



# 2021 INFORME

## EXPORTACIONES ARGENTINAS DE EQUIPAMIENTO MÉDICO

---

Elaborado por

Franco Godino  
Franco Mizzau  
Eliana Glasiuk  
Caroll Lema Cuesta



## Introducción

---

El principal objetivo de este trabajo es realizar una segmentación de mercados que permita agrupar a los distintos países considerando las exportaciones argentinas de equipamientos médicos hacia ellos. Con este fin, se trabajó con información sobre este tipo de comercio, juntamente con variables consideradas de interés para caracterizar a cada una de las naciones incluídas. Dentro de la información comercial se encontraba el valor y participación de importaciones desde Argentina (además de otros países relevantes). Por otro lado se ofrecían 50 variables que resumen las condiciones de salud, demográficas y económicas de los países.

Dado que existen más de medio centenar de variables para la caracterización de los destinos de las exportaciones argentinas, primeramente es necesario reducir la dimensionalidad. Una vez establecida la reducción de dimensionalidad se han de clasificar a los países de manera que se pueda comparar su relación comercial con Argentina y en función de ella obtener grupos de países similares en sus características.

De ser exitosa la clasificación, podríamos obtener a partir de ella una lista de ‘clientes potenciales’ para las exportaciones argentinas de equipamiento médico basada en su relativa similitud con clientes actuales de igual porte. Aquí es necesario aclarar que, independientemente del éxito en esta clasificación desde un punto de vista matemático, extrapolar las características de grandes clientes actuales no es necesariamente correcto desde un punto de vista teórico y, en particular, se trata de un análisis estático dado que los datos no son temporales sino que son una foto en un momento determinado del tiempo.

---

<sup>1</sup> En justicia se trata de un promedio y no tanto de una foto, pues muchas de las variables utilizadas son los “promedios de los últimos 5 años” para sus valores. Esto no cambia sin embargo el carácter estático de los datos.

# Tratamiento de los Datos

---

La base se encontraba dividida en dos, una con información de comercio exterior y otra con variables agrupadas con distinta información: demográficas, económicas, institucionales, salud, gasto en salud, exportaciones/importaciones para cada país. Para la segunda base, fue necesario buscar información complementaria, ya que había gran cantidad de datos faltantes. Estos se obtuvieron de la base de datos del Banco Mundial, siempre respetando los mismos criterios de la base original en cuanto al periodo seleccionado.

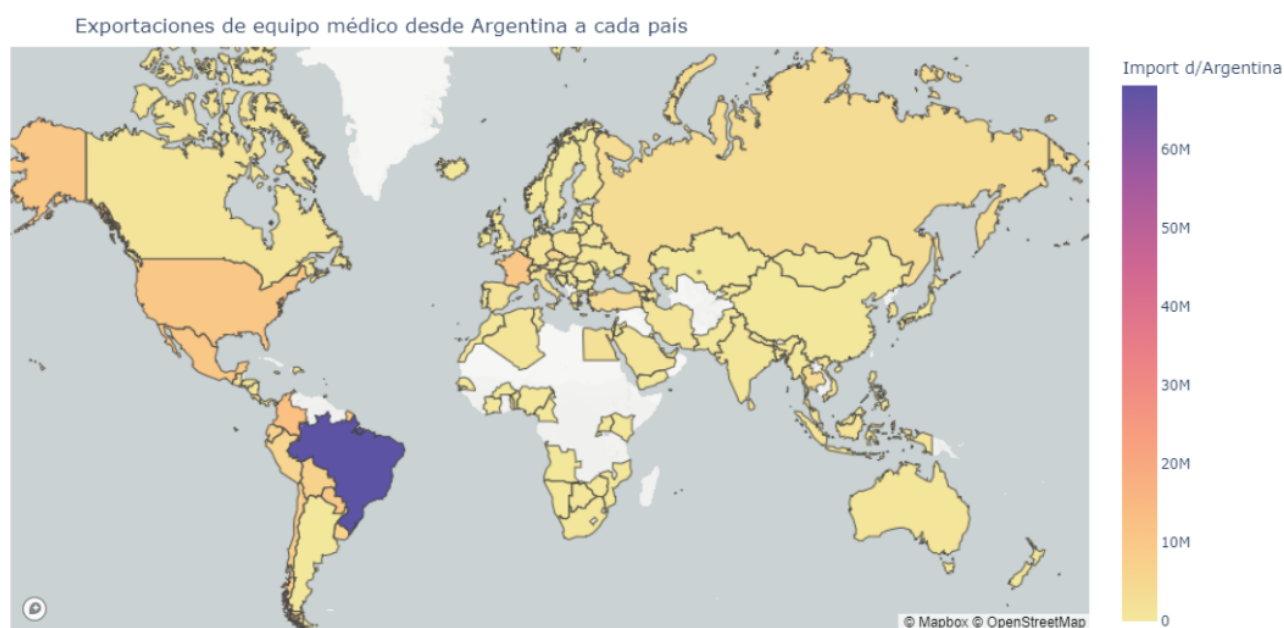
Se corrigieron problemas relacionados con los códigos ISO (utilizando como control los códigos provistos por el Banco Mundial), y además se agregó una variable de distancia de cada país a Argentina. Además, se abandonaron en primera instancia todas las variables que contaran con más del 30% de los datos nulos, con la excepción de la recaudación por impuestos al comercio exterior por considerarla potencialmente relevante por sí misma. Posteriormente se unificaron ambas bases en función del código ISO y, como las variables se encuentran en unidades muy distintas y de rango radicalmente diferente, se las escaló para hacerlas comparables. Además de lo anterior se corrigieron algunas variables que se encontraban en un tipo de dato distinto del correcto y se removieron duplicados.

Dentro de cada categoría de variables, se encontraban algunas que eran muy similares entre sí en términos conceptuales y que por ende aportan información relativamente equivalente. En general, aunque no siempre, estas variables se encuentran también altamente correlacionadas (se mueven en dirección y magnitud similares en el mismo momento del tiempo). Por ende, algunas de las 51 variables con que se contaba eran esencialmente redundantes (e.g. las variables de desempeño logístico poseían entre ellas correlaciones siempre mayores a 0.9), por lo que haciendo un análisis de correlación entre las variables por categoría (de manera de asegurarse cierta similitud conceptual) se descartaron aquellas variables que poseyeran una correlación mayor al 85% con otra de mayor relevancia según nuestro juicio. Finalmente, se trabajó con 32 variables (se descartaron 19), una reducción de casi el 40% sobre la cantidad inicial.

Posteriormente se unificaron ambas bases en función del código ISO y, como las variables se encuentran en unidades muy distintas y de rango radicalmente diferente, se las escaló para hacerlas comparables. Además de lo anterior se corrigieron algunas variables que se encontraban en un tipo de dato distinto del correcto y se removieron duplicados.

Considerando que lo que se busca es una categorización de las exportaciones argentinas, fue necesario crear categorías para esta variable. Se decidió crear 6 categorías para los países en función de sus importaciones de equipamiento médico. Las líneas divisorias de cada categoría son los cuartiles de la distribución de sus importaciones desde Argentina, con dos salvedades: la primera es que se creó una etiqueta exclusiva para Brasil por ser un outlier muy grande; la segunda es que se eliminó del cálculo de los cuartiles a los países que no importaban desde Argentina y se les asignó a estos una categoría propia (0Q). Por ende se crearon 6 etiquetas para los países inicialmente.

## Gráfico 1



Un análisis exploratorio demostró que existe gran variabilidad entre los países para cada característica, pero fundamentalmente para la más relevante: las importaciones de equipamiento médico desde Argentina. Existen outliers muy grandes (el caso de Estados Unidos como importador por ejemplo) y en el caso de las importaciones argentinas, como se ve en el Gráfico 1, Brasil constituye el principal destino de las exportaciones de equipamiento médico por una amplia diferencia. Otro resultado relevante del análisis exploratorio es que los demás países se encuentran muy cercanos y hacia la parte baja de la distribución en volumen de importaciones desde Argentina, siendo países del continente americano los principales destinos.



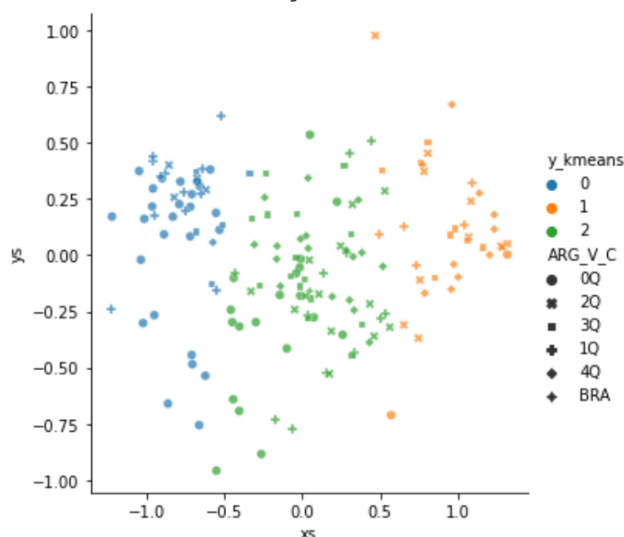
# Análisis de Cluster

Contando con la clasificación y con más de 30 variables explicativas, las técnicas de categorización partieron primeramente de una reducción de dimensionalidad. Para ello se usó el método de Componentes Principales, que posteriormente fue el insumo para la determinación de los clusters. En lugar de elegir variables relevantes *a priori*, la reducción de dimensionalidad nos permite capturar el espectro de variación de las variables. Esto significa que se intenta abarcar la variabilidad total de nuestras variables en solo unas pocas variables “ficticias”, llamadas componentes principales.

Luego, la clusterización se consideró el método principal más útil para categorizar los países. Con el análisis de cluster intentamos encontrar países que, por su características, son similares entre sí y verificar si existe algún patrón que se ajuste a las exportaciones argentinas en relación a su similitud. Es de esperar que los países a los que Argentina exporte más sean entre ellos similares, pero no necesariamente sabemos en qué. Entonces podemos obtener la clasificación de los países primero, para luego comparar algunas variables seleccionadas.

En función de la capacidad de la clusterización de asignar correctamente las etiquetas a los países, **la mejor clasificación se logra con tres clusters sobre la base completa de 32 variables**, previamente reducidas en dimensionalidad, y es presentada en el Gráfico 2.

**Gráfico 2**



Se probó con otras condiciones para la clusterización (eliminando variables que presentan multicolinealidad o eliminando los países a los que Argentina no exporta) pero no se logra una mejor clasificación para compensar el costo de eliminar observaciones. En el Gráfico 2 puede verse cómo se realiza el agrupamiento de los países por cercanía (en distancia euclídea). Es evidente que no existen clusters simples, nítidamente separados de los demás, por lo tanto y como era de esperar no es posible decir que los países a los que Argentina exporta equipamiento médico puedan ser fácilmente separados en función de algunas variables generales que los caracterizan.

La Tabla 1 provee información resumen de cada cluster. Para los valores numéricos, provee el promedio de todos los países, mientras que para los valores cualitativos provee la moda de las observaciones. Puede verse que los clusters logran una clasificación bastante aceptable de las características de los países. En el primero de ellos, tenemos países de ingresos bajos, mayormente africanos, con economías relativamente cerradas y bajos gastos en salud. En el Cluster 0 estaríamos en presencia de los países de altos ingresos, mayormente europeos, que son quienes más gastan en salud y presentan mayor apertura comercial. Finalmente, el cluster 2 contiene países llamados de ingreso medio, mayormente americanos, que son en casi todos los aspectos un caso intermedio entre los dos extremos anteriores, a juzgar por las medidas resumen presentadas

Tabla 1

	Cluster 1	Cluster 2	Cluster 3
Continente	África	Europa	América
Distancia (ARG)	10777.82419	12398.20673	10433.90768
Población	32,291,014.56	27,856,994.53	58,949,746.35
PIB per cápita	\$1,717.78	\$45,227.63	\$10,914.52
Balanza Comercial	-\$3,497,239,434.04	\$13,100,573,506.35	\$3,632,496,492.41
Importaciones (% PIB)	39.23%	59.18%	47.04%
Impuestos al Comercio (% Recaudación)	11.97%	0.79%	6.43%
Gastos de bolsillo en salud (PPP)	\$73.34	\$799.77	\$310.54
Gasto corriente en salud per cápita	\$85.86	\$3,987.65	\$545.23

Una característica destacable que se desprende de la tabla anterior es la variable que fue añadida por nosotros: la distancia geográfica. Para todas las otras variables numéricas, se cumple que estas crecen desde los países de ingresos bajos hasta los de ingresos altos, es decir que el PIB aumenta, también las importaciones, etc. Esto no es cierto para la distancia, los países del cluster 2 (ingresos medios) son los más cercanos geográficamente a la Argentina, mientras que los del 1 son los más distantes. Naturalmente esto tiene sentido considerando que Latinoamérica es una región de ingresos medios, por lo que es de esperar que el cluster 2 tenga países cercanos geográficamente. Sin embargo es un resultado interesante a la luz de lo siguiente.

Tabla 2

	Cluster 1	Cluster 2	Cluster 3	TOTAL
Sin exportaciones	26	12	19	57
Cuartil 1	12	10	18	40
Cuartil 2	5	8	15	28
Cuartil 3	4	5	13	22
Cuartil 4	1	3	12	16
Brasil	0	0	1	1
TOTAL	48	38	78	164

La Tabla 2 muestra a qué cuartil de las exportaciones de equipamiento médico argentinas pertenecen las observaciones asignadas a cada cluster. Lo primero que es destacable es que el cluster 2 es notablemente más grande que los demás (50% del total de observaciones aprox.). Lo segundo es que todos los cluster poseen mayoría de países a los que Argentina no exporta, es decir que algunos de estos países constituyen probablemente casos donde se pueden expandir mercados, sobre todo aquellos que pertenecen al cluster 2, como se verá más adelante.

El cluster 2 contiene el grueso de los países del cuartil 4 (75%) y la mitad del cuartil 3 (59%). Si bien es cierto que es el cluster más grande, estos porcentajes son mayores que para los demás cuartiles. Además, el cluster 2 también contiene a Brasil, con mucho el mayor destino de las exportaciones argentinas analizadas. Por ende, podemos concluir que el tercer cluster caracteriza al menos tentativamente los destinos principales de las exportaciones de equipo médico argentinas. Esto es, son países de ingresos medios, relativamente abiertos al comercio internacional, con gastos en salud moderados y cercanos geográficamente a la Argentina.

## Algoritmos de clasificación

Realizada la clusterización se avanzó hacia otros métodos de clasificación.<sup>2</sup> Los algoritmos de clasificación intentan, por diversos mecanismos, asignar a cada observación una etiqueta de característica, en nuestro caso las etiquetas de cuartiles. Un cambio que se realizó respecto del análisis de cluster es que se incluyó a Brasil en el cuartil 4, quedando entonces con 5 etiquetas posibles. La razón es que, a diferencia de los clusters que no comparan las etiquetas sino las variables, los algoritmos de clasificación si hacen uso de la etiqueta para “aprender”, por ende mantener una etiqueta con una sola observación es imposible. Para verificar además que el aprendizaje sea predictivo, se separa la base en dos: una para entrenar los algoritmos y una para testar su éxito.

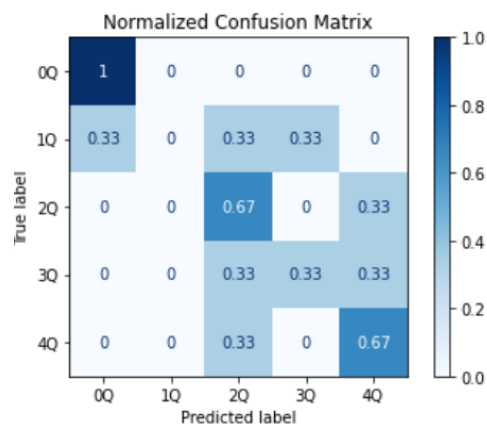
Se usaron 4 algoritmos de clasificación: K vecinos más cercanos, árbol de decisión, regresión logística y Random Forest. No es posible detenerse en las particularidades de cada uno, simplemente nos abocaremos a los que mejores resultados proveyeron. Random Forest alcanzó el mayor nivel de éxito, con un accuracy del 57%. Esto significa que provisto el conjunto de test, el algoritmo es capaz de clasificar correctamente el 57% de las observaciones. El Gráfico 3 muestra la matriz de confusión para el algoritmo, ella visualiza las clasificaciones correctas e incorrectas para cada una de las observaciones con la que se testeó la capacidad predictiva del modelo. Puede verse que la etiqueta del tercer cuartil es la más difícil de predecir correctamente por ejemplo. Los demás algoritmos lograron accuracies, en orden decreciente de 48% para el árbol de decisión, 42% para la regresión logística y 30% para k-vecinos, claramente menores al Random Forest.

Random Forest es un algoritmo que esencialmente repite, un número predeterminado de veces, el análisis de los árboles de decisión y devuelve para cada observación la etiqueta predicha por el mayor número de árboles. Esto implica que no es simple observar qué variables son las que determinan la clasificación, pues en realidad es un conteo de resultados de caminos distintos. Entonces, podemos usar el algoritmo de Random Forest

<sup>2</sup> Se intentó también un análisis factorial exploratorio para descubrir variables relevantes. Sin embargo, se prefirió utilizar clusters.

para **predecir qué países Argentina puede exportar equipamiento médico en mayor cantidad, pero no observar fácilmente qué caracteriza a estos países.**

Gráfico 3



La Tabla 3 resume este orden para las exportaciones argentinas. De ella se desprende que, en función de su valor de importaciones totales, **Arabia Saudí, Sudáfrica y Grecia constituyen destinos potenciales de las exportaciones argentinas aún no explotados al máximo.** El algoritmo predice que estos países deberían pertenecer al cuartil 4, de mayor importaciones desde Argentina pero no lo hacen y juzgando por que poseen un valor total de importaciones importante, constituyen potenciales destinos relevantes.<sup>3</sup>

Tabla 3

	Valor	Dist	Cuartil Real	Cuartil Predicho	Salto
Arabia Saudi	\$6,209,070,769.00	13202.67	3.00	4.00	1.00
Sudafrica	\$3,701,103,417.00	8106.38	3.00	4.00	1.00
Grecia	\$2,690,822,573.00	11981.50	2.00	4.00	2.00
Argentina	\$2,146,707,859.00	0.00	2.00	4.00	2.00
Croacia	\$1,478,260,202.00	11864.29	1.00	2.00	1.00
Eslovenia	\$888,736,027.00	11879.35	1.00	2.00	1.00
Myanmar	\$444,653,565.00	17812.06	1.00	3.00	2.00
Kyrgyzstan	\$136,965,874.00	16350.41	1.00	3.00	2.00

En un orden menor, Croacia y Eslovenia pueden ser de un interés secundario, países a los que Argentina podría exportar algo más de lo que actualmente hace. Más extraño es el caso de Myanmar y Kirguistán que el algoritmo predice que deberían pertenecer al tercer cuartil y por ende se les podría exportar mucho más, pero tienen valores totales de importaciones bajos además de encontrarse lejanos y ser relativamente pobres.

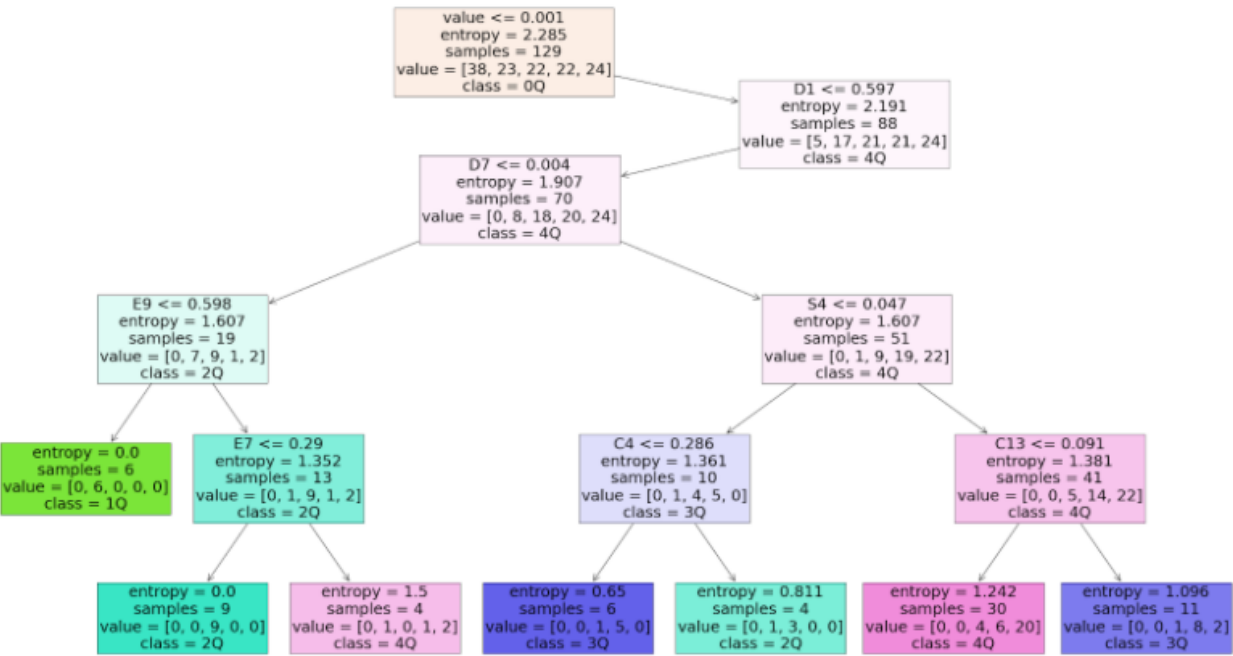
Como el Random Forest no permite observar el mecanismo de clasificación, mostramos el árbol de decisión que construimos para ejemplificar el análisis en el Gráfico 4, recortado solo para llegar a las clasificaciones del cuartil 4 por razones expositivas. El árbol usa las variables diversas para clasificar las observaciones según su etiqueta de pertenencia a cada cuartil de las exportaciones argentinas de equipamiento médico, pero esencialmente

3 La presencia de Argentina en la lista de la Tabla 3 puede ser desconcertante. Sin embargo hay dos opciones: una es que la mera existencia en la base sea un error; la otra es que la contabilidad del comercio internacional esté considerando el componente argentino de las exportaciones de otros países hacia la Argentina. Como no podemos saberlo con la información provista originalmente, decidimos mantener la observación pero ignorarla en materia de conclusiones.



observa el valor de la importaciones (value), el crecimiento poblacional (D1), la población total (D7), la balanza comercial (E9), muertes por enfermedades transmisibles (S4), el índice de desempleo (E7) y la recaudación por impuestos al comercio exterior (C13) para clasificar a los países en el cuartil 4.

Gráfico 4



## Conclusión

El objetivo planteado era caracterizar los países a los que Argentina exporta equipamiento médico y posteriormente rankear qué países pueden ser destinos no explotados. El análisis de cluster nos permitió alcanzar una caracterización donde los países a los que **Argentina exporta preferentemente son países de ingresos medios (con las características usualmente asociadas) y geográficamente cercanos**, siendo los países ricos un segundo grupo de menor relevancia. Por otro lado, se pudo rankear a los destinos potenciales muy tentativamente utilizando las predicciones de un algoritmo Random Forest, donde **tres países parecen ser las mejores opciones: Arabia Saudi, Sudáfrica y Grecia**.

Una línea de profundización abierta es rankear dentro de cada cuartil los países. Hay diferencias sustanciales entre los elementos de ellos, en particular en el más alto al que pertenece Brasil, por lo que **existe espacio para aumentar las exportaciones intracuartil y no solo en aquellos casos donde se exporta poco a un país que se predice pertenece al cuartil más alto**. Esto permitiría obviar las predicciones algo extrañas del ranqueo intercuartil, donde se sugeriría exportar a países poco usuales y de baja relevancia en el comercio internacional. También puede usarse el “camino” provisto por los árboles de decisión para identificar variables de mayor relevancia y determinar países objetivo en los que centrarse en la promoción exportadora.