

**Correlación entre la piratería del software y los índices de desarrollo económico y humano:
análisis de componentes principales y agrupación.**

Palacios-Hernandez, Omar *
Programa de Doctorado
Transdisciplinario en Desarrollo
Científico y Tecnológico para la
Sociedad, Centro de Investigación y de
Estudios Avanzados del Instituto
Politécnico Nacional (Cinvestav), México
omar.palacios@cinvestav.mx

Padilla-Viveros, America A.
Programa de Doctorado
Transdisciplinario en Desarrollo
Científico y Tecnológico para la
Sociedad, Centro de Investigación y de
Estudios Avanzados del Instituto
Politécnico Nacional, México
aviveros@cinvestav.mx

Lopez-Fernandez, R.
Departamento de Física,
Centro de Investigación y de
Estudios Avanzados del
Instituto Politécnico
Nacional, México
ricardo.lopez@cinvestav.mx

*Autor de correspondencia

Palabras clave: Machine Learning, Propiedad Intelectual, Componentes Principales,
Cumplimiento, Clustering.

1. Abstract

En esta investigación se analizó la correlación que existe entre el desarrollo económico, el desarrollo humano y el uso de software pirata mediante un enfoque en la ciencia de datos. Se aplicó el método de componentes principales usando seis variables de cuatro tipos para 104 países, obteniendo la correlación entre: desarrollo económico, desarrollo humano, estado de derecho y piratería de software. El análisis de componentes principales y el proceso de agrupación jerárquica (HCPC) fueron realizados con el programa R Studio (R Studio Team, 2020). Derivado del análisis se identificó que la piratería (BSA) está más anti-correlacionada con la integridad legal de los

contratos. Un mayor nivel de integridad legal está correlacionado con un nivel de piratería más bajo. El resto de las variables, que están relacionadas con temas legales, también están correlacionadas con la integridad legal de los contratos y correlacionadas de manera inversa con el nivel de piratería de software. Otro hallazgo interesante fue que el PIB per cápita también está correlacionado con las variables de cumplimiento legal, es decir, los países en los que se observa un desarrollo económico más alto también tienen niveles de cumplimiento legal y estado de derecho más elevados. Se encontró que el índice de desarrollo humano está decorrelacionado con el resto de las variables. Finalmente, mediante el enfoque de agrupación jerárquica en componentes principales se identificó la formación de cuatro grupos de individuos (países) en el espacio de componentes principales donde se aprecian similitudes en los rasgos de las variables. Los resultados de este análisis sugieren que hay una correlación baja del desarrollo humano con el desarrollo económico general.

2. Introducción

En esta investigación se estudió la relación entre el índice de piratería de BSA-The Software Alliance (BSA, 2018) en 104 países del mundo con el desarrollo económico general, medido por el Producto Interno Bruto (PIB) per cápita, con el desarrollo humano, medido por el Índice de Desarrollo Humano de las Naciones Unidas (UNDP, 2020) y el cumplimiento del estado de derecho. Se busca comprender cómo las violaciones de los derechos de propiedad intelectual, específicamente el uso de software pirata, pueden afectar el desarrollo de un país, incluyendo el desarrollo económico y el desarrollo humano. Aunque la piratería de software puede aportar importantes bienes públicos o capital productivo a un país, lo que podría superar la pérdida de ingresos que presenta la industria del software por la piratería (Nobre-Fernandez et al., 2018), también existe evidencia que sugiere que los países con mayores niveles de piratería de software

tienen niveles más bajos de desarrollo económico general (Andrés & Goel, 2012). Estos resultados mostraron que la aplicación de los derechos de propiedad intelectual aumentó la actividad económica (Andrés & Goel, 2012).

Para realizar un análisis más robusto, se integró también el "índice de protección de derechos de propiedad, (Protection of property rights), el índice de integridad del sistema legal (Integrity of the legal system) y el índice de aplicación legal en contratos (Legal enforcement of contracts), estos últimos tres emitidos por "The Fraser Institute". Se considera que la protección de los derechos de propiedad intelectual (DPI) y el marco legal nacional son componentes fundamentales de una política económica sólida (Gwartney et al., 2019). La propiedad intelectual tiene las características de un bien público si produce derrames de conocimiento, la falta de propiedad privada o en su caso de un marco regulatorio y sistema comercial de dicha producción puede disminuir el incentivo para producir nuevos productos y tecnologías. Los avances tecnológicos recientes han generado una mayor disponibilidad de productos basados en la información, así como la tecnología utilizada para copiar o piratear dichos productos (OECD, 2016). A la luz de estos cambios, es relevante comprender el papel de la protección de los DPI en la promoción de los avances tecnológicos, la innovación, el desarrollo económico y humano. En esta investigación de alcance global se analizó la relación que existe entre el desarrollo económico, el desarrollo humano y el software pirata mediante un enfoque en la ciencia de datos.

3. Recursos y métodos.

En este documento se investigó empíricamente la relación entre los índices de piratería de software y el desarrollo, medido por el PIB per cápita, con el Índice de Desarrollo Humano (IDH)

de las Naciones Unidas y con el índice de protección a la propiedad intelectual cada país. El análisis de componentes principales (PCA) y el proceso de agrupación jerárquica (HCPC) fueron realizados con el programa R Studio Versión 1.2.5033 (R Studio Team, 2020).

El análisis de componentes principales es probablemente la más antigua y mejor conocida de las técnicas de análisis multivariante. Fue introducido por primera vez por Pearson (1901) y desarrollado independientemente por Hotelling (1933) (Jolliffe, 2002). La idea central del análisis de componentes principales es reducir la dimensionalidad de un conjunto de datos en el que hay una gran cantidad de variables interrelacionadas, mientras se retiene la mayor cantidad posible de la variación presente en el conjunto de datos. Esta reducción se logra transformando a un nuevo conjunto de variables, los componentes principales, que no están correlacionados, y que están ordenados de manera que los primeros retienen la mayor parte de la variación presente en todas las variables originales. El cálculo de los componentes principales se reduce a la solución de un problema de valores y vectores propios para una matriz simétrica positiva-semidefinida (Jolliffe, 2002). Por lo tanto, la definición y el cálculo de los componentes principales no es complicado, pero esta técnica aparentemente simple tiene gran aplicación y utilidad para análisis en los que se busca encontrar una correlación entre múltiples variables.

Recursos

Las variables consideradas para este estudio se seleccionaron con base en tres criterios: representatividad, disponibilidad y estandarización. Los datos relacionados con piratería se tomaron de la página oficial de la Alianza de Software de Negocios (BSA–Business Software Alliance <https://www.bsa.org/>). Por otra parte, los datos relacionados con el cumplimiento de leyes

y Estado de Derecho fueron tomados de la página de “The Fraser Institute” un centro de investigación canadiense que define su misión como “medir, estudiar, y comunicar el impacto de los mercados competitivos y la intervención gubernamental en el bienestar de los individuos” (<https://www.fraserinstitute.org/about>). Por último, los indicadores de desarrollo humano (HDI) y PIB per cápita (GNI) se tomaron de la página de datos de libre acceso del Banco Mundial (<https://datos.bancomundial.org>).

Se realizó un análisis de componentes principales (PCA) considerando seis variables de sistema legal en el párrafo anterior para 104 países del mundo y las variables de la estimación se desarrollaron de conformidad con la siguiente lista:

1. El índice de piratería de la BSA.
2. El Producto Interno Bruto per cápita (GNI).
3. El índice de desarrollo humano de la Organización de las Naciones Unidas (HDI).
4. El índice de protección de los derechos de propiedad del Instituto Fraser.
5. El índice de integridad del sistema legal del Instituto Fraser.
6. El índice de ejecución legal de contratos del Instituto Fraser.

En el Anexo 1 se muestran los 104 estados nacionales que se consideraron en el presente estudio.

Métodos

Análisis de componentes principales (PCA)

- a) La tabla 1 es una tabla pivote o matriz de correlaciones que muestra los seis componentes principales (vectores), llamados PC1, PC2, PC3, PC4, PC5 y PC6, que se calculan a partir de las correlaciones que tienen los seis indicadores originales en sus varianzas.

Tabla 1. Matriz de correlaciones de los componentes principales.

	PC1	PC2	PC3	PC4	PC5	PC6
Protection Of Property Rights	0.449	- 0.139	0.096	- 0.482	- 0.158	0.716
Integrity Of Legal System	0.443	- 0.035	0.274	- 0.413	0.579	- 0.471
Legal Enf Of Contracts	0.374	- 0.299	- 0.866	0.054	0.037	- 0.132
Índice BSA	0.448	0.029	- 0.242	- 0.087	0.743	0.425
HDI	0.282	0.932	- 0.206	0.048	0.031	0.072
GNI	0.426	- 0.141	0.255	0.765	0.293	0.250

Fuente: Elaboración propia

- b) Los valores de cada variable están representados en un espacio de seis dimensiones (matriz de 6 x 6), sin embargo, usando los resultados de la Tabla 2 y la figura 1, se puede observar que es posible seleccionar el número de componentes principales ya que su relevancia está determinada por la varianza.

Tabla 2. Importancia (ponderación) de los componentes principales.

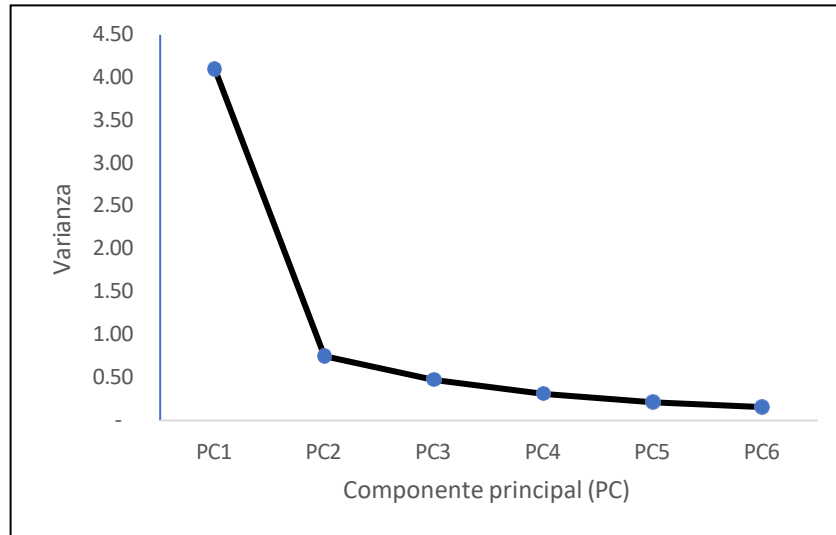
	PC1	PC2	PC3	PC4	PC5	PC6
Varianza	4.096	0.749	0.475	0.311	0.213	0.155
Proporción de varianza	0.6827	0.1249	0.07922	0.05181	0.03558	0.02583
Proporción de varianza acumulada	0.6827	0.8076	0.88679	0.93859	0.97417	1.0000

Fuente: Elaboración propia

Se observa que la proporción de varianza acumulada de PC1 y PC2 es de 0.80, mientras que la contribución de las varianzas de las otras componentes principales es marginal. A partir de la figura 2 sólo se representarán gráficamente las correlaciones que existan entre los indicadores en esos dos vectores como un plano cartesiano con PC1 y PC2 en el eje de las x y y, respectivamente.

- c) La figura 1 muestra las varianzas de las componentes principales (PC) y permite apreciar la proporción de cada una con respecto de las demás.

Figura 1. Varianzas de los componentes principales.

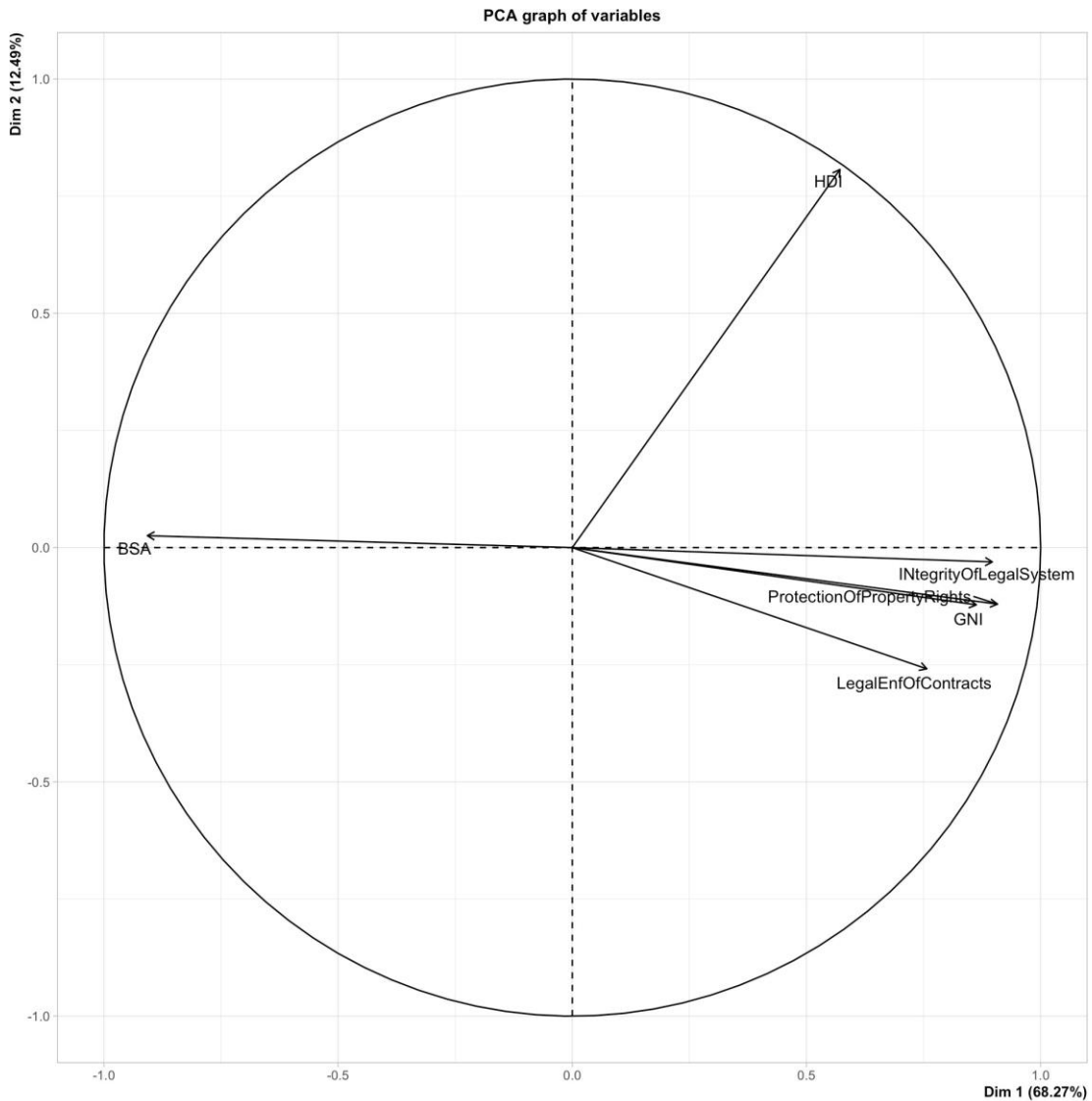


Fuente: Elaboración propia

- d) La figura 2 muestra a cada una de las variables como un vector con dirección y magnitud en el plano de las dos primeras componentes principales, representadas en un plano cartesiano las PC en el eje de las x y y, respectivamente.

Usando este tipo de representación se puede extraer información a partir de simples disposiciones geométricas. Por ejemplo, entre más pequeño sea el ángulo formado por los vectores, mayor será la correlación directa entre esas variables. Entre más cercano sea el ángulo entre dos vectores a 180° , mayor será su anti-correlación, como es el caso de "integridad del sistema legal" y "BSA". Si el ángulo es cercano a los 90° , entonces la correlación con las variables es baja, como es el caso del HDI con respecto al resto de las demás variables. La longitud de las flechas indica qué tan bien está representada una variable; mientras más cercana esté al 1.00 (perímetro del círculo), mejor representada gráficamente está esa variable.

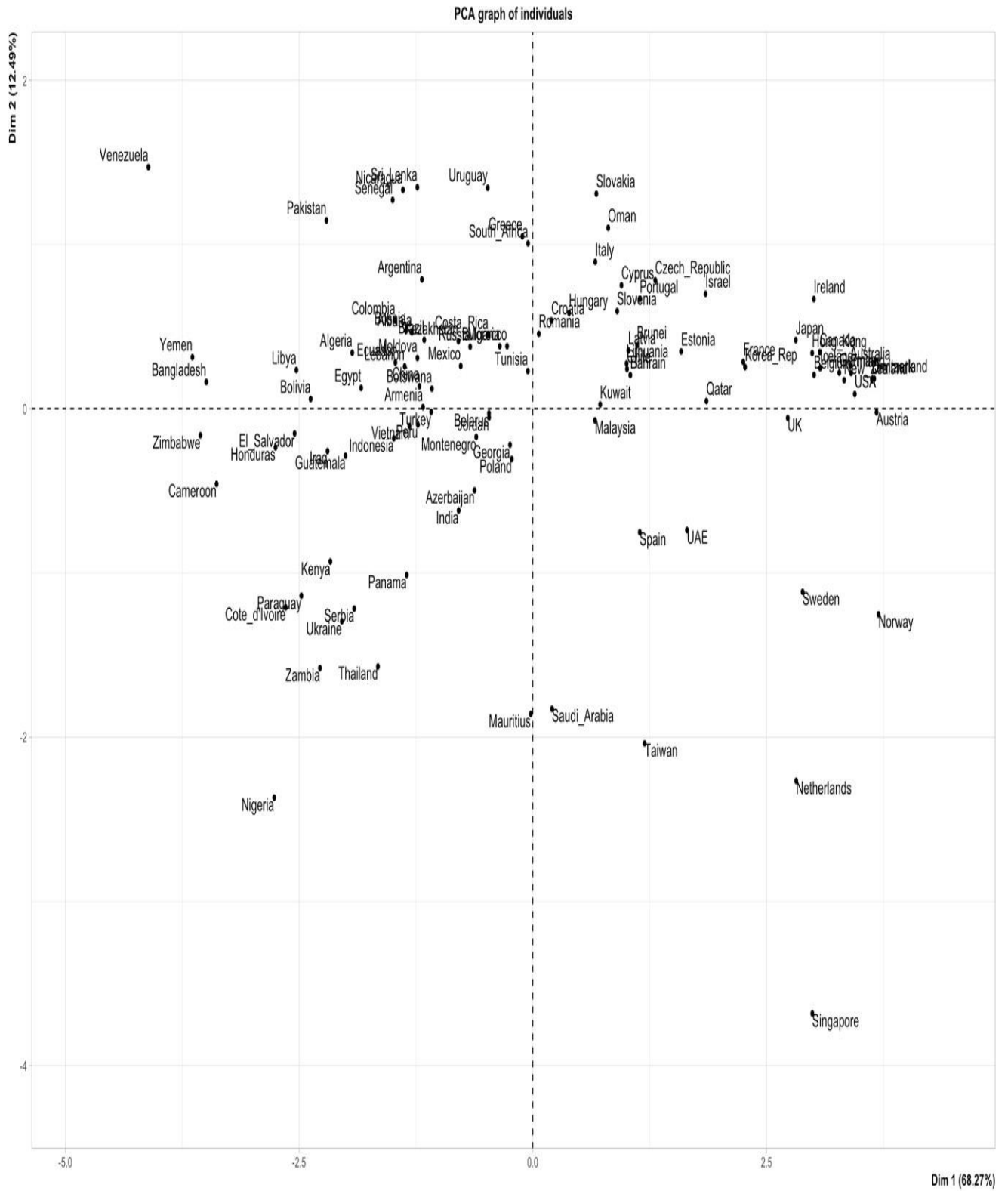
Figura 2: Variables representadas como vectores en el plano de las dos primeras componentes principales.



Fuente: Elaboración propia

- e) En la figura 3 se muestra la ubicación de cada uno de los 104 individuos en el mismo plano de los dos primeros componentes principales. A primera vista se observa una correspondencia entre los indicadores de estado de derecho, posición en el plano y ubicación geográfica.

Figura 3. Ubicación de los individuos (países) por componente principal.



Fuente: Elaboración propia

Análisis de agrupamiento en componentes principales

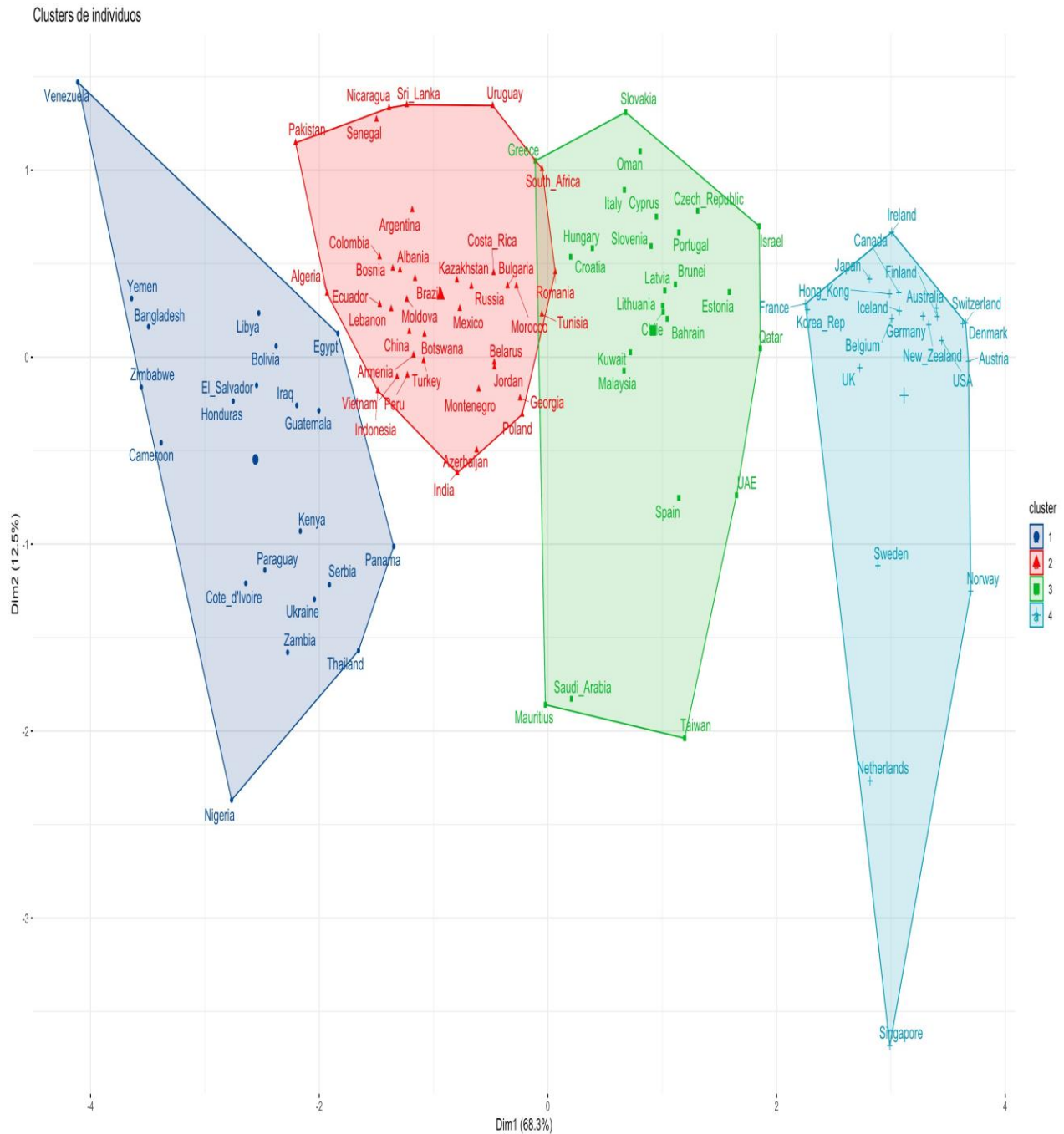
El análisis de agrupamiento es uno de los métodos de minería de datos más importantes para obtener hallazgos relevantes en conjuntos de datos multivariados. El objetivo es identificar grupos de objetos similares dentro de un conjunto de datos de interés. Las dos estrategias de agrupamiento más comunes son: a) Agrupación jerárquica, utilizada para identificar grupos de observaciones similares en un conjunto de datos y b) Agrupación de particiones como el algoritmo k-means (medias), que se utiliza para dividir un conjunto de datos en varios grupos.

El enfoque HCPC nos permite combinar los tres métodos estándar utilizados en los análisis de datos multivariados (Husson, Josse y J. 2010): a) Métodos de componentes principales; b) Agrupación jerárquica y; c) Agrupación de particiones, particularmente el método de k-medias.

Después de realizar el PCA para reducir la dimensión de los datos en pocas variables continuas que contengan la información más importante en los datos, se puede realizar un análisis de agrupamiento en los resultados del PCA. El PCA también se puede considerar como un paso de eliminación de ruido que puede conducir a un agrupamiento más estable. Esto puede resultar muy útil si tiene un gran conjunto de datos con múltiples variables, como es el caso de la correlación que se estudió en esta investigación.

En la figura 4 se muestran nuevamente los 104 individuos (países) en el plano de los dos primeros componentes principales agrupándolos por su distancia euclidiana.

Figura 4. Agrupación de individuos en el plano de los componentes principales 1 y 2.



Fuente: Elaboración propia

Los individuos se consideran en un espacio euclidiano de alta dimensión. Estudiar las similitudes entre los individuos significa estudiar la forma de la nube de puntos. Luego, los métodos de componentes principales aproximan esta nube de puntos a un sub-espacio euclidiano

de dimensiones más bajas mientras preservan tanto como sea posible las distancias entre individuos (Husson et al., 2010).

En la figura 4 se observa que en dos dimensiones existen cuatro grupos de países y la división está determinada por el eje del PC1. A pesar de que pudiera parecer un agrupamiento trivial determinado por el eje del PC1, es pertinente recordar que el PC1 tiene contribuciones de todas las variables, por lo tanto, este agrupamiento no hubiera sido posible sin el análisis previo de componentes principales.

4. Resultados

Derivado del PCA, y usando la información de la figura 2, se identificó que la piratería (BSA) está más anti-correlacionada con la integridad legal de los contratos. Un mayor nivel de integridad legal está correlacionado con un nivel de piratería más bajo. El resto de las variables, que están relacionadas con temas legales, también están correlacionadas con la integridad legal de los contratos y correlacionadas de manera inversa con el nivel de piratería de software.

Otro hallazgo relevante es que el PIB per cápita también está correlacionado con las variables de cumplimiento legal, es decir, los países en los que se observa un desarrollo económico más alto también tienen niveles de cumplimiento legal y estado de derecho más elevados.

El índice de desarrollo humano, por su parte, está decorrelacionado con el resto de las variables, es decir, su comportamiento no tiene relación con las demás.

Se podría pensar que una mayor actividad en la piratería de software se correlaciona con un nivel más bajo de IDH, mientras que el ingreso per cápita y la libertad económica se correlacionan positivamente con el IDH. Sin embargo, los resultados de este análisis sugieren que,

con los datos aportados, hay una correlación baja del desarrollo humano con el desarrollo económico general.

La mayoría de los estudios sobre desarrollo económico emplean una de dos medidas comunes de desarrollo económico: ingreso per cápita o crecimiento del ingreso. Sin embargo, estas medidas utilizadas tradicionalmente no miden necesariamente el desarrollo social. El término “desarrollo social” a menudo se refiere a un concepto de mejora más completo o más amplio que los ingresos por sí solos.

Como señala Sen et al., (1999), “una concepción adecuada del desarrollo debe ir mucho más allá de la acumulación de riqueza y el crecimiento del producto nacional bruto y otras variables relacionadas con el ingreso”. En otras palabras, las tasas de crecimiento no tienen por qué ser sinónimo de desarrollo económico. En este sentido, los hallazgos de esta investigación permiten observar que, aunque no necesariamente existe una correlación con el HDI, sí existe una correlación directa con las variables de aplicación de estado de derecho y con la de PIB per cápita.

Si las estimaciones de Business Software Alliance (BSA-The Software Alliance, 2018) son precisas, el valor del software robado en todo el mundo fue de aproximadamente 13.000 millones de dólares anuales para 2018. Aunque no impacta de manera nominal, sí existe una correlación. Por lo tanto, es poco probable que sea posible detectar un cambio en el producto interno bruto solo por la piratería de software. Sin embargo, con el tiempo es posible que la piratería de software influya en el desarrollo económico general. Revisar esta posible influencia podría representar una línea de investigación extendida de este estudio.

Dado que el PIB es un componente del HDI (UNDP, 2020), se podría asumir que debería existir una correlación positiva entre las dos medidas. Sin embargo, un hallazgo adicional inesperado consistió en que, al compararlo con el PIB per cápita, en este análisis encontramos una correlación baja entre el PIB per cápita y el HDI, esto podría deberse al hecho de considerar otros componentes para el cálculo del HDI, incluida la esperanza de vida, el acceso al agua potable, a la atención médica y educación. Por lo tanto, si la piratería de software se correlaciona con una mejora de cualquiera de los componentes no basados en los ingresos del índice de desarrollo humano, es posible que el efecto total de la piratería de software no se refleje únicamente en cambios en el producto interno bruto. Una ampliación interesante de este trabajo en el futuro sería un análisis de un posible “efecto multiplicador” de la piratería de software para revelarnos más información sobre impactos no considerados en el alcance de este trabajo exploratorio.

5. Agradecimientos

Omar Palacios-Hernandez agradece al Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav) por el apoyo DCTS-CN-2021-295 para la asistencia al XIX Congreso Latino-Iberoamericano de Gestión Tecnológica y de la Innovación ALTEC 2021 y al Consejo Nacional de Ciencia y Tecnología (Conacyt) por la beca.

America A. Padilla-Viveros agradece al Programa de Doctorado Transdisciplinario en Desarrollo Científico y Tecnológico para la Sociedad del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav) por el apoyo proporcionado para la asistencia al congreso.

6. Referencias

- Andrés, A. R., & Goel, R. K. (2012). Does software piracy affect economic growth? Evidence across countries. *Journal of Policy Modeling*. <https://doi.org/10.1016/j.jpolmod.2011.08.014>
- BSA-The Software Alliance. (2018). 2018 BSA GLOBAL CLOUD COMPUTING SCORECARD. *Business Software Alliance*.
- Gwartney, J. D., Lawson, R. A., Hall, J., & Murphy, R. (2019). Economic freedom of the world annual report. In *Fraser Institute*.
- Husson, F., Josse, J., & Pages, J. (2010). Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? *Agrocampus*, 18(1), 17–26. <https://doi.org/10.2478/v10037-010-0003-0>
- Jolliffe, I. T. (2002). Principal Components Analysis. In *Principal Components Analysis*. Elsevier. <https://doi.org/10.1016/B978-0-08-044894-7.01358-0>
- Nobre Fernandez, R., Garcia Ribeiro, F., & Del Ponte Duarte, J. M. (2018). Effects of Software Piracy on Economic Growth. *International Journal of Economics and Finance*. <https://doi.org/10.5539/ijef.v10n6p1>
- OECD. (2016). *Trade in Counterfeit and Pirated Goods: Mapping the Economic Impact*.
- R Studio Team. (2020). *RStudio: Integrated Development for R* (1.2.5033). RStudio, PBC. <http://www.rstudio.com/>
- Sen, A., Miletzki, J., Broten, N., Rifkin, S. B., & Visser, W. (1999). Development as freedom. In *Econometrica*.
- UNDP. (2020). *Human Development Index (HDI) | Human Development Reports*. United Nations Development Programme. <http://hdr.undp.org/en/content/human-development-index-hdi>

ANEXO 1:

Lista de los 104 estados nacionales que se consideraron en el presente estudio.

Albania	Colombia	India
Algeria	Costa_Rica	Indonesia
Argentina	Coted'Ivoire	Iraq
Armenia	Croatia	Ireland
Australia	Cyprus	Israel
Austria	Czech Republic	Italy
Azerbaijan	Denmark	Japan
Bahrain	Ecuador	Jordan
Bangladesh	Egypt	Kazakhstan
Belarus	El_Salvador	Kenya
Belgium	Estonia	Korea_Rep
Bolivia	Finland	Kuwait
Bosnia	France	Latvia
Botswana	Georgia	Lebanon
Brazil	Germany	Libya
Brunei	Greece	Lithuania
Bulgaria	Guatemala	Malaysia
Cameroon	Honduras	Mauritius
Canada	Hong_Kong	Mexico
Chile	Hungary	Moldova
China	Iceland	Montenegro

Morocco	Romania	Thailand
Netherlands	Russia	Tunisia
New Zealand	SaudiArabia	Turkey
Nicaragua	Senegal	Ukraine
Nigeria	Serbia	UAE
Norway	Singapore	UK
Oman	Slovakia	USA
Pakistan	Slovenia	Uruguay
Panama	South_Africa	Venezuela
Paraguay	Spain	Vietnam
Peru	Sri_Lanka	Yemen
Poland	Sweden	Zambia
Portugal	Switzerland	Zimbabwe
Qatar	Taiwan	