Q164	Q199	Q200	Q201	Q203	Q204	Q205	Q207
Q208	Q209	Q210	Q211	Q212	Q213	Q214	Q215	Q216	Q217
Q241	Q242	Q244	Q247	Q248	Q279	Q286	Q287		

Tabla 10. Preguntas con mayor prevalencia clúster 2

#### 6.4. MODELACION SEM

Hasta el momento se ha estudiado la relación entre los valores sociales y la aceptación de sobornos como una conducta corrupta, utilizando datos de la Encuesta Mundial de Valores (WVS) (Haerpfer et al., 2022). Iniciamos con un planteamiento de hipótesis basado en el marco teórico de los tres pilares del institucionalismo: normativo, regulativo y cognitivo (Scott, 1995), los cuales proporcionan una comprensión profunda de los factores que influyen en la conducta corrupta. En la fase de análisis exploratorio de datos (EDA), se eliminaron variables con datos faltantes significativos, se realizó evaluación de correlaciones y la prueba Mann-Whitney U para seleccionar las preguntas que no fueron afectadas por la imputación de datos.

Posteriormente, se empleó la técnica de reducción de dimensionalidad UMAP para analizar las 215 preguntas con respuestas ordinales, enfocándose en aquellas que mostraban una clara relación con la variable de interés, Q118, relacionada con la aceptación de sobornos. Al no estandarizar los datos, se encontró la formación de tres clústeres distintos y se analizaron los resultados de clustering con su composición y preguntas con mayor prevalencia, lo cual validó la hipótesis inicial de que los tres pilares institucionales tienen una influencia directa en la aceptación de sobornos en un contexto de evaluación de valores sociales. Estos hallazgos proporcionaron una base sólida para avanzar con el estudio y seleccionar las preguntas más indicadas.

En esta etapa, se procede con la implementación del Modelos de Ecuaciones Estructurales (SEM, Structured Equations Model), una metodología robusta y adecuada para explorar relaciones complejas entre variables latentes y observadas. Estos modelos no generan un análisis causal, pero si generan constructos para medir relaciones entre variables. El modelado SEM permite modelar las interrelaciones entre múltiples variables dependientes e independientes simultáneamente, ofreciendo una representación gráfica clara y comprensible de estas relaciones (Hair Jr et al., 2021). Además, es especialmente útil en estudios donde se analizan constructos teóricos que no pueden ser medidos directamente, como los valores sociales, y su influencia sobre conductas específicas.

El modelado de ecuaciones estructurales (SEM) es una metodología estadística avanzada que permite examinar las relaciones complejas entre variables latentes y observadas. SEM combina aspectos de análisis factorial y modelos de regresión, proporcionando una representación gráfica de las relaciones entre múltiples variables. Su base teórica se sustenta en la capacidad de modelar constructos teóricos que no pueden ser medidos directamente, usando variables latentes representadas por indicadores observables (Kline, 2016).

SEM utiliza los siguientes componentes (Joos, 2021):

- **Variables Latentes y Observadas:** SEM trabaja con variables latentes (constructos teóricos no observables directamente) y variables observadas (indicadores medibles). Las variables latentes representan conceptos abstractos como actitudes, valores o rasgos, que se miden a través de múltiples variables observadas.
- **Modelos de Medición:** Evalúan cómo las variables latentes son indicadas por las variables observadas. El análisis factorial confirmatorio (CFA) se utiliza para validar que los indicadores medidos reflejan correctamente los constructos latentes.
- **Modelos Estructurales:** Examina las relaciones causales entre las variables latentes. Aquí se establecen caminos que representan las hipótesis de causalidad entre diferentes constructos. Esto permite modelar interacciones complejas y efectos directos e indirectos.
- **Modelos Mixtos:** Integran los modelos de medición y estructurales, proporcionando una evaluación completa del modelo teórico. Estos modelos permiten simultáneamente medir constructos latentes y probar las relaciones causales entre ellos.

El proceso para realizar una modelación SEM tiene los siguientes pasos principales (Joos, 2021):

- a. **Especificación del Modelo:** Se define un modelo teórico basado en la literatura y se especifican las relaciones entre variables latentes y observadas.
- b. **Identificación del Modelo:** Se asegura que el modelo es identificable, es decir, que hay suficientes datos para estimar los parámetros.
- c. **Estimación del Modelo:** Se utilizan métodos como la Máxima Verosimilitud (ML) para estimar los parámetros del modelo.
- d. **Evaluación del Modelo:** Se analizan los índices de ajuste y las cargas factoriales para determinar la validez del modelo.
- e. **Revisión y Mejora del Modelo:** Se ajusta el modelo basado en los resultados obtenidos, refinando las hipótesis y las relaciones propuestas.

#### 6.4.1. Especificación del modelo

Para especificar el modelo SEM de este estudio, se definen las relaciones entre variables latentes y observadas. Tomando como referencia la hipótesis teórica en la que se plantea que los tres pilares del institucionalismo (regulatorio, normativo y cognitivo) influyen significativamente en la aceptación de sobornos como conducta corrupta, se define que la pregunta Q118 es la variable medible que se usará. La pregunta Q118 hace parte del bloque temático de corrupción y se encuentra enunciada de la siguiente manera:

Q118. Queremos saber acerca de su experiencia con funcionarios locales y proveedores de servicios, como policías, abogados, médicos, maestros y funcionarios públicos en su comunidad. ¿Con qué frecuencia piensa que la gente común como Ud. o las personas de su barrio tienen que pagar un soborno, dar un regalo o hacer un favor a estas personas con el fin de obtener los servicios que necesita? ¿Ocurre nunca, raramente, frecuentemente o siempre? (Codificar sólo un número)

Nunca	Raramente	Frecuentemente	Siempre
1	2	3	4

Ilustración 14. Enunciado pregunta Q118

Ahora, la definición de variables latentes es:

- **Regulativo (R):** Representa leyes formales, autoridad exógena, explícita y escrita que regulan la conducta social.
- **Normativo (N):** Refleja normas y valores sociales compartidos, de la índole cultural, moral, costumbres compartidas y que son inercia social.
- **Cognitivo (C):** Incluye marcos de pensamiento menos conscientes y más interiorizados.
- **Aceptación de sobornos (S):** Variable latente dependiente de los pilares Regulativo, Normativo, Cognitivo, otras variables latentes, y de la variable observada Q118.

Definición de variable observada:

- **Corrupción (Q118):** Indicador de la aceptación de sobornos

Para nuestro estudio, la corrupción está representada, según la hipótesis, por una variable observada que es la Q118, la cual tiene valores ordinales del 1 al 4. Adicionalmente, la hipótesis establece que la corrupción está también influenciada por variables latentes, Regulativas, Normativas, Cognitivas, que no son medibles directamente, pero cada una de ellas si pueden expresarse por variables observadas de las 215 preguntas que se encuentran disponibles para analizar. A continuación, se enumeran las preguntas seleccionadas bajo el criterio del marco teórico establecido en el estudio:

<b>Pregunta</b>	<b>Enunciado</b>	<b>Bloque temático</b>
Q45	Futuros cambios: Que haya un mayor respeto por la autoridad	Valores sociales, actitudes y estereotipos
Q70	Confianza: Las cortes de justicia	Capital social, confianza y pertenencia a asociaciones
Q71	Confianza: El gobierno	
Q74	Confianza: Los funcionarios públicos	
Q132	Frecuencia de hechos en su barrio: Robos	Seguridad
Q134	Frecuencia de hechos en su barrio: Interferencia de la policía o los militares en la vida privada de la gente	

*Tabla 11. Preguntas seleccionadas para el pilar Regulativo*

<b>Normativas</b>		
Q40	El trabajo es un deber que se tiene para con la sociedad	Valores sociales, actitudes y estereotipos
Q177	Justificable: Reclamar beneficios del Estado a los que no se tiene derecho	Valores éticos y normas
Q178	Justificable: Evitar pagar el pasaje en algún transporte público	
Q179	Justificable: Robar	
Q180	Justificable: Engañar en el pago de impuestos, si se puede	
Q181	Justificable: Que alguien acepte un soborno en el cumplimiento de sus funciones	
Q195	Justificable: La pena de muerte	

*Tabla 12. Preguntas seleccionadas para el pilar Normativo*

<b>Pregunta</b>	<b>Enunciado</b>	<b>Bloque temático</b>
Q2	Importancia en la vida: Amigos	Valores sociales, actitudes y estereotipos
Q12	Importancia en educar a los hijos: Tolerancia y respeto hacia otros	
Q16	Importancia en educar a los hijos: Generosidad, altruismo	
Q17	Importancia en educar a los hijos: Obediencia	
Q164	Importancia de Dios en su vida	Valores religiosos
Q165	Creencia en: Dios	
Q166	Creencia en: Vida después de la muerte	
Q167	Creencia en: El infierno	
Q168	Creencia en: El cielo	
Q173	Es una persona religiosa	
Q174	Significado de la religión: Cumplir normas y ceremonias VS Hacer el bien a la gente	
Q253	Respeto a los derechos humanos hoy en día	Cultura y sistemas políticos

*Tabla 13. Preguntas seleccionadas para el pilar Cognitivo*

Por lo tanto, tenemos la definición de variables latentes expresadas en variables observadas de la siguiente manera:

**Regulativo (R):** Q45 + Q70 + Q71 + Q74 + Q132 + Q134

**Normativo (N):** Q40 + Q177 + Q178 + Q179 + Q180 + Q181 + Q195

**Cognitivo (C):** Q2 + Q12 + Q16 + Q17 + Q164 + Q165 + Q166 + Q167 + Q168 + Q173 + Q174 + Q253

#### **6.4.2. Identificación del modelo**

Existe un mecanismo sencillo para diagramar las relaciones entre variables latentes y observadas, sus relaciones planteadas en las hipótesis, y las variaciones y residuales generados. Las variables latentes se indican con un círculo. Las variables medidas se indican con un cuadrado. Las relaciones se indican con flechas. Las varianzas y los residuos se indican mediante flechas desde una variable hacia sí misma (Joos, 2021). A continuación, la ilustración 15 presenta el modelo SEM propuesto.

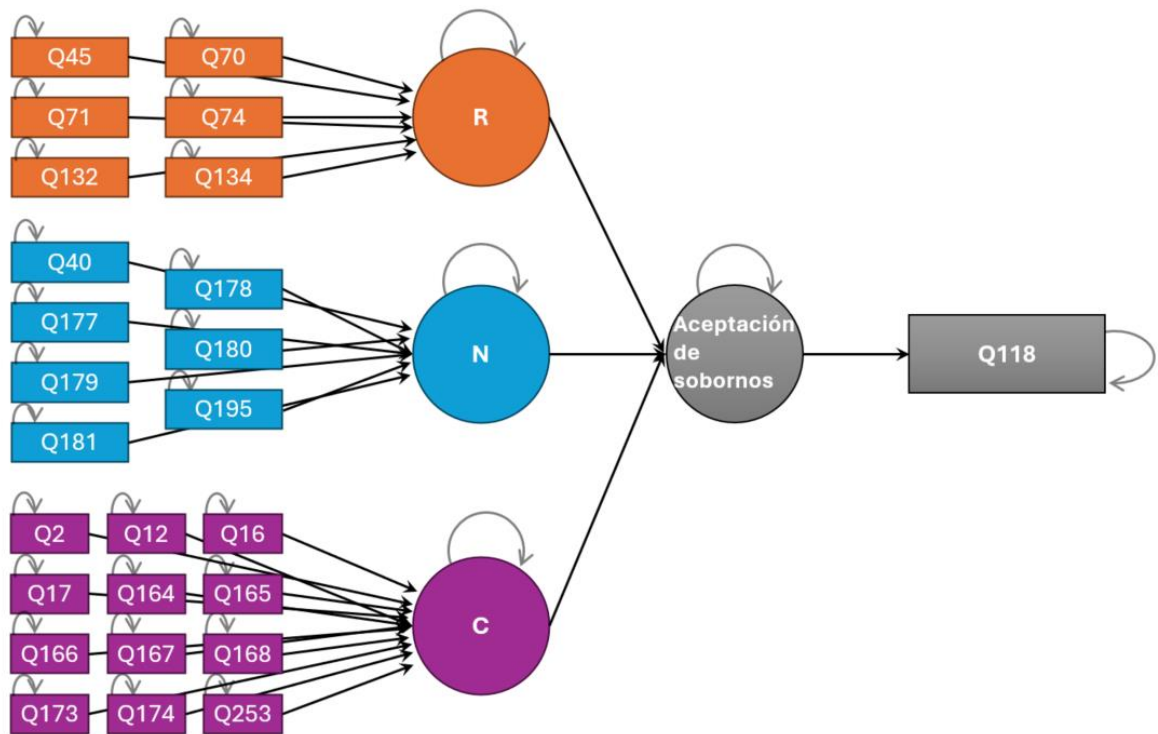


Ilustración 15. Modelo SEM propuesto

La notación gráfica es un estándar (Joos, 2021) que sigue los siguientes lineamientos:

- Variables latentes son representadas por un círculo
- Variables observables son representadas por un cuadrado
- Relaciones planteadas son representadas por flechas
- Varianzas y residuales son representadas por flechas bucles de variables a sí mismas

#### 6.4.3. Estimación y evaluación del modelo

Se realiza la estimación y evaluación del Modelo de Ecuaciones Estructurales (SEM) utilizando el paquete de Python **semopy**. Este paquete permite especificar el modelo teórico, ajustar los datos empíricos y evaluar la calidad del ajuste del modelo mediante diversos índices de ajuste. A través de **semopy**, se puede cuantificar las relaciones complejas entre las variables latentes y observadas, validando la hipótesis sobre la influencia de los pilares regulativos, normativos y cognitivos en la aceptación de sobornos como una conducta corrupta.

La relación entre la variable dependiente **Corrupción** y las variables latentes **Regulativa (R)**, **Normativa (N)** y **Cognitiva (C)** se puede expresar con el planteamiento de hipótesis nula e hipótesis alternativa de la siguiente manera:

#### Hipótesis Nula ( $H_0$ )

No hay una relación significativa entre las variables latentes (Regulativa (R), Normativa (N) o Cognitiva (C)) y la variable dependiente Corrupción.

### Hipótesis Alternativa (H<sub>1</sub>)

Existe una relación significativa entre las variables latentes (Regulativa (R), Normativa (N) o Cognitiva (C)) y la variable dependiente Corrupción.

A continuación, se presentan los resultados del modelo:

<b>lval</b>	<b>op</b>	<b>rval</b>	<b>Estimate</b>	<b>Std. Err</b>	<b>z-value</b>	<b>p-value</b>
Corruption	~	Regulative	8.433182	1.047753	8.048824	0.0
Corruption	~	Normative	2.157768	0.206583	10.445056	0.0
Corruption	~	Cognitive	0.746129	0.107149	6.963467	0.0

*Tabla 14. Resultados estadísticos del modelo SEM*

Se establece un nivel de significancia o alfa de 5%. Para la relación entre **Corrupción** y **Regulativa (R)**, se observa un p-value de 0.0, lo cual es menor que el p-value de 0.05, Por lo cual, se rechaza la hipótesis nula (H<sub>0</sub>) y se observa que hay evidencia suficiente para concluir que los factores regulativos (leyes formales y autoridad exógena) tienen un impacto notable en la aceptación de sobornos. Para la relación entre **Corrupción** y **Normativa (N)**, se observa un p-value de 0.0, lo cual es menor que el p-value de 0.05, Por lo cual, se rechaza la hipótesis nula (H<sub>0</sub>) y se observa evidencia suficiente para concluir que los factores normativos (Cultura, moral, costumbres, inercia social) tienen un impacto notable en la aceptación de sobornos.

Para la relación entre **Corrupción** y **Cognitivo (C)**, se observa un p-value de 0.0, lo cual es menor que el p-value de 0.05, Por lo cual, se rechaza la hipótesis nula (H<sub>0</sub>) y se observa que hay evidencia suficiente para concluir que los factores cognitivos (marcos de pensamiento menos conscientes y más interiorizados) tienen un impacto notable en la aceptación de sobornos. En conclusión, el modelo construido con las variables observadas seleccionadas confirma la significancia e influencia de los pilares institucionales en la aceptación de sobornos como una conducta de corrupción. En especial, los coeficientes estimados indican un mayor peso o influencia sobre la variable dependiente "**Corrupción**", en donde el pilar **Regulativo** tiene una mayor influencia con un 8.4331, seguido del pilar **Normativo** con un 2.1577 y finalmente por el pilar **Cognitivo** con un 0.1071.

A continuación, se observan los resultados de las métricas resultantes para el modelo:

**Chi-square:** 123904.1798 **DoF (df):** 293 **RMSEA:** 0.0668 **p-value:** 0.0

A pesar de un Chi-square elevado y un p-value de 0.0, lo cual es esperado dado el gran tamaño de la muestra, el RMSEA de 0.0669 sugiere que el modelo SEM posee

un ajuste razonable a los datos, capturando de manera efectiva las relaciones esenciales sin sobreajuste.

#### 6.4.4. Revisión y mejoras del modelo

Como paso final se realiza la evaluación de las cargas factoriales. Las cargas factoriales son indicadores de la relación entre las variables observables y las variables latentes, y su análisis permite determinar la validez y la fiabilidad del modelo (Kenny, 2015). Específicamente, una carga factorial alta indica que la variable observable tiene una fuerte correlación con la variable latente, sugiriendo que es un buen indicador de dicha variable latente. Por el contrario, cargas factoriales bajas pueden señalar variables observables que no contribuyen significativamente a la variable latente y que podrían ser candidatas para ser revisadas o eliminadas. Este análisis es fundamental para asegurar que cada constructo del modelo esté adecuadamente representado por sus variables observables, mejorando así la precisión y la interpretabilidad del modelo SEM en el estudio de la relación entre los valores sociales y la aceptación de sobornos como una forma de corrupción. Las ilustraciones 16, 17 y 18, presentan los resultados encontrados:

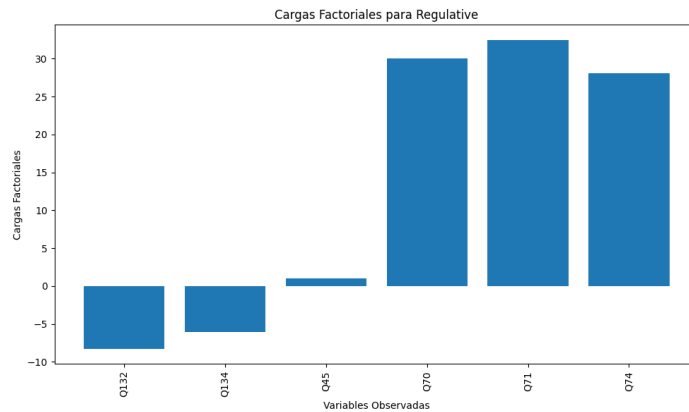


Ilustración 16. Cargas factoriales para preguntas seleccionadas pilar Regulative

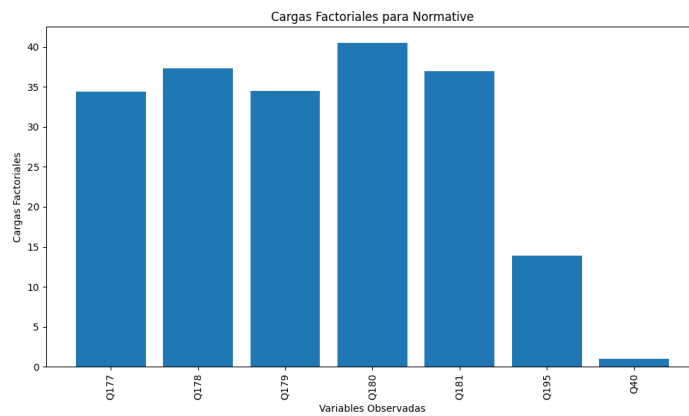


Ilustración 17. Cargas factoriales para preguntas seleccionadas pilar Normativo

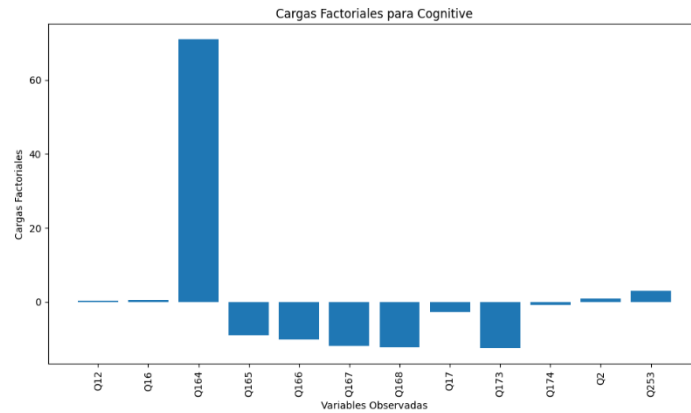


Ilustración 18. Cargas factoriales para preguntas seleccionadas pilar Cognitivo

Se observa principalmente que las preguntas seleccionadas en el modelo SEM para el pilar Cognitivo tienen cargas factoriales negativas, lo que indica que están inversamente relacionadas, estas variables disminuyen el valor del factor Cognitivo cuando sus valores aumentan. En contraste, la variable Q164 es la pregunta más influyente en el pilar Cognitivo, contribuyendo positivamente. Para los otros pilares se observa una influencia positiva en todas las preguntas del pilar Normativo y de 4 de 6 preguntas para el pilar Regulatorio.

Finalmente, para evaluar los resultados del clustering y los resultados del modelo SEM se pueden encontrar aproximaciones importantes. Se resaltan en las ilustraciones 19, 20 y 21 que las preguntas seleccionadas para el modelo SEM en el pilar Regulatorio (R) se encuentran distribuidas en igual proporción entre los clústeres 0 y 1, para el pilar Normativo (N) se encuentran en mayor proporción en el clúster 1 y en menor proporción en el clúster 0 y para el pilar Cognitivo (C) se encuentran distribuidas en los tres clústeres, con mayor proporción en el clúster 0.

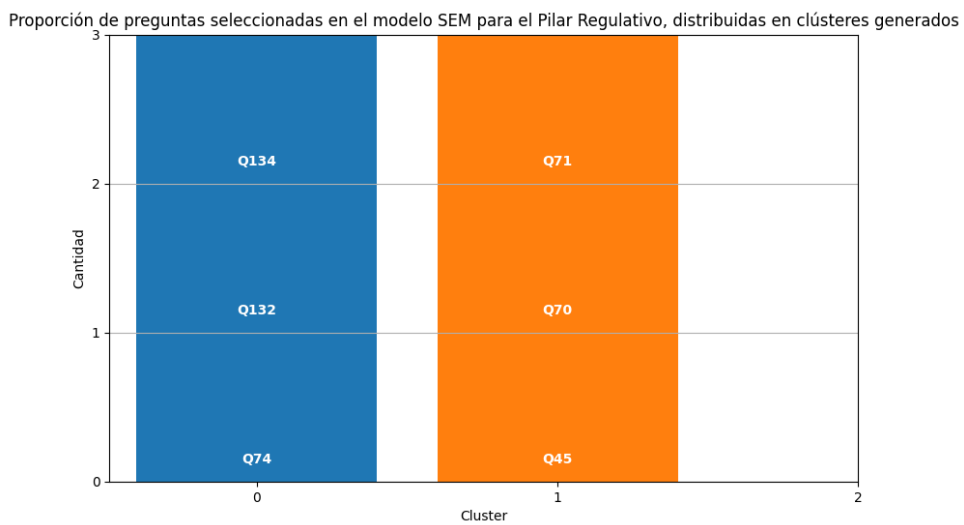


Ilustración 19. Proporción de preguntas seleccionadas en el modelo SEM para el Pilar Regulatorio, distribuidas en clústeres generados

Proporción de preguntas seleccionadas en el modelo SEM para el Pilar Normativo, distribuidas en clústeres generados

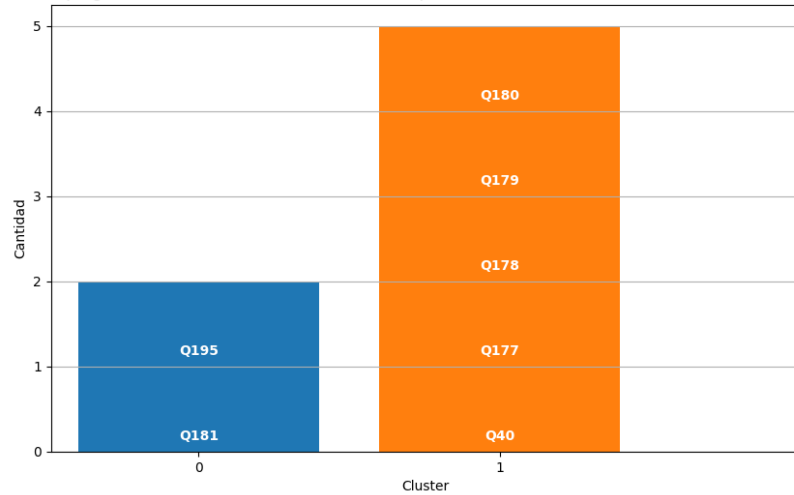


Ilustración 20. Proporción de preguntas seleccionadas en el modelo SEM para el Pilar Normativo, distribuidas en clústeres generados

Proporción de preguntas seleccionadas en el modelo SEM para el Pilar Cognitivo, distribuidas en clústeres generados

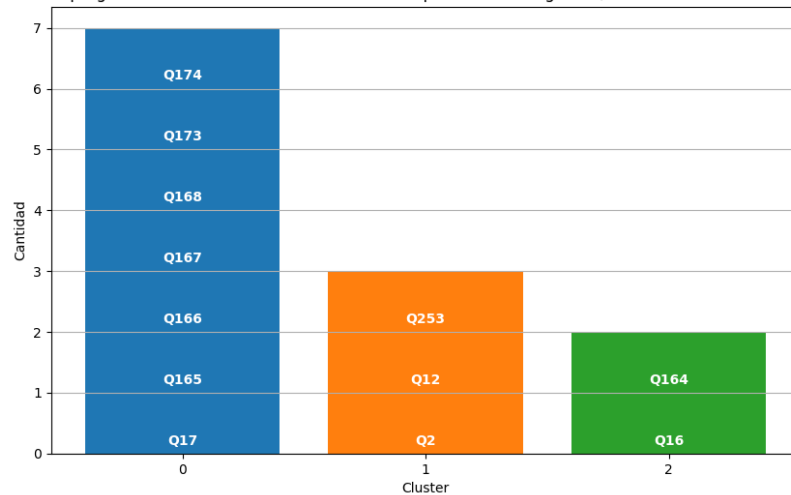


Ilustración 21. Proporción de preguntas seleccionadas en el modelo SEM para el Pilar Cognitivo, distribuidas en clústeres generados

Estos resultados nos revelan que las preguntas seleccionadas para el modelo SEM bajo la mirada experta y el criterio del marco teórico se encuentran cercanas a los resultados generados por la reducción de dimensionalidad UMAP y la clusterización implementada conformadas a partir de los patrones subyacentes.

## 6.5. DESPLIEGUE

El despliegue del estudio se realizó mediante una composición de herramientas avanzadas y prácticas recomendadas en el campo de la ciencia de datos y analítica. Los datos fueron descargados del sitio dispuesto por la Encuesta Mundial de

Valores (WVS) (Haerpfer et al., 2022). El código fuente fue escrito en Python, aprovechando su extensa biblioteca de herramientas para análisis de datos y Machine Learning. Los notebooks interactivos fueron desarrollados en Google Colab, permitiendo el uso de recursos computacionales en la nube. Todo el código fuente, junto con los gráficos de resultados del análisis exploratorio de datos (EDA) y procesamiento, fue almacenado y gestionado en un repositorio de GitHub.

Para la implementación, se utilizaron paquetes especializados como SEM (Structural Equation Modeling) para el modelado de ecuaciones estructurales, UMAP (Uniform Manifold Approximation and Projection) para la reducción de dimensionalidad, K-Means para la agrupación de datos, y Seaborn para la visualización avanzada de datos. Además, se hizo uso extensivo de Scikit-learn para tareas de Machine Learning y análisis estadístico. De esta manera se garantiza la trazabilidad y reproducibilidad del estudio, facilitando el acceso a la fuente de datos, código, y resultados visuales, con los cuales se dispone para futuras colaboraciones.

El acceso al repositorio se encuentra disponible en el siguiente enlace:

- [https://github.com/giovy79/MCDA\\_TesisGrado](https://github.com/giovy79/MCDA_TesisGrado)

## 7. CONCLUSIONES

La Encuesta Mundial de Valores (WVS) ofrece una fuente rica en datos que abarca múltiples dimensiones de las actitudes y valores sociales en todo el mundo. Los bloques temáticos de la encuesta incluyen aspectos como valores económicos, religiosos, éticos y sociales, percepciones sobre la participación política, seguridad y migración. Estos temas proporcionan un marco general para analizar los valores y creencias de diferentes culturas y en diversos contextos sociales.

Para este estudio ha sido relevante y de gran utilidad contar con la percepción de corrupción para explorar la relación entre los valores sociales y la aceptación de sobornos como una conducta corrupta. Los pilares del institucionalismo propuestos por Scott (Scott, 1995) sirven como un marco teórico sólido para interpretar distintas formas mediante las cuales se puede observar el fenómeno de la corrupción en algunas dimensiones, como por ejemplo, en las estructuras sociales que, vistas desde el pilar **regulativo**, pueden conformar mecanismos legales de validación de conductas corruptas, desde el pilar **normativo** se puede generar una 'normalidad' el hecho de aceptar sobornos para obtener un servicio y desde el pilar **cognitivo** se puede constituir una práctica social aceptada sin cuestionamientos, y patrón mental en el cual es válido aceptar sobornos; aclarando que la validez es resultado de la costumbre y percepción normalizada, y no el resultado de un discernimiento y escrutinio racional. Las instituciones y normas sociales no solo existen en la realidad externa, también son construidas y mantenidas en las mentes de las personas, quienes en su entorno interactúan y refuerzan estos conceptos y estructuras.

En el contexto de los valores sociales, establecer una causalidad directa, es decir, una conexión inequívoca, entre diversas variables sociales y la aceptación de sobornos no podría establecerse sin ambigüedades, debido a la multiplicidad de factores, variables y componentes que se relacionan en distintos horizontes temporales, y muchas veces de maneras poco visibles. Esta complejidad de factores diversos, externos e internos, porosos e incrustados (*embedded*) en dinámicas de interacción cognitivas, normativas y regulativas, solo podrían aislarse de forma artificial.

Para la elaboración de este estudio se han identificado asociaciones significativas entre algunos valores sociales y la aceptación de sobornos, aclarando que esto no lo interpretamos necesariamente como indicadores de causalidades directas. Es decir, es posible que estas asociaciones estén influenciadas simultáneamente por otras variables no consideradas, tales como ciclos electorales, implementación de leyes locales, configuraciones de acuerdos internacionales, cambios demográficos, etc. Conviene precisar que el WVS se realiza simultáneamente en los países auscultados, sin que se estén ajustando los tiempos de las encuestas a las vicisitudes locales.

Desde la perspectiva de este estudio, se ha abordado una comprensión holística para explorar aquellos datos relevantes de la Encuesta Mundial de Valores (WVS),

de tal forma que al examinar posibles agrupaciones y relaciones con la aceptación de sobornos como una conducta corrupta. De todas formas, esta aproximación preliminar no pretende ofrecer interpretaciones definitivas; pero si servir como base. En todos los casos, se requieren estudios adicionales para establecer causales definitivas y comprender en detalle mecanismos y factores involucrados; en los que además convendría considerar las especificidades de contextos socio-políticos, económicos y electorales locales. Por ejemplo, se identifica con claridad lo determinante que es la pregunta Q164 en el pilar cognitivo y su ubicación en el clúster 2, pero, no es muy clara la influencia de las preguntas seleccionadas en el modelo SEM para el pilar regulativo y su ubicación en los clústeres 1 o 0; podrán ubicarse en uno u otro.

En entornos o sociedades en las cuales se normaliza o justifica la aceptación de sobornos, la corrupción tiene una mayor influencia y genera un mayor deterioro en la confianza de las personas. Estas prácticas deshonestas se pueden perpetuar y lograr violar el principio de “mantener cierta distancia”, la cual, por ejemplo, desde una mirada en la gestión pública, se refiere al principio según el cual las decisiones económicas y administrativas deben tomarse de manera imparcial, sin la influencia de relaciones personales o de parentesco y menos aún motivadas por sobornos o prebendas. Para una sociedad que no quiera colapsar, resulta evidente que no se debe permitir la normalización de la corrupción pues su impacto afecta todos los campos de acción de una sociedad.

Desde la interpretación de los hallazgos en este estudio, se confirma que fortalecer los marcos legales y regulatorios que promueven la transparencia, la rendición de cuentas y el establecimiento de códigos de conducta claros y políticas de “cero tolerancia” hacia la aceptación de sobornos, es una acción que tiene influencia en la reducción de la corrupción pero que aún así no resulta ser la más determinante.

El entorno cultural y las costumbres representan una mayor influencia y es necesario contar con una promoción de una cultura ética en la cual existan programas de educación y conciencia para fomentar valores éticos y la capacidad de razonamiento moral (más allá de una inercia cultural), desde una edad temprana realzando la importancia de la integridad y la honestidad, así mismo, crear mecanismos de reconocimiento y recompensa en donde se premien y exalten estos comportamientos y recalando y promoviendo la tolerancia cero hacia la corrupción, esto genera un gran impacto en el entorno.

Finalmente, se evidencia como la mayor influencia se realiza desde el sistema de identidades y creencias de cada individuo el cual debe ser desafiado constantemente a través de campañas de comunicación que cambien las percepciones frente a la ‘normalización’ de la corrupción. Esta base fundamental, que suele pasar desapercibida puede ser la más significativa en el corto plazo.

En el presente estudio se ha logrado cumplir con el objetivo general de desarrollar un modelo de **aprendizaje de máquina no supervisado** utilizando modelos de ecuaciones estructurales (SEM) y técnicas de reducción de dimensionalidad para

entender la relación entre los valores sociales y la propensión a la aceptación de sobornos como una forma de corrupción, basándose en los datos de la Ola 7 (2017-2022) de la Encuesta Mundial de Valores (WVS). Se ha explicado el contexto del problema de la corrupción, apoyado desde el marco teórico del institucionalismo expuesto, mediante un análisis exploratorio descriptivo, seguido de la preparación de los datos utilizando técnicas de limpieza, transformación y reducción de dimensionalidad UMAP y *clustering*. Además, se implementaron y ajustaron modelos de aprendizaje de máquina no supervisados de modelos SEM, evaluando y contrastando los resultados de la composición de clústeres con las cargas factoriales de cada pregunta seleccionada.

Finalmente, es importante evaluar las implicaciones de los hallazgos desde diferentes puntos de vista. Desde lo teórico, la interdisciplinariedad del enfoque adoptado ha permitido una comprensión más completa y diversa del problema de la corrupción. Al abordar el estudio sin sesgos sociales, culturales o políticos, y con una mirada técnica propia de un ingeniero de sistemas y estudiante de la maestría de ciencia de datos y analítica, se ha procurado garantizar una imparcialidad que fortalece la validez de los resultados obtenidos; a la vez que se ofrecen resultados para enriquecer el debate social, mediante los ejes solidez y conversaciones interdisciplinarias que permiten los pilares del institucionalismo, un marco teórico incluyente y compatibles con la interpretación de entornos VUCA (volátiles, inciertos, complejos y ambiguos), dónde suelen desplegarse los intercambios sociales en que se gestan prácticas corruptas, y simultáneamente, los mecanismos sociales de control y prevención.

Desde lo metodológico, las ciencias de los datos no son solo “análisis de datos” o tecnologías “computacionales” modernas aplicadas a la resolución de problemas reales, las ciencias de datos constituyen un nuevo campo interdisciplinario que usa un número relevante de disciplinas de conocimiento tales como la estadística, computación, informática, comunicación, administración, sociología con el fin de proveer soluciones sistemáticas y holísticas a partir del estudio de los datos correspondientes al dominio del conocimiento en el cual será aplicado, siguiendo un pensamiento y método científico.

Durante la fase de preparación y modelación se observó como la selección de una sola variable tuvo implicaciones directas sobre los resultados estadísticos e interpretaciones realizadas. Esto destaca la importancia de un enfoque meticuloso y exhaustivo en la selección y análisis de variables, asegurando que los modelos resultantes sean robustos y las conclusiones sean precisas y significativas.

Desde lo empírico, es posible neutralizar la corrupción implementando sistemas de monitoreo con algoritmos de *Machine Learning* para detectar anomalías en tiempo real, cuando se infrinjan los marcos regulatorios y legales, realizar campañas de concienciación y programas educativos, que cambien las percepciones culturales y normativas sobre la corrupción, Además, es crucial desarrollar campañas de comunicación que permitan cuestionar, criticar y eventualmente cambiar las

percepciones individuales y las prácticas colectivas frente a la ‘normalización’ de la corrupción.

Las limitaciones y dificultades vistas durante este estudio han sido diversas. Desde el punto de vista estadístico los datos faltantes, la diversidad de tipos de preguntas y respuestas, las correlaciones entre preguntas y la evaluación de los procesos de imputación han impuesto rigurosidad en el tratamiento de los datos y en pensar la necesidad de contrastar con más fuentes de información. Adicionalmente, la discrepancia de valores entre diferentes culturas, idiomas y significados regionales dificulta la interpretación. También existe el riesgo de que las encuestas puedan ser según sesgos de lo que se perciba como lo políticamente correcto en lugar de expresar sus verdaderos pensamientos y sentimientos. Esta tendencia a dar respuestas socialmente aceptables distorsiona los resultados y limita la capacidad de encontrar su origen. Sin embargo, si suponemos que todos los encuestados tienen el mismo sesgo, se legitiman las comparaciones. A pesar de todas las limitaciones, encontramos que la WVS ofrece los datos más completos y legítimos para explorar de manera global y comparativa los valores a nivel mundial.

Este estudio ha permitido plantearse futuras preguntas de investigación que merecen ser exploradas en profundidad en diversas áreas y variables del comportamiento humanos, tales como las culturales, colectivas e individuales; examinando como cada dimensión y nivel de recursividad específica puede influir de forma particular, y en combinaciones con las demás, en términos de la aceptación prácticas corruptas, tales como las dinámicas, frecuencias y magnitudes de sobornos en diferentes regiones. Más importante aún, sería deseable profundizar investigaciones respecto a cómo mitigar, prevenir y compensar manifestaciones corruptas indeseables, mediate acciones e intervenciones sustentadas para movilizar personas y grupos humanos, por ejemplo, apalancándonos en las comprensiones profundas que aportan los pilares del institucionalismo.

En consecuencia, resulta imprescindible investigar el impacto de la educación y las campañas de concientización sobre la reducción de la corrupción, así como la creación de marcos teóricos integrados que combinen perspectivas sociológicas y tecnológicas para comprender mejor el fenómeno de la corrupción y lograr reducirla.

## 8. REFERENCIAS

- 2023 Corruption Perceptions Index: Explore the...* - Transparency.org. (n.d.). Retrieved August 6, 2024, from <https://www.transparency.org/en/cpi/2023>
- Achim, M. V. (2016). Cultural Dimension of Corruption: A Cross-Country Survey. *International Advances in Economic Research*, 22(3), 333–345. <https://doi.org/10.1007/s11294-016-9592-x>
- Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25(5), 464–469. <https://doi.org/10.1111/J.1467-842X.2001.TB00294.X>
- Berger, P. L., & Luckmann, T. (2011). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Open Road Media. <https://books.google.com.co/books?id=Jcma84waN3AC>
- Bourdieu, P., & Nice, R. (1977). Outline of a Theory of Practice. In *Cambridge Studies in Social and Cultural Anthropology*. Cambridge University Press. <https://doi.org/DOI: 10.1017/CBO9780511812507>
- Consequences of Corruption at the Sector Level and Implications for Economic Growth and Development. (2015). *Consequences of Corruption at the Sector Level and Implications for Economic Growth and Development*. <https://doi.org/10.1787/9789264230781-EN>
- Davidescu, A., Manta, E., Stoica-Ungureanu, A., & (Musat), M. (2022). Could Religiosity and Religion Influence the Tax Morale of Individuals? An Empirical Analysis Based on Variable Selection Methods. *Mathematics*, 10, 4497. <https://doi.org/10.3390/math10234497>
- Donald B. Rubin, & Roderick J. A. Little. (2014). *Statistical Analysis with Missing Data* (John Wiley & Sons, Ed.; 2nd ed.).

- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.  
<http://www.sup.org/books/title/?id=3850>
- Fisher, R. A. (1922). The Goodness of Fit of Regression Formulæ, and the Distribution of Regression Coefficients. *Journal of the Royal Statistical Society*, 85(4), 597–612. <https://doi.org/10.1111/j.2397-2335.1922.tb00832.x>
- Gelman, A., Carlin, J. B. B., Stern, H. S. S., & Rubin, D. B. B. (2014). Bayesian Data Analysis. In *Book*. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Haerper, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., & Ponarin, E. & P. B. (2022). *WVS Database*. World Values Survey Wave 7 (2017-2022) Cross-National Data-Set. Version: 4.0.0. World Values Survey Association. <https://doi.org/doi.org/10.14281/18241.18>
- Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295X>
- Hair, J. F. , B. W. C. , B. B. J. and A. R. E. (2010). *Multivariate Data Analysis*. 761.
- Hair Jr, J. F., M. Hult, G. T., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray Soumya. (2021). *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R* (Springer, Ed.; Open Access). <https://doi.org/https://doi.org/10.1007/978-3-030-80519-7>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. In *The Elements of Statistical Learning* (Vol. 27, Issue 2). <https://doi.org/10.1007/b94608>
- Hoffman, A. J. (2001). *From Heresy to Dogma*. 287. [https://books.google.com/books/about/From\\_Heresy\\_to\\_Dogma.html?hl=el&id=xFq-uHIGhtwC](https://books.google.com/books/about/From_Heresy_to_Dogma.html?hl=el&id=xFq-uHIGhtwC)

- Hofstede, G. (2001). Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. In *Behaviour Research and Therapy - BEHAV RES THER* (Vol. 41). [https://doi.org/10.1016/S0005-7967\(02\)00184-5](https://doi.org/10.1016/S0005-7967(02)00184-5)
- Hood, C. (1990). J. G. March and J. P. Olsen Rediscovering Institutions: The Organizational Basis of Politics, New York, Free Press, 1989. *Journal of Public Policy*, 10(3), 349–351. <https://doi.org/DOI: 10.1017/S0143814X00005869>
- Horkheimer, M., Adorno, T. W., & Schmid Noerr, G. (n.d.). *Dialectic of enlightenment : philosophical fragments*.
- Hume, D. (1777). *An Enquiry Concerning the Principles of Morals*. <https://www.gutenberg.org/files/4320/4320-h/4320-h.htm>
- IBM. (2021). *Conceptos básicos de ayuda de CRISP-DM - Documentación de IBM*. <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning*. Springer New York, NY. <https://doi.org/https://doi.org/10.1007/978-1-0716-1418-1>
- Johnston, M. (1986). The Political Consequences of Corruption: A Reassessment. *Comparative Politics*, 18(4), 459–477. <https://doi.org/10.2307/421694>
- Jones, F. E. (1959). Men and Their Work. The Free Press. 1958. Pp. 184. *Canadian Journal of Economics and Political Science*, 25(4), 523–524. <https://doi.org/DOI: 10.2307/139003>
- Joos, K. (2021, July). *Structural Equation Modeling | Towards Data Science*. <https://towardsdatascience.com/structural-equation-modeling-dca298798f4d>
- Kenny, D. A. (2015). *Confirmatory Factor Analysis for Applied Research Methodology in the Social Sciences*. [www.guilford.com/MSS](http://www.guilford.com/MSS)

- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (T. D. Little, Ed.; Fourth Edition). The Guilford Press. [www.guilford.com/kline](http://www.guilford.com/kline)
- Klitgaard, R. E. (1988). *Controlling corruption*. 220. [https://books.google.com/books/about/Controlling\\_Corruption.html?hl=es&id=L7YwDwAAQBAJ](https://books.google.com/books/about/Controlling_Corruption.html?hl=es&id=L7YwDwAAQBAJ)
- Li, S. (2019). *Bribery and Corruption in Weak Institutional Environments: Connecting the Dots from a Comparative Perspective*. Cambridge University Press. [https://doi.org/DOI: 10.1017/9781108631440](https://doi.org/DOI:10.1017/9781108631440)
- Lima, M. S. M., & Delen, D. (2020). Predicting and explaining corruption across countries: A machine learning approach. *Government Information Quarterly*, 37(1), 101407. <https://doi.org/https://doi.org/10.1016/j.giq.2019.101407>
- Little, R. J. A., & Rubin, D. B. (2002). Single Imputation Methods. In *Statistical Analysis with Missing Data* (pp. 59–74). <https://doi.org/https://doi.org/10.1002/9781119013563.ch4>
- Mauro, P. (1995). Corruption and Growth\*. *The Quarterly Journal of Economics*, 110(3), 681–712. <https://doi.org/10.2307/2946696>
- McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*.
- McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. *The Corsini Encyclopedia of Psychology*, 1.
- Medina y Navarro, E. (1873). *Aristóteles, Política*. <https://www.filosofia.org/bol/bib/nb028.htm>
- Montesquieu, C. (1748). *Del espíritu de las leyes. Volumen. 1*. <https://archive.org/details/del-espiritu-de-las-leyes-volumen-1-montesquieu>
- Mungiu-Pippidi, A. (2015). *The Quest for Good Governance: How Societies Develop Control of Corruption*. <https://doi.org/10.1017/CBO9781316286937>

- NEYMAN, J., & PEARSON, E. S. (1928). ON THE USE AND INTERPRETATION OF CERTAIN TEST CRITERIA FOR PURPOSES OF STATISTICAL INFERENCE PART I. *Biometrika*, 20A(1–2), 175–240. <https://doi.org/10.1093/biomet/20A.1-2.175>
- North, D. C. (1990). Institutions, Institutional Change and Economic Performance (pp. 1-2). In *Political Economy of Institutions and Decisions*. Cambridge University Press. [https://doi.org/DOI: 10.1017/CBO9780511808678](https://doi.org/DOI:10.1017/CBO9780511808678)
- Nye, J. S. (1967). Corruption and Political Development: A Cost-Benefit Analysis\*. *American Political Science Review*, 61(2), 417–427. [https://EconPapers.repec.org/RePEc:cup:apsrev:v:61:y:1967:i:02:p:417-427\\_13](https://EconPapers.repec.org/RePEc:cup:apsrev:v:61:y:1967:i:02:p:417-427_13)
- Pope, J. (2000). Confronting corruption: The elements of a national integrity system. In *TI source book 2000*. Transparency International.
- Rose-Ackerman, S., & Palifka, B. J. (2016). *Corruption and Government Causes, Consequences, and Reform* (2nd Edition). Cambridge University Press.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Salcedo Albarán, Eduardo. (2007). *Corrupción, cerebro y sentimientos: una indagación neuropsicológica en torno a la corrupción*. Método.
- Scott, W. Richard. (1995). *Institutions and organizations* (SAGE Publications, Ed.; ilustrada, reimpressa). [https://books.google.com/books/about/Institutions\\_and\\_Organizations.html?hl=es&id=BfMSSgAACAAJ](https://books.google.com/books/about/Institutions_and_Organizations.html?hl=es&id=BfMSSgAACAAJ)
- Sobol, I., & Kucherenko, S. (2005). Global sensitivity indices for nonlinear mathematical models. Review. *Wilmott*, 2005, 56–61. <https://doi.org/10.1002/wilm.42820050114>

- Tenbrunsel, A. E., & Smith-Crowe, K. (2008). Ethical Decision Making: Where We've Been and Where We're Going. *Https://Doi.Org/10.5465/19416520802211677*, 2(1), 545–607. <https://doi.org/10.5465/19416520802211677>
- Tomás de Aquino, S. (2005). *Suma teológica*. [https://archive.org/details/4.-santo-tomas-de-aquino-suma-teologica-vol.-iv-ed.-loyola\\_202301](https://archive.org/details/4.-santo-tomas-de-aquino-suma-teologica-vol.-iv-ed.-loyola_202301)
- Tukey, J. W. (1977). *Exploratory data analysis*. [https://archive.org/details/exploratorydataa0000tuke\\_7616/page/n9/mode/2up](https://archive.org/details/exploratorydataa0000tuke_7616/page/n9/mode/2up)
- UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction — umap 0.5 documentation*. (n.d.). Retrieved August 6, 2024, from <https://umap-learn.readthedocs.io/en/latest/>
- WVS. (2024). *WVS What we do*. <https://www.worldvaluessurvey.org/WVSContents.jsp>

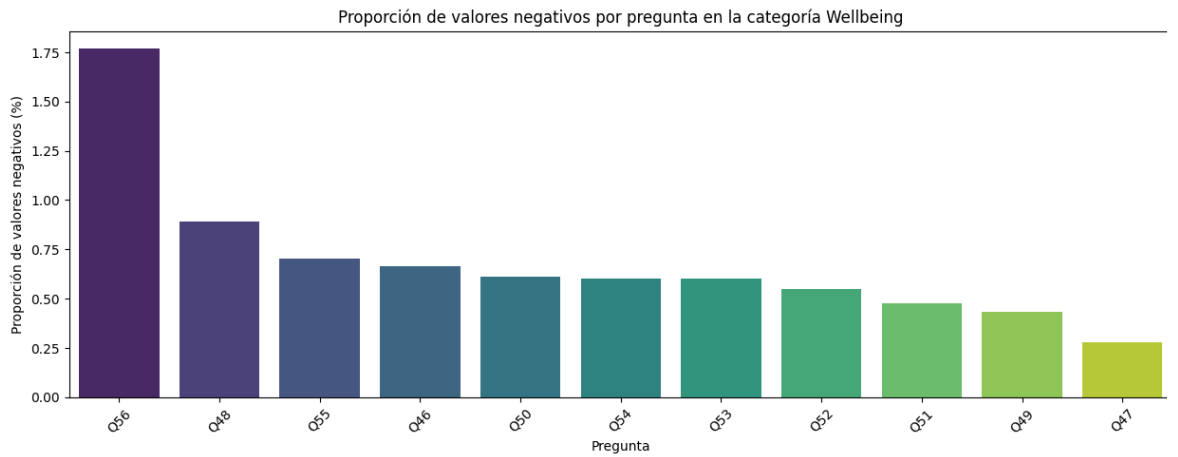
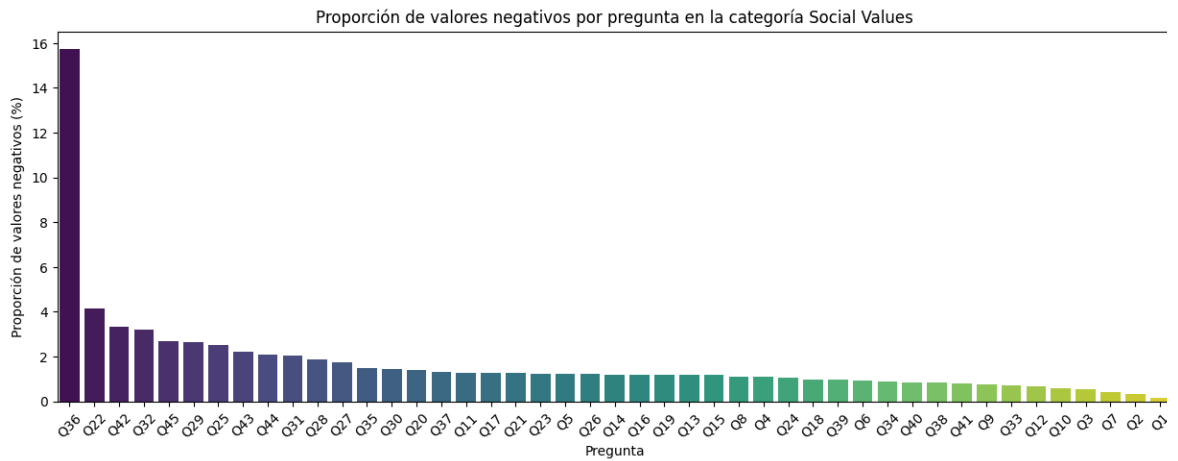
## 9. ANEXOS

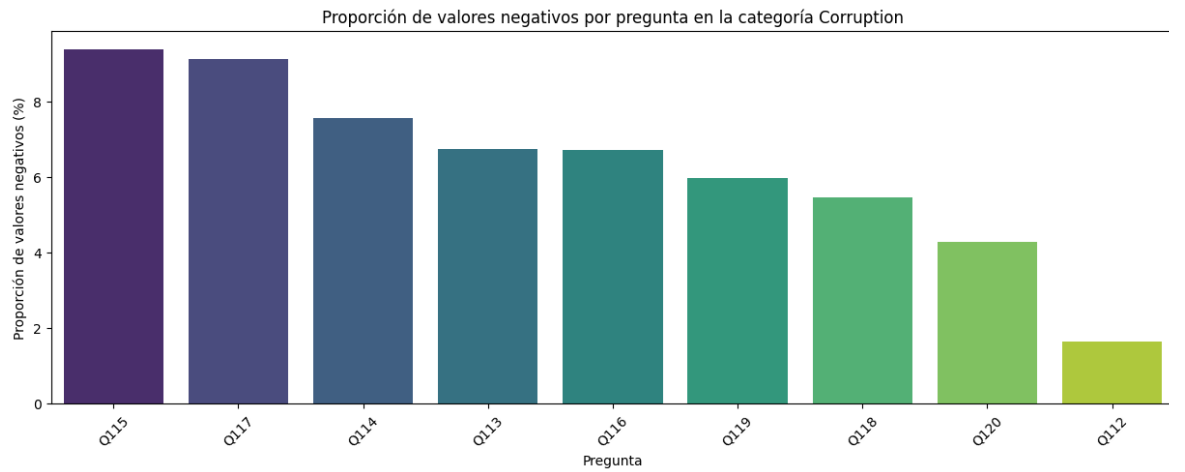
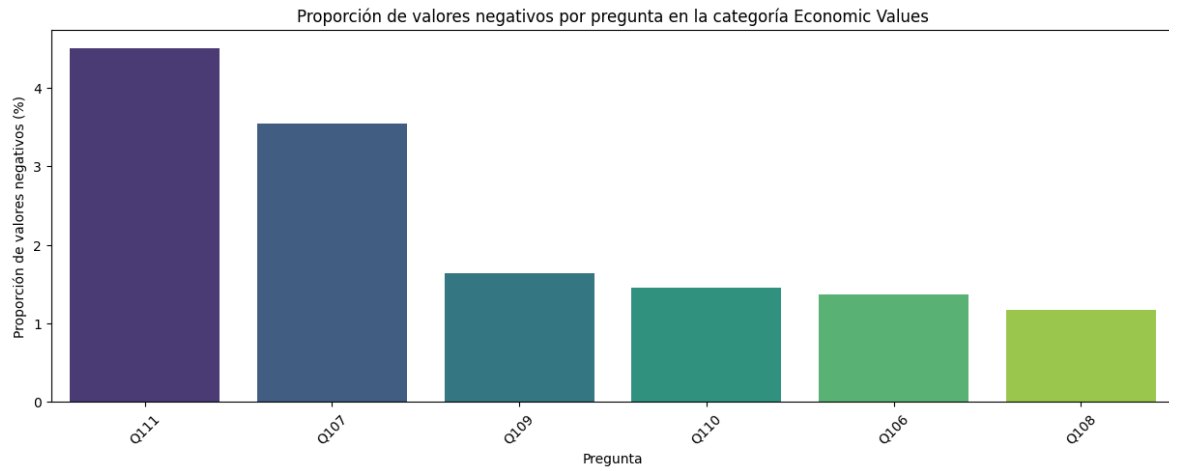
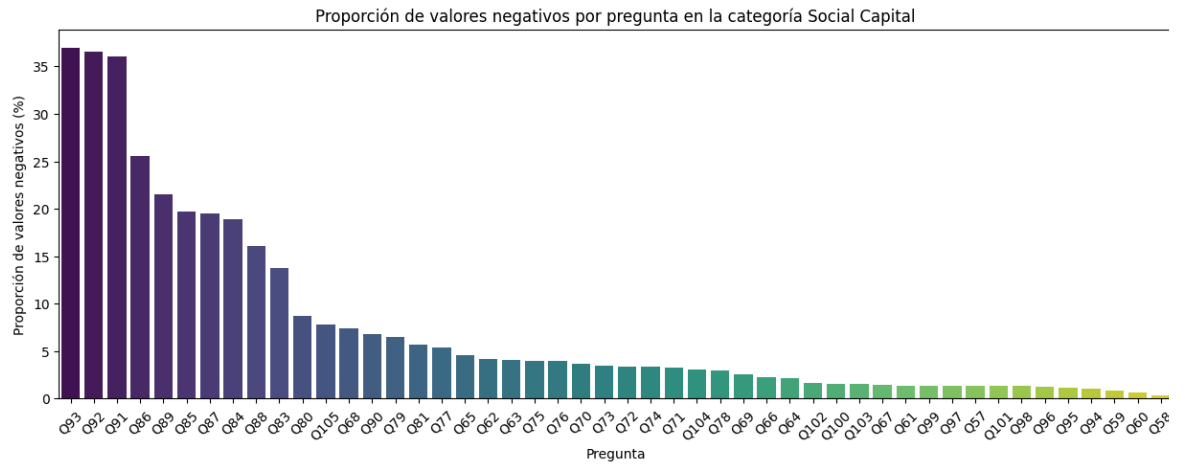
### 9.1. FORMULARIO MAESTRO DE LA ENCUESTA MUNDIAL DE VALORES, OLA 7, 2017-2020

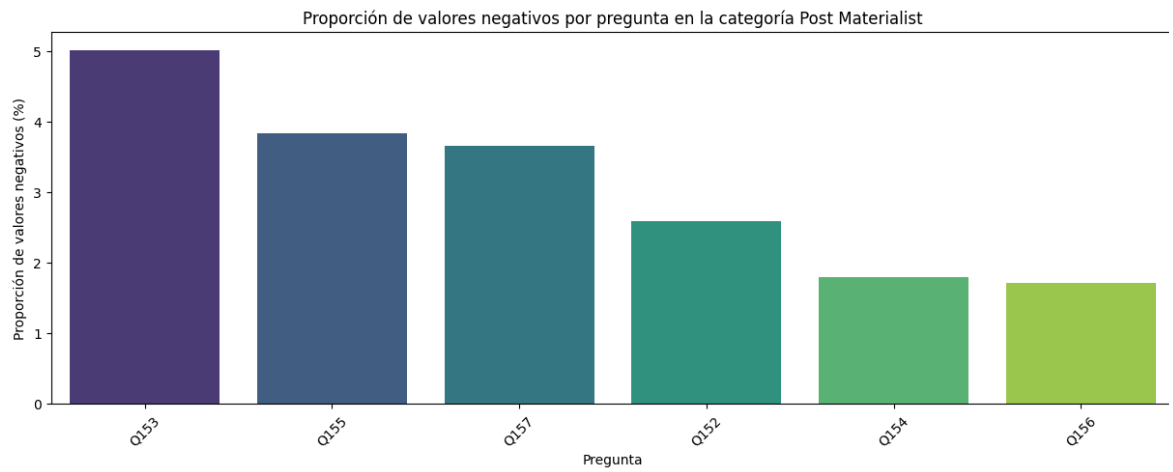
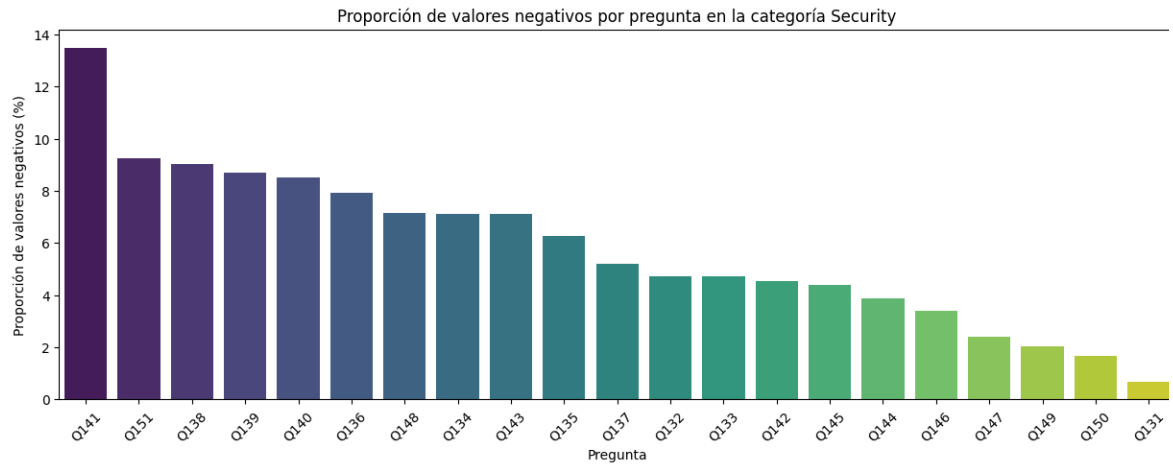
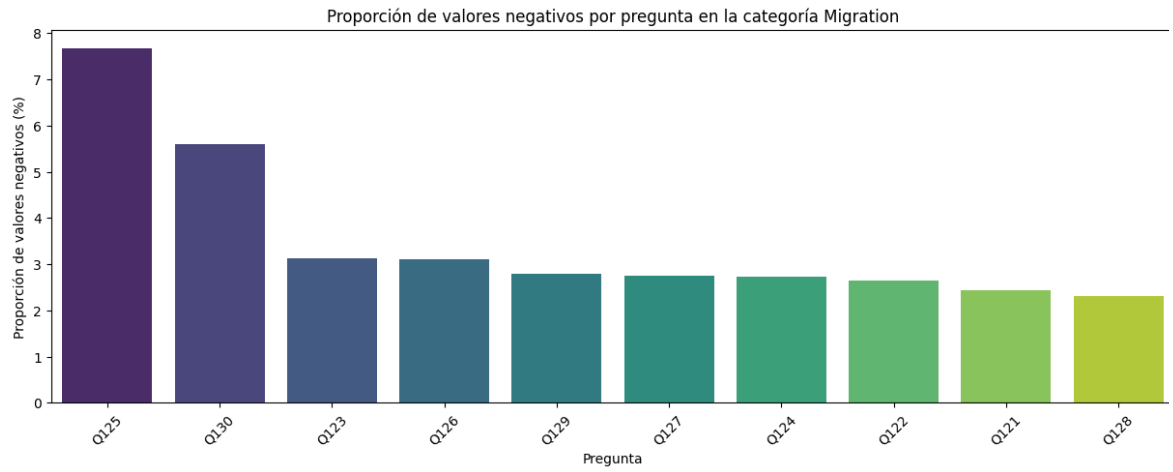


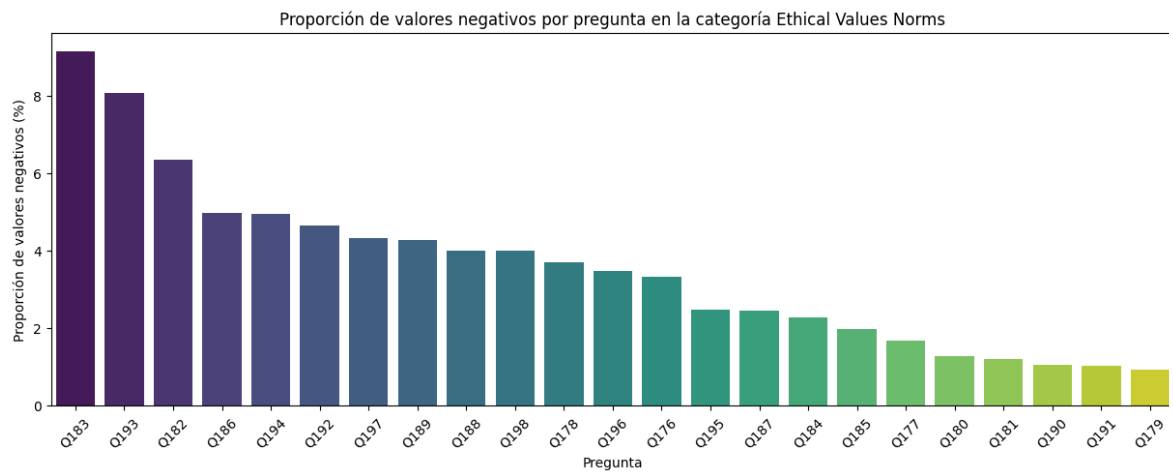
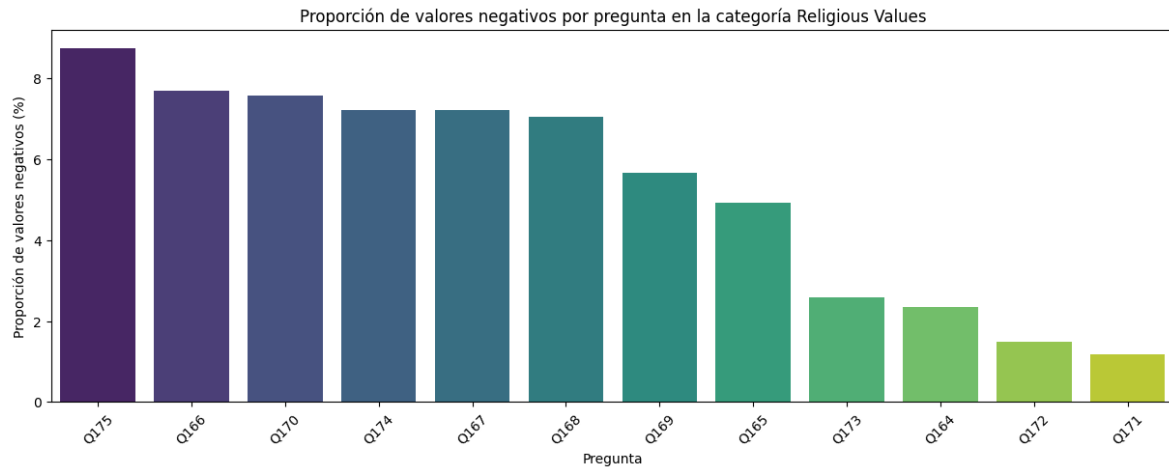
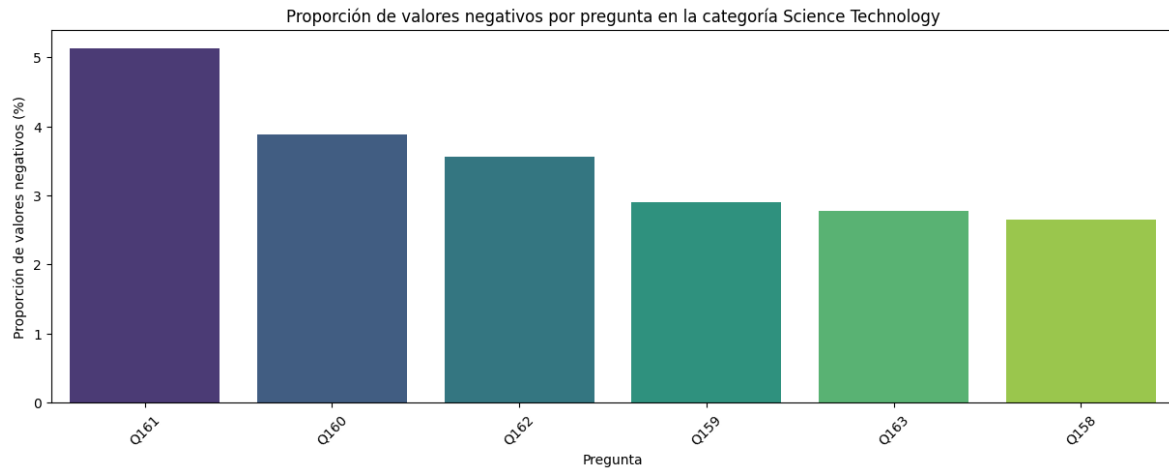
F00010738-WVS-7\_M  
aster\_Questionnaire\_2

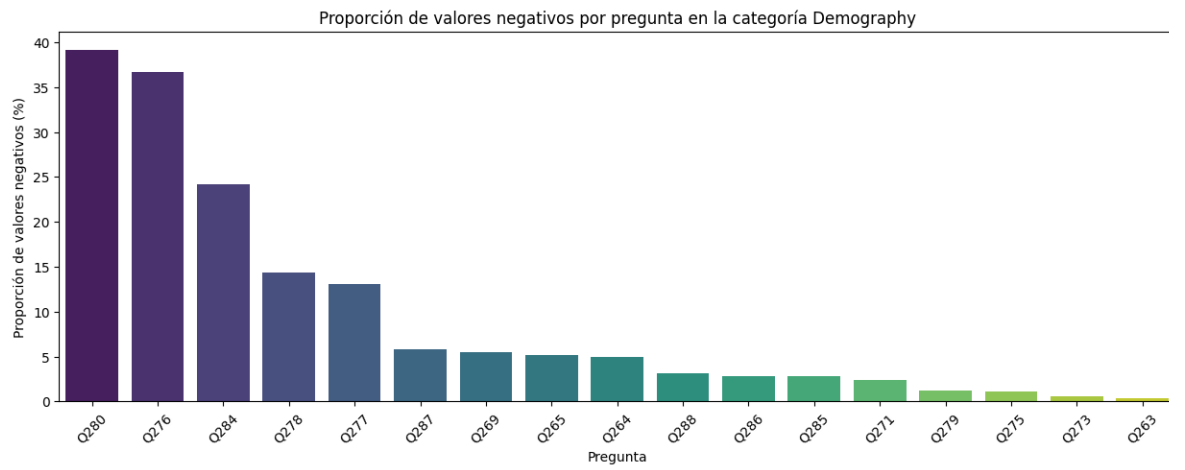
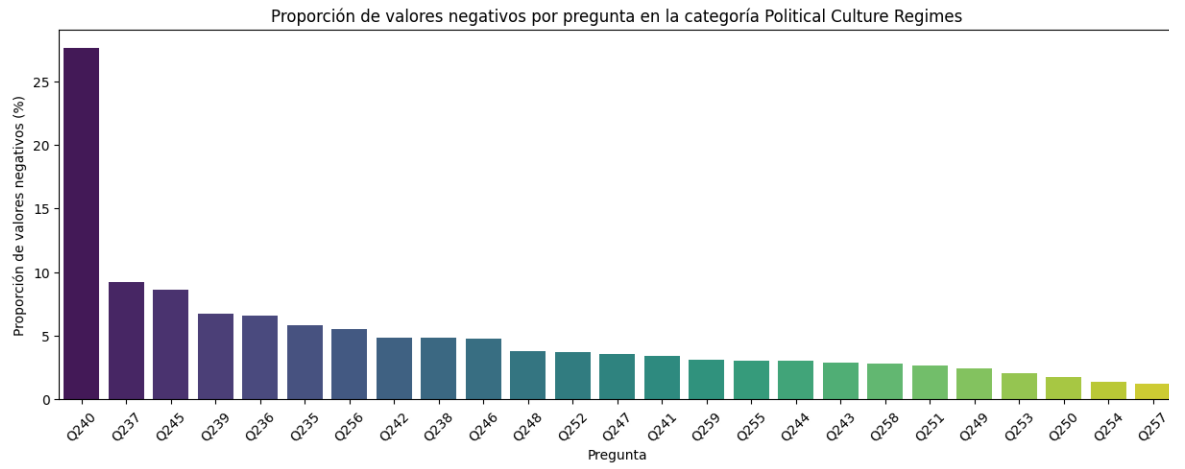
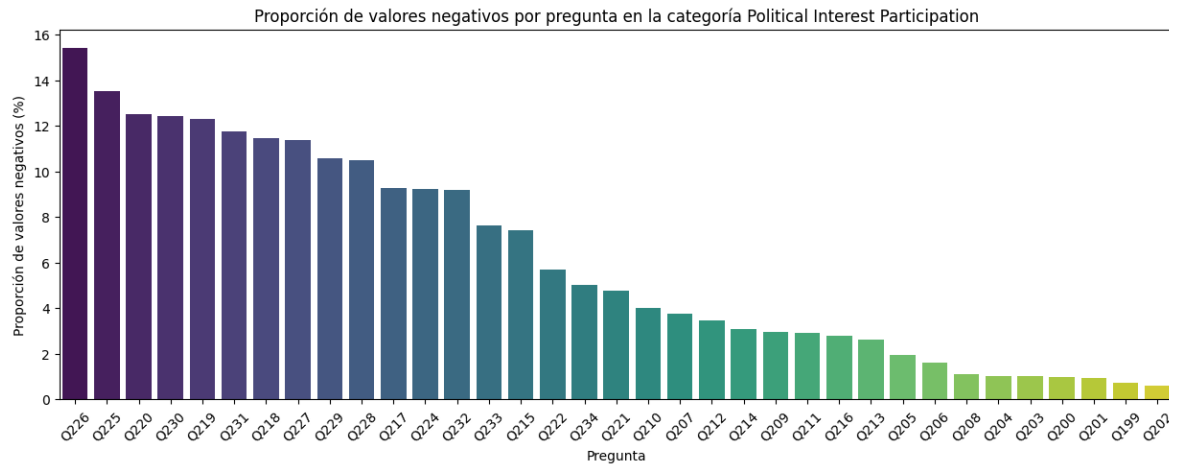
### 9.2. PROPORCIÓN DE RESPUESTAS CON VALOR NEGATIVO POR BLOQUE TEMÁTICO





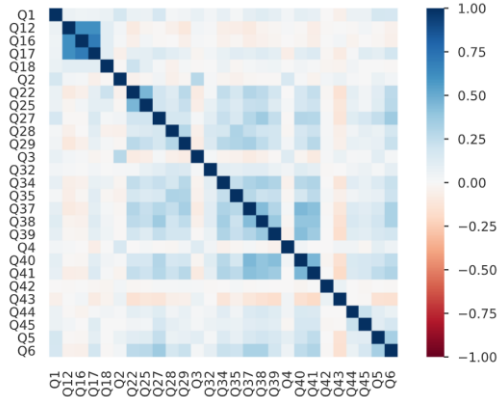




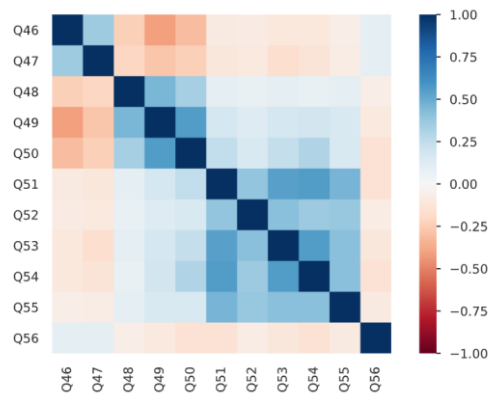


### 9.3. MATRICES DE CORRELACIÓN SIN DATOS FALTANTES POR BLOQUE TEMÁTICO

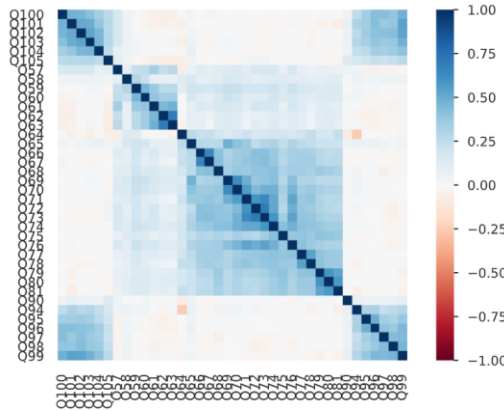
Valores sociales



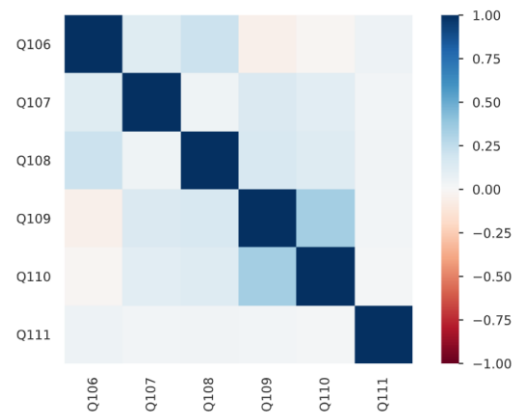
Bienestar social



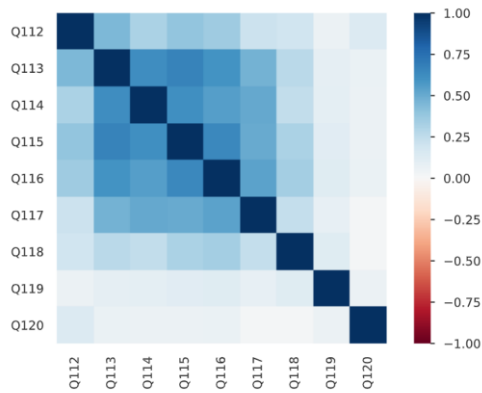
Capital social



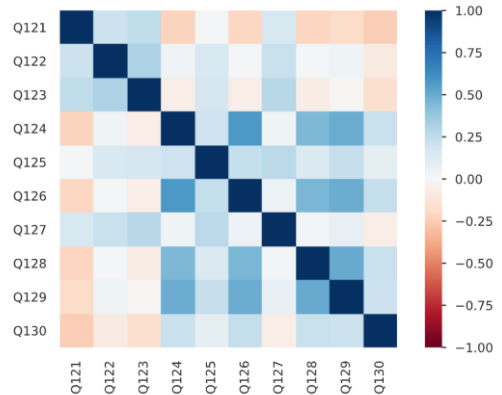
Valores económicos



Corrupción

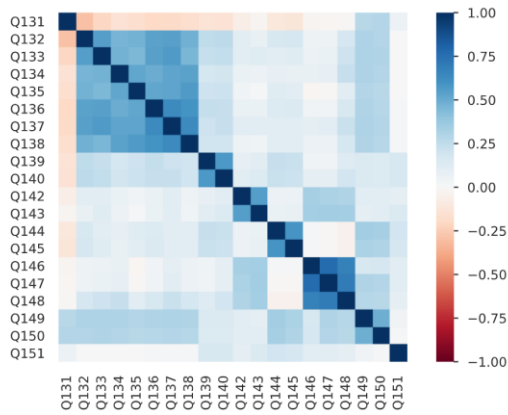


Migración

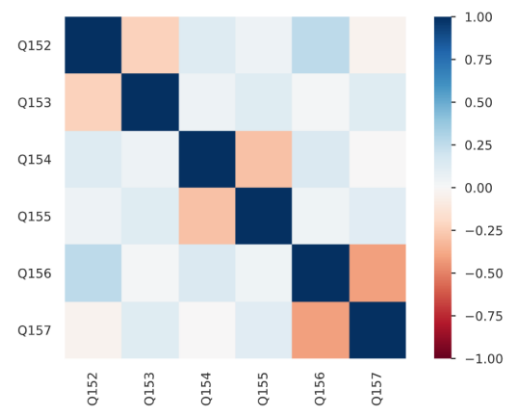


Seguridad

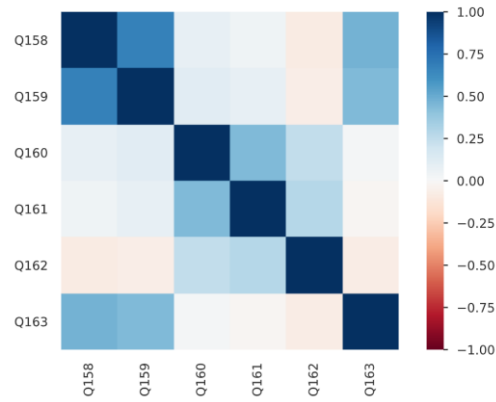
Índice postmaterialismo



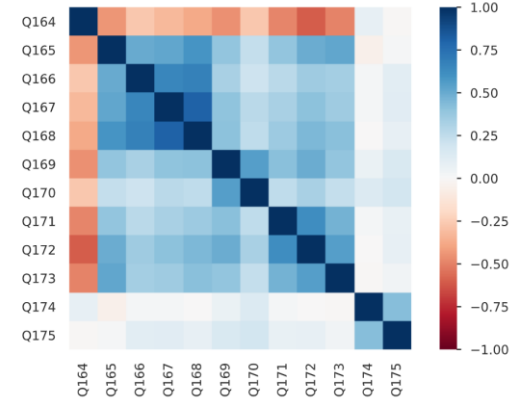
Ciencia y tecnología



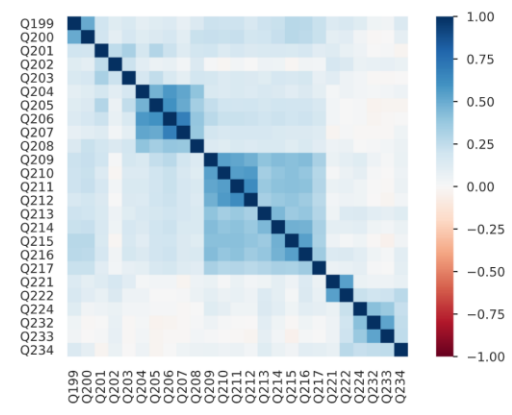
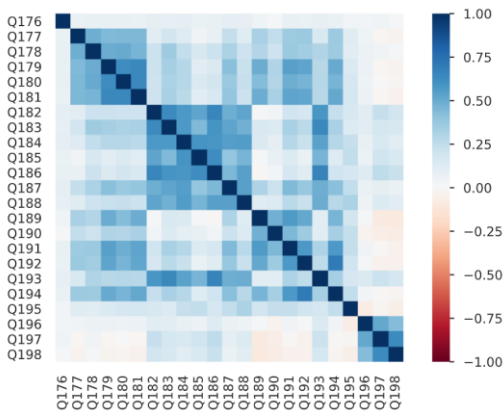
Valores religiosos



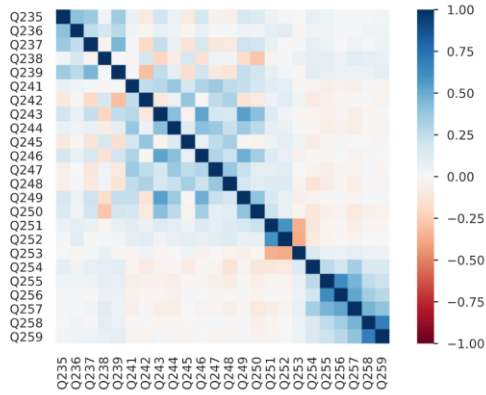
Valores éticos y normas



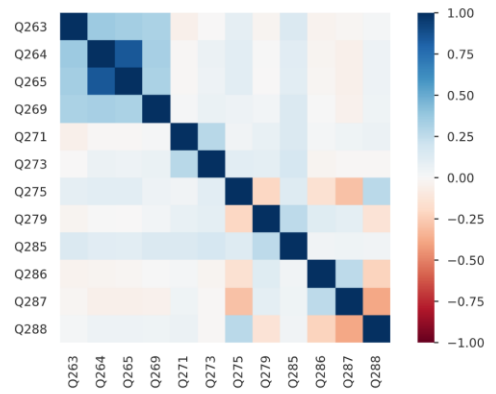
Interés por la política



Cultura y sistemas políticos

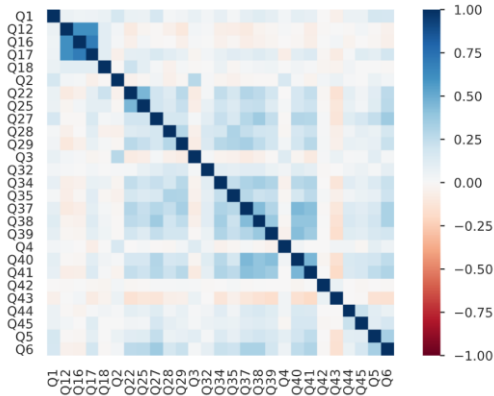


Datos sociodemográficos

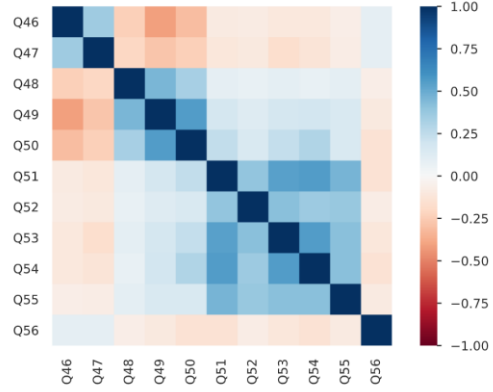


**9.4. MATRICES DE CORRELACIÓN FINAL POR BLOQUE TEMÁTICO**

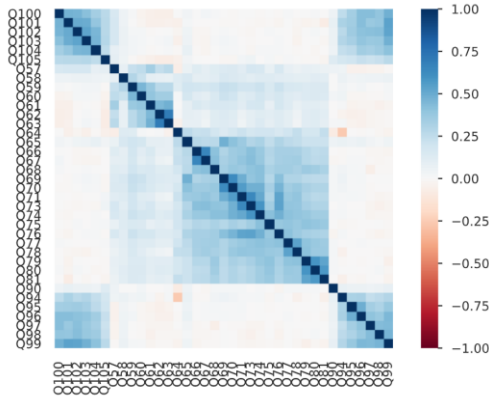
Valores sociales



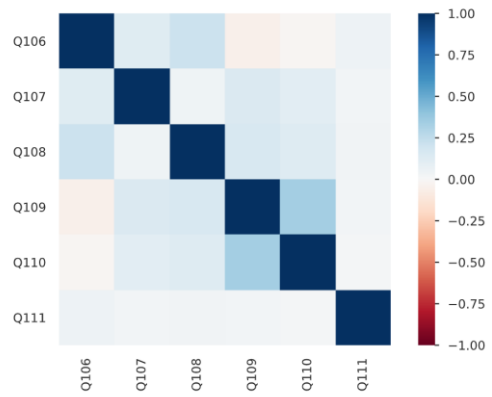
Bienestar social



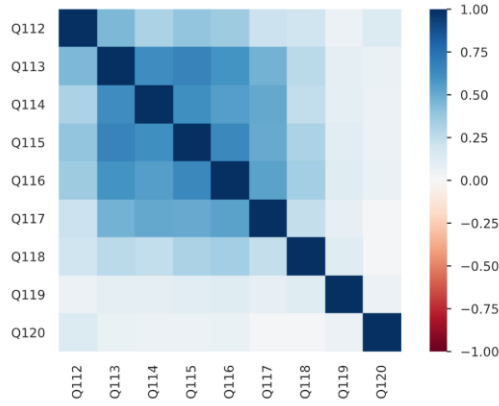
Capital social



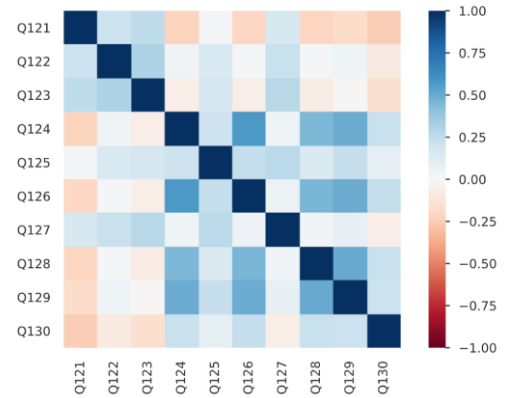
Valores económicos



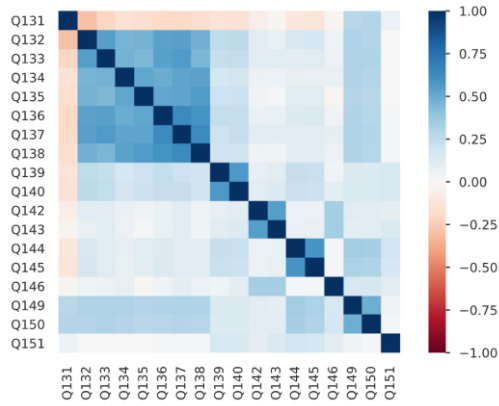
Corrupción



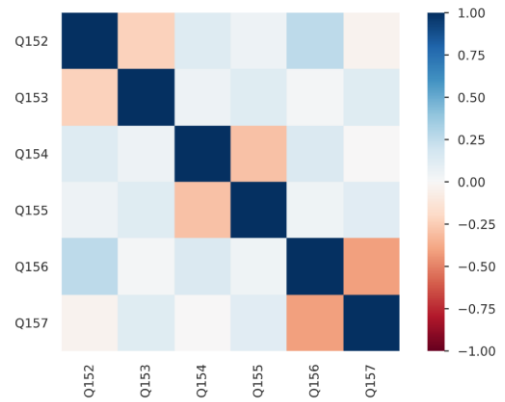
Migración



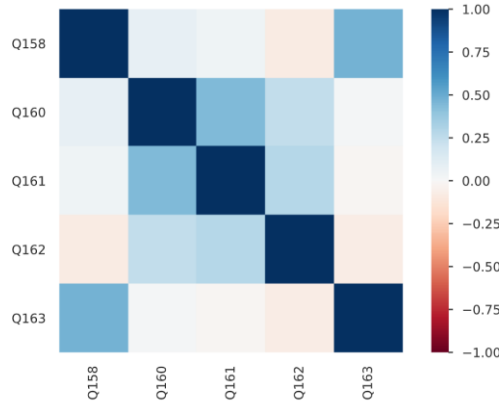
Seguridad



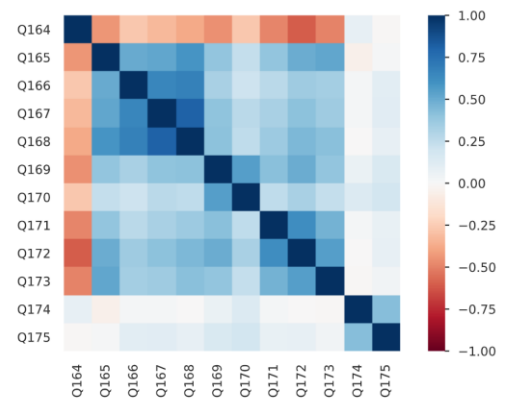
Índice postmaterialismo



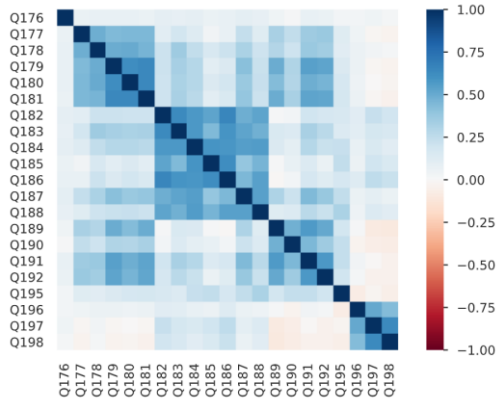
Ciencia y tecnología



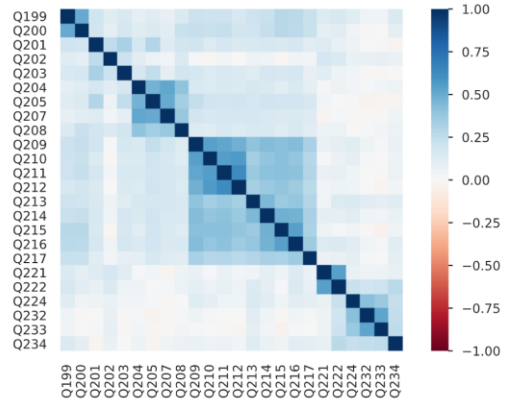
Valores religiosos



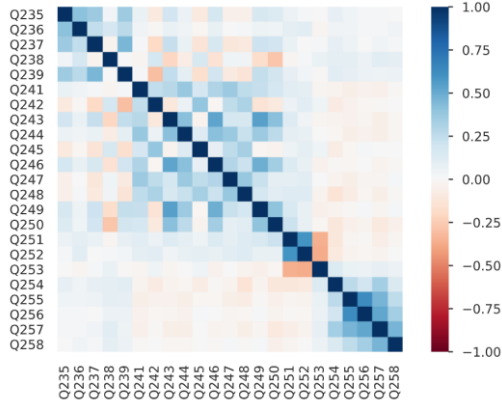
Valores éticos y normas



Interés por la política



Cultura y sistemas políticos



Datos sociodemográficos

