
Análisis de Minería de Datos sobre el Impacto y Gestión Global del COVID-19



ASIGNATURA: ALMACENES Y MINERÍA DE DATOS

TRABAJO

CURSO: 2025/26

Pascual Albericio, Irene (NIP 871627)

Vázquez Martín, Lucía (NIP 871886)

1. Metodología

El proyecto sigue un flujo de trabajo estructurado en tres fases analíticas ejecutadas en R.

En primer lugar, para la segmentación, se estandarizaron los datos y se determinó mediante el método del codo que $K=5$ era el número óptimo de grupos. Se aplicó el algoritmo K-Means y se programó una asignación lógica de etiquetas para asegurar que cada cluster mantuviera siempre su misma identidad independientemente de la ejecución. Esta estructura de identidades se validó mediante un dendrograma jerárquico, que mostró que claramente había 5 grupos de países de forma natural.

En segundo lugar, se realizó una validación estadística. Se inició con un análisis de correlación y modelos de regresión lineal simple para evaluar el impacto individual de cada variable (Rigor, Edad y PIB). Posteriormente, se ejecutó una regresión lineal múltiple específica combinando Edad y PIB para aislar el factor demográfico. Esto permitió demostrar que, al descontar el efecto de la edad, la riqueza se revela como un factor protector (reduciendo la mortalidad), desmintiendo así la aparente paradoja de 'a más dinero, más muertes'.

Finalmente, la interpretación se apoyó en técnicas de visualización avanzada (gráficas con líneas de tendencia y puntos, diagramas de caja, etc) desarrolladas en lenguaje R.

2. Repositorios de información utilizados

La fuente primaria de información ha sido el repositorio oficial de **Our World in Data** (OWID), del cual se extrajo una instantánea consolidada de los datos globales de la pandemia. El dataset final integra variables de impacto sanitario (muertes totales por millón), gestión política (índice de rigurosidad), demografía (edad media y densidad) y economía (PIB per cápita). En la fase de limpieza, se filtró la muestra a 150 países representativos, eliminando agregados continentales para evitar sesgos. Se trataron los valores nulos mediante imputación por medias regionales y, fundamentalmente, se aplicó una normalización a todas las variables numéricas para neutralizar las diferencias de magnitud y asegurar el correcto funcionamiento de los algoritmos de distancia.

3. Esfuerzos invertidos

A continuación, se detallan las tareas realizadas para el desarrollo del trabajo. De acuerdo con la metodología establecida, todas las actividades se llevaron a cabo de forma conjunta, mediante programación en pareja y revisión simultánea de los resultados.

Apartados del trabajo realizados	Lucía	Irene
Análisis del enfoque a seguir y extracción de datos	2	2
Desarrollo de código en R	6	6
Análisis de los resultados obtenidos en R	3	3
Desarrollo de memoria y presentación	3	3
	14	14

4. Webgrafía

Bonaccio, M., Iacoviello, L., Donati, M. B., & De Gaetano, G. (2020). A Socioeconomic Paradox in the COVID-19 Pandemic in Italy: a Call to Study Determinants of Disease Severity in High and Low-Income Countries. [pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC7340247/).

<https://pmc.ncbi.nlm.nih.gov/articles/PMC7340247/>

COVID-19 Government Response Tracker. (2020, 1 marzo). Blavatnik School Of Government. <https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker>

Dowd, J. B., Andriano, L., Brazel, D. M., Rotondi, V., Block, P., Ding, X., Liu, Y., & Mills, M. C. (2020). Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings Of The National Academy Of Sciences*, 117(18), 9696-9698. <https://doi.org/10.1073/pnas.2004911117>

Frias, L. (s. f.). ISCIII CNE - RENAVE. <https://cnecovid.isciii.es/>

Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Gavrilov, D., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., & Roser, M. (2020, 7 marzo). COVID-19 pandemic. Our World In Data. <https://ourworldindata.org/coronavirus>

Naciones Unidas. (2020, mayo). Informe de políticas: Los efectos de la COVID-19 en las personas de edad. https://www.un.org/sites/un2.un.org/files/2020/10/old_persons_spanish.pdf