

EL USO DE DATOS PARA CARACTERIZAR LA POBLACIÓN OBJETIVO DE UNA POLÍTICA PÚBLICA: EL CASO DEL REGISTRO FEDERAL DE CULTURA

Paula Luvini

Científica de Datos, FUNDAR.

1. Introducción

Toda política pública tiene una población objetivo. A menudo, necesitamos distinguir en subgrupos dentro de esta población objetivo para conocerla mejor y diseñar políticas más precisas y mejor orientadas hacia estos subgrupos.

En la literatura, existe un creciente consenso en la necesidad de distinguir entre subgrupos para hacer que las políticas sean más efectivas. Una misma política pública podría tener resultados muy diversos cuando se aplica a distintas personas. Poder distinguir entre subgrupos nos ayuda a identificar mejor qué tipo de intervención genera algún tipo de efecto según el grupo al que se le aplique. Para ello, podemos dividir a la población objetivo en diversos subgrupos basados en diferencias demográficas, geográficas, de situación familiar, ocupacional, entre otras.

Evaluar los efectos de una intervención entre distintos subgrupos (en lugar del impacto promedio de toda la población) permite evaluar la posibilidad de que una misma intervención tenga un impacto diferente entre estos grupos y conocer mejor una población objetivo que puede ser muy diversa (Beecroft y Lee 2000).

Identificar subgrupos puede ser necesario en las evaluaciones de impacto para identificar efectos heterogéneos en el tratamiento. Una misma intervención podría tener un impacto diferenciado entre subgrupos (Bloom y Michaloupulos 2010), la probabilidad de participar de un programa podría diferir entre subgrupos (Gibson 2003).

Por ejemplo, Gibson (2003) utiliza las técnicas de *clustering* y *propensity score matching* para identificar distintos subgrupos dentro de la población objetivo de una política de asistencia estatal a sectores necesitados encontrando tasas de impacto muy distintas entre subgrupos. Peck (2003) identifica distintas respuestas entre subgrupos beneficiarios del *New York State's Child's Assistance Program*, un programa estatal destinado incorporar al mercado laboral a madres solteras que dependían de la ayuda estatal. A través de este análisis puede identificar un mayor impacto del programa en el subgrupo de madres mayormente afroamericanas, de edad más avanzada y mayor trayectoria en el mercado laboral que entre el subgrupo de madres, en mayor proporción hispanas, de menor edad y menor trayectoria en el mercado laboral.

Pero la necesidad de dividir en subgrupos no es únicamente en función de identificar efectos causales heterogéneos en el caso de estudios de impacto. También es necesario poder conocer la población objetivo de una política pública y para poder conocerla es necesario a menudo poder caracterizar subgrupos. Estos subgrupos pueden tener distintas características, intereses y necesidades. En contextos críticos como la emergencia de la pandemia, la restricción de actividades se vuelve necesario dirigir las políticas de asistencia a los sectores que tienen mayor necesidad.

La identificación de subgrupos con características comunes permite identificar patrones dentro de una población heterogénea. En el caso de la población objetivo del Ministerio de Cultura, existen perfiles tan diversos como el de escenógrafos/as, artesanos/as, músicos/as, docentes, etc. Existe naturalmente también un amplio espectro de destinatarios/as en términos geográficos, de situación laboral, etarios.

Todo esto hace que la segmentación en diversos perfiles pueda ayudar a agregar precisión a los objetivos que se proponga cualquier política pública para el sector. Se trata de un uso de datos distinto al de las evaluaciones de impacto de políticas públicas al utilizar datos y evidencia para conocer el público receptor de una política. Esto nos ayuda a identificar las características y necesidades de diversos grupos y así hacer un uso más eficiente de los recursos públicos. Podemos, por ejemplo, identificar a aquellos grupos que han recibido más apoyo en las políticas públicas y a aquellos que han recibido menor atención. También podemos conocer la localización geográfica, rama de actividad y situación ocupacional de los diversos subgrupos y así diseñar políticas más precisas y eficientes para favorecerlos.

3. Primer paso: Análisis exploratorio, evaluar la calidad de datos

Anonimización de datos

Un paso importante a la hora de iniciar el intercambio de información consistió en la anonimización de los datos existentes. Para ello se identificaron las variables de interés que se necesitaban del RFC así como los posibles riesgos de cruzamiento con otras bases existentes. De esta manera se hicieron una serie de cambios en pos de preservar los datos personales de los miembros del registro. La anonimización que se llevó a cabo fue pensada para intercambiar de manera segura los datos entre Fundar y el Ministerio de Cultura, considerando que la exhaustividad de la anonimización necesaria para que esta base sea completamente pública es mayor. Se eliminaron entonces variables personales y se modificaron o limpiaron otras.

Este procedimiento se hizo teniendo en cuenta las variables que se necesitaban para llevar a cabo la caracterización de perfiles. Así se eliminaron los nombres y apellidos, direcciones de email, teléfonos y CUIL. El CUIL no era necesario para llevar a cabo la caracterización pero era útil para obtener la edad estimada de los inscriptos de manera que antes de eliminarlo se estimó y agregó esta información. Otras variables personales que también se eliminaron fueron todos los usuarios de redes sociales, reemplazándolos por una variable binaria en los casos en que la persona contaba con redes sociales.

Para realizar las modificaciones a las variables de respuesta abierta, se utilizó la librería de Python *scrubadub* para remover datos personales que los participantes pueden haber escrito en las respuestas de campo abierto. Las demás modificaciones fueron hechas en código de Python nativo.

Aunque la base no fue compartida públicamente, llevar a cabo este ejercicio de preservación de datos sensibles es una práctica deseable incluso en ambientes de confidencialidad. En este caso particular, analizamos las pérdidas por las cuales anonimizar de manera más estricta más campos o variables trae como consecuencia la pérdida de información y por ende perjudica el análisis posterior. Esto mismo no quita importancia al punto de que, más allá de que vamos a tratar la información confidencialmente, no debemos vulnerar la identidad de los involucrados.

Evaluar la calidad de datos

En primer lugar y siempre que comenzamos cualquier análisis es necesario **revisar los datos** que tenemos de manera de **conocerlos y evaluar su calidad**. En nuestro caso, al tener como insumo respuestas de tipo declarativo de los participantes del RFC es posible que las mismas tengan inconsistencias o errores. El Registro Federal de Cultura tiene un total de 81 variables de las cuales 52 corresponden a preguntas respondidas al momento de la inscripción (variables declarativas) y 29 que constituyen cruzamientos con otras bases internas del Ministerio. La posibilidad de contar con cruzamientos con otras bases disponibles es algo deseable y cuya práctica debería extenderse. En este sentido, la apertura de datos de parte del sector público permite un mayor aprovechamiento cuando una base individual se abre, ya que el cruzamiento con otras bases permite generar nuevas preguntas y análisis más ricos para el diseño y evaluación de políticas públicas.

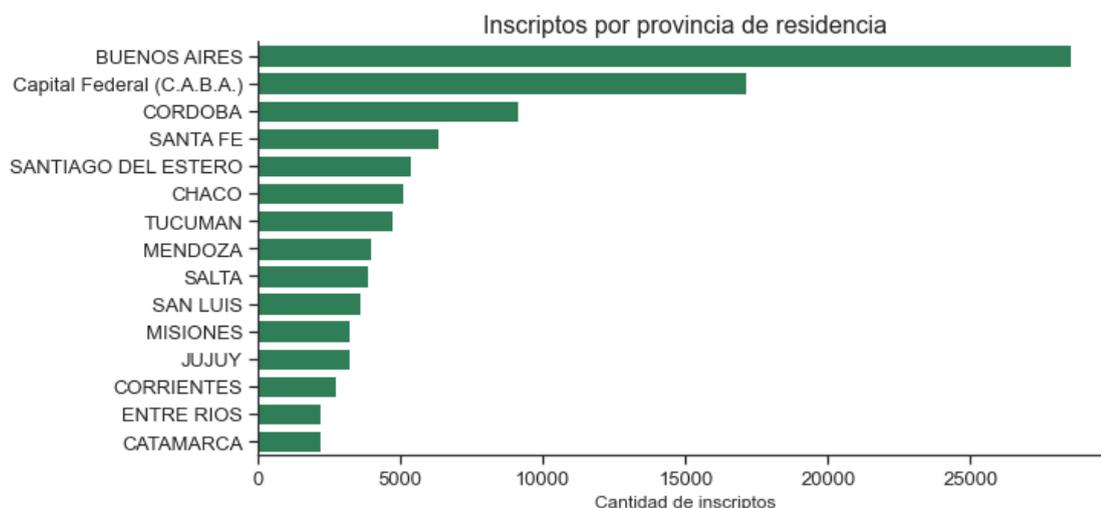
De esta manera, se empezó revisando el formato de las variables y si se tenían valores nulos o incompletos. Se encontraron algunas variables con campos sin información, cosa que se debe revisar particularmente en aquellas preguntas de respuesta obligatoria. También se hizo una evaluación de cosas a mejorar en la interfaz de carga y que son potenciales problemas futuros como la inclusión de más listas desplegables para completar donde sea posible o la exigencia de un formato específico en la respuesta a ciertas preguntas, por ejemplo en las que se espera un número con una cantidad de caracteres preestablecidos como son las fechas.

Análisis exploratorio

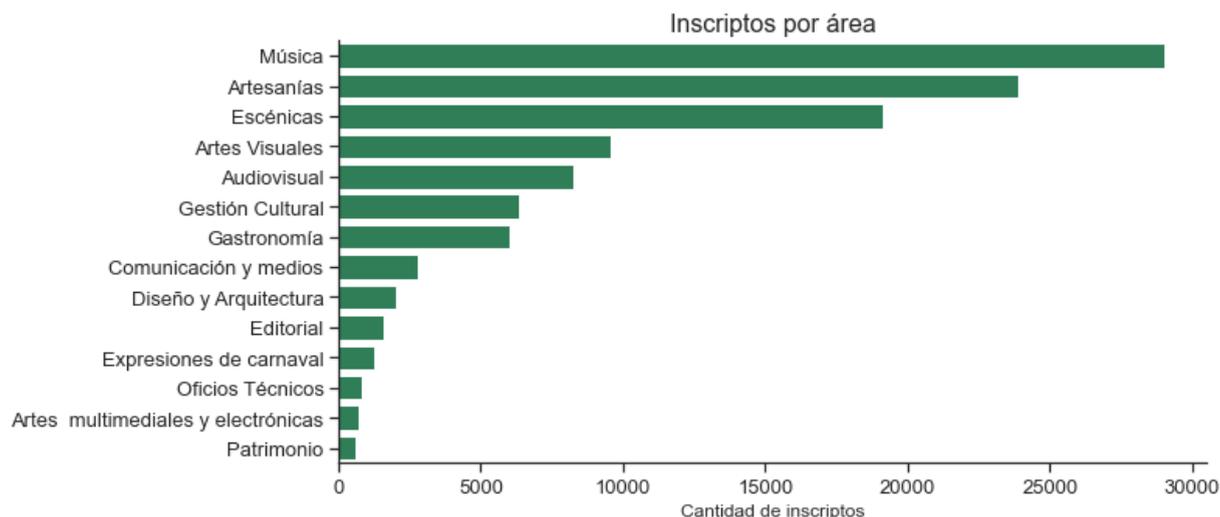
A la hora de iniciar un trabajo con datos necesitamos llevar a cabo una **exploración** de los mismos para obtener distribuciones, métricas de resumen y otras herramientas que permitan identificar patrones o anomalías en los datos. Además, el análisis exploratorio es un insumo para saber si con estos datos podemos responder las preguntas establecidas al inicio del proyecto. Para ello se hace uso de algunas herramientas gráficas y descriptivas, de las cuales algunos resultados se enumeran a continuación.

La base está constituida por inscriptos que se inscribieron al registro entre Marzo (mes de inicio del mismo) y Noviembre de 2021. El 75% de los inscriptos que completaron su alta lo hicieron en el mes de Marzo de 2021. En los cuatro meses subsiguientes la cantidad de inscriptos rondó el 0,95% de inscriptos por mes, tendencia que cambió a partir de agosto de ese año donde se incrementaron los formularios recibidos.

Un 49.5% de los inscriptos en el RFC **identificaron su género** como mujeres y un 47.9% como hombres. Otras identidades también incluidas en estas preguntas son no binarias, varones trans y mujeres trans. Además, los participantes contaban con una opción para aquellos que preferían no decirlo (donde contestaron un 1% del total) y una respuesta abierta en caso de que su identidad no perteneciera a ninguna de las anteriores ([Ver Tabla 1](#)). En cuanto a sus países de origen, el 97% de los inscriptos es de **nacionalidad argentina**. Las siguientes nacionalidades más frecuentes son la chilena, venezolana, uruguaya, colombiana y paraguaya ([Ver Tabla 2](#)). En términos de la edad de los participantes, vemos que es más frecuente la participación de **jóvenes y adultos jóvenes** ([Ver Gráfico 1](#)). Las cuatro **provincias** con más inscriptos son Buenos Aires, Capital Federal, Córdoba y Santa Fe. Buenos Aires y Capital Federal tienen una cantidad de inscriptos que contribuye al 41% del total ([Ver Tabla 3](#)).

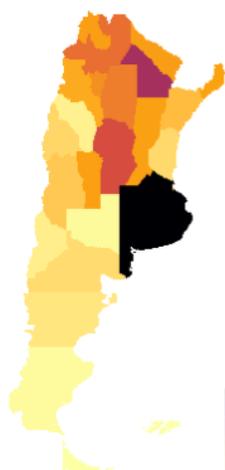


En términos de inscriptos por **área cultural** Música es la más popular, seguida de Artesanías y Escénicas. Estas tres áreas tienen particularmente muchos más inscriptos que el resto.



En la mayoría de las provincias las áreas más populares son Artesanías o Música, a excepción de Capital Federal, Provincia de Buenos Aires, Mendoza y Santa Fe, donde las artes Escénicas integran el segundo lugar más recurrente. En las Artesanías observamos que hay más diversidad de provincias y están más atomizadas que en las otras dos, habiendo una gran cantidad de provincias del Norte con participación.

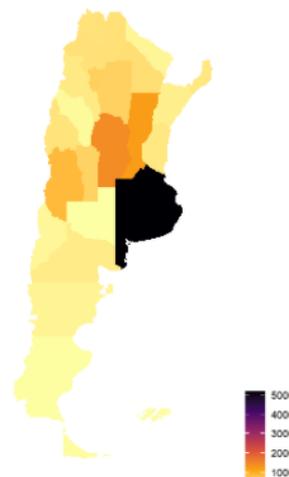
Artesanías



Música



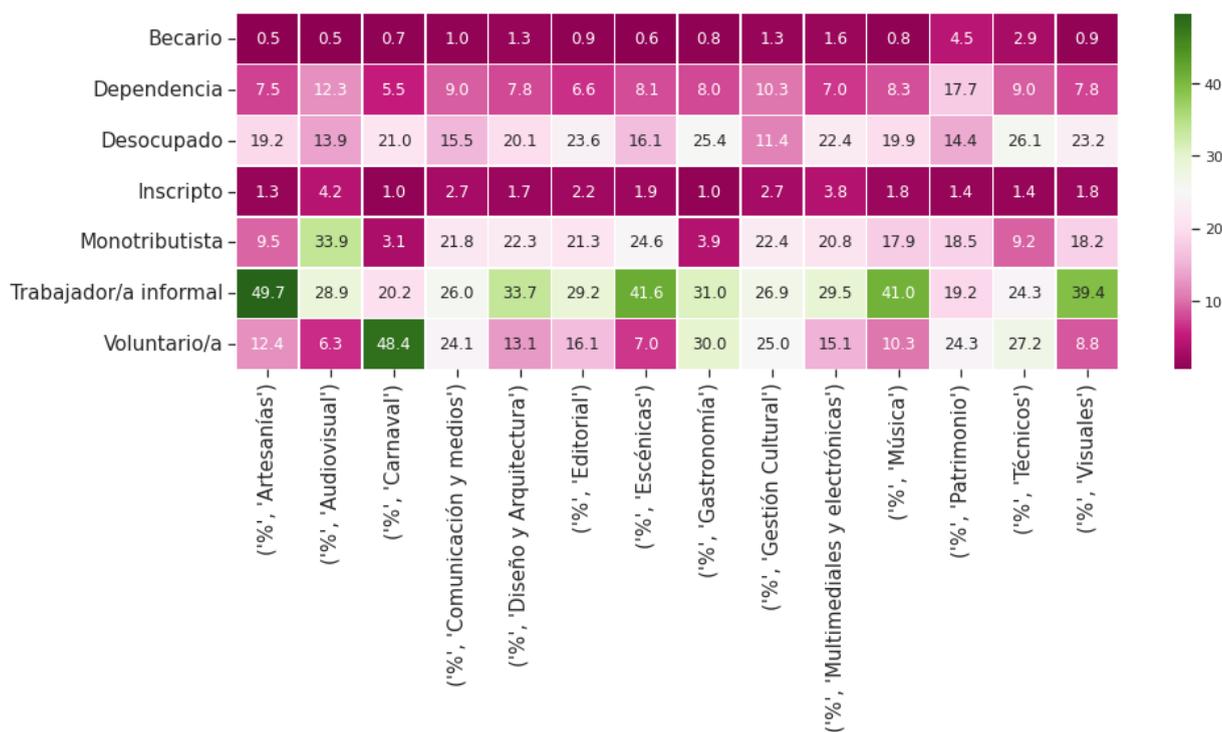
Escénicas



La **relación laboral** más común es la de los trabajadores informales, que son el 39% del total, seguidos por desocupados, monotributistas y voluntarios ([Ver Tabla 4](#)). Los empleados en relación de dependencia son menos del 10% del total de inscriptos. En el porcentaje de ingresos recibidos por la actividad cultural, hay una distribución bastante bimodal en los mismos. Aquellos que reciben entre un 0 y un 25% de **ingresos por la actividad cultural** son los más frecuentes, mientras que en segundo lugar están los que reciben más del 70% de ingresos de la actividad cultural, los ingresos más altos. Los terceros más frecuentes son aquellos que no reportan ingresos por la actividad: un 17% que está constituido mayormente por los desocupados ([Ver Tabla 5](#)).

Observando los cruzamientos entre las dos variables anteriores se encuentra que aquellos que están desocupados son los que tienen como ingreso “No corresponde”. Por otro lado, un 8.2% de los inscriptos es Monotributista y gana más del 70% por la actividad cultural, así como un 10.67% es trabajador informal y tiene esos ingresos. Es decir, que **los altos ingresos son mayormente monotributistas y trabajadores informales**. Por otro lado, los ingresos de 0 a 25% se explican por trabajadores informales y por voluntarios ([Ver Tabla 6](#)).

Cruzando ahora estas variables laborales con las áreas culturales a las que pertenecen los inscriptos, se puede observar que los **Trabajadores informales son la categoría más repetida principalmente por las áreas de Música, Artesanías y Escénicas**. En Artesanías, los trabajadores informales son los más importantes, representando un 50% del total. El área Audiovisual por otro lado tiene más trabajadores Monotributistas que informales y proporcionalmente más que cualquier otra categoría. Asimismo, el área con más desocupados es la de Técnicos, seguidos por los Gastronómicos, Editorial y Visuales. Los becarios son más frecuentes en el área de Patrimonio y los voluntarios son más frecuentes en el área de Carnaval.



En las dos áreas más populares encontramos una diferencia bastante marcada en las **edades de los inscriptos**. Aún cuando los inscriptos en Música eran más que los inscriptos en Artesanías, podemos notar que los primeros tienen más presencia en los más jóvenes, mientras que en los artesanos predominan aquellos nacidos antes de 1975 ([Ver Gráfico 3](#)).

El 40% de los participantes forma parte de algún **Registro**, siendo los más populares el Registro Nacional del Fondo Nacional de las Artes, el Registro Nacional de la Música, el Registro Nacional del Teatro Independiente y el Registro Nacional de Trabajadores y Trabajadoras de la Economía Popular ([Ver Tabla 7](#)).

4. Resultados del clustering.

Para seleccionar la cantidad de clusters a encontrar se utilizó el método del codo descrito en el apéndice metodológico. Así se encontraron 6 clusters con las siguientes características:

- Cluster 1: Es el cluster más numeroso, contando con el 30% del total de participantes. Provincias más repetidas: Buenos Aires, Capital Federal (C.A.B.A.), Santiago Del Estero, Chaco, Santa Fe. Varón y mujer más balanceados que en otros casos. No corresponden ingresos (56.3%) y desocupados constituyen el (59%) del grupo. Áreas más repetidas: Música, Artesanías, Visuales. 83.3% no inscriptos en ningún registro.
- Cluster 2 (18.1% de todos los inscriptos): Provincias más repetidas: Capital Federal (C.A.B.A.), Buenos Aires, Córdoba, Santa Fe. El 87% de este grupo son mujeres y el 9.7% varones. Más del 70% de los ingresos, trabajadores informales seguidos de monotributistas. Áreas más repetidas: Escénicas, seguidas de Artesanías y Música. Inscriptos a registros (82%).
- Cluster 3: (16% del total) Provincias más repetidas: Buenos Aires, Córdoba, Capital Federal (C.A.B.A.). El 92% son varones y el 5% son mujeres. Ingresos repartidos

en: Más del 70%, 0 a 25 %, 26 a 50%, 51 a 70%. El 63% trabajador informal y el 15% monotributistas. Inscriptos a registros (86%)

- Cluster 4: El segundo cluster más numeroso, cuenta con un 18.4% de los datos utilizados. Provincias más repetidas: Buenos Aires, Capital Federal (C.A.B.A.), Santa Fe. 83% mujeres, 15% varones. Este cluster es el que concentra ingresos culturales más bajos, con un 72% de sus integrantes teniendo ingresos entre 0 y 25% y un 14% entre 26% y 50%. Si bien el grupo más importante son los trabajadores informales aquí hay una gran participación de voluntarios, en porcentaje el cluster con la mayor cantidad de ellos. La mayoría (95%) no está inscripto en los registros. Este cluster está más repartido y menos atomizado en sus áreas: Escénicas, Visuales, Gastronomía, Artesanías, Gestión Cultural, Música, Audiovisual
- Cluster 5: Este es el cluster más pequeño, con solo un 6.6% de los datos (7440 participantes). Provincias más repetidas: Capital Federal (C.A.B.A.), Córdoba, Buenos Aires. 95% varones y 0 mujeres. El 58% de este cluster cuenta con ingresos mayores al 70% por la actividad cultural: son en su mayoría monotributistas (83%). Áreas más repetidas: Escénicas, Audiovisual, Música Hay un 72% de inscriptos en registros.
- Cluster 6 (11.9% del total de inscriptos): Provincias más repetidas: Córdoba, Chaco, Santiago Del Estero, Salta, Formosa. 75% mujeres y 23% varones. El área más grande con un 85% es Artesanías. Un 60% de trabajadores informales y un 18% de voluntarios de ingresos bajos por la actividad cultural (pero ingresos al fin, no hay no corresponde). No inscriptos a registros son la mayoría

6. Apéndice metodológico

El **aprendizaje no supervisado** utiliza algoritmos de machine learning para agrupar datos. El objetivo de estos algoritmos es descubrir patrones en los datos para agruparlos sin la necesidad de setear más que algunos parámetros correspondientes a cada modelo.

Llevar a cabo un clustering implica dividir nuestros datos en grupos (llamados clusters) que están conformados por objetos que son más similares a sus pares de cluster que a los de los demás grupos, dada una medida elegida previamente. La conformación de estos grupos va a depender de las variables que elijamos como inputs de nuestro modelo. Si incluimos variables como la edad, el género, trabajo o localización de individuos, el algoritmo leerá todo esto y buscará patrones transversales a estas variables.

Algoritmos de particiones

Dentro del aprendizaje no supervisado hay muchos tipos de metodologías que se pueden utilizar: con algoritmos de particiones, mixturas, jerárquicos, de densidad, espectrales, entre otros. En el presente trabajo hemos utilizado algoritmos de partición, los cuales van a determinar dado un conjunto de observaciones en un espacio una partición de los datos en K grupos de manera tal que las observaciones dentro de cada partición son similares entre sí y diferentes a las de otras particiones.

De esto último se desprende que necesitamos un **criterio para evaluar qué significa que dos observaciones sean similares entre sí**. Un criterio posible es aquel usado en K Medias: **distancias**. De esta manera, observaciones que se encuentren cerca unas de otras van a considerarse similares, mientras que aquellas que estén más alejadas en el espacio van a considerarse distintas. Este criterio de distancias entonces va a encontrar

las particiones que minimicen la dispersión dentro de los grupos y maximice la dispersión entre grupos.

K Medias es uno de los algoritmos partitivos de aprendizaje no supervisado más usados para hallar grupos dentro de datos continuos y descubrir patrones en ellos. El número de grupos o k que encontrará es definido de antemano e indicado al algoritmo. Con ello, K Medias identificará k centroides en los datos y a partir de ellos clasificará los demás puntos dentro de los grupos correspondientes. Esto lo hará colocando a cada punto en el cluster más cercano, es decir minimizando la distancia de los puntos con los centroides en una serie de iteraciones hasta encontrar la mejor configuración. Para realizar el agrupamiento, K Medias considera como métrica de distancia entre puntos a la distancia Euclídea al cuadrado.

K Modas

En el caso de contar con variables categóricas, utilizar la distancia Euclídea ya no es posible sin binarizar de alguna manera las variables. Otra opción es utilizar un algoritmo de partición que tome como función de evaluación alguna específica para este tipo de datos: K Modas. Este reemplaza las medias de los clusters con modas, utiliza la frecuencia de las variables para la asignación a los clusters y como función de costos toma una medida de disimilaridad. La moda es el valor que tiene mayor frecuencia absoluta en una categoría determinada. Específicamente, los pasos del algoritmo son los siguientes:

1. Se seleccionan aleatoriamente las k modas iniciales que van a ser los centroides de los cluster.
2. Utilizando la función de disimilaridad se adjudica cada punto a un cluster y luego se actualiza su moda (centroide).
3. Se vuelven a evaluar los puntos según la función de disimilaridad y los nuevos centroides y si se encuentra que son más cercanos a otro cluster se reasignan.
4. Esto se repite hasta que ninguno de los puntos cambia de cluster.

A la hora de evaluar políticas públicas o analizar datos de individuos, en la mayoría de los casos vamos a contar con variables categóricas por lo que contar con una herramienta específica para este tipo de datos es importante. Además, K Modas cuenta la ventaja de que es eficiente y puede procesar rápidamente un gran volumen de datos, a diferencia de otros métodos que también pueden procesar datos categóricos, como los jerárquicos.

¿Cómo encontrar el número de particiones?

Los algoritmos de K Medias y de K Modas necesitan que les indiquemos la cantidad de clusters que deseamos encontrar. Aunque el número indicado no sea el correcto, los algoritmos se forzarán a encontrar esos k clusters. Por esto es importante hallar el número correcto. El método del codo (o elbow) va a ayudarnos a elegir el número correcto. Lo que hacemos es correr el algoritmo para un rango de valores k y evaluar en cada uno el costo asociado.

Idealmente queremos un costo bajo para que los clusters sean más bien homogéneos, pero el costo va disminuyendo a medida que aumenta k hasta ser 0 cuando tenemos 1 cluster por cada observación. Para “saber dónde parar” vamos a elegir el punto de

inflexión del gráfico: donde comenzamos a tener rendimientos decrecientes por aumentar k . Sería algo así como “el codo” de la curva graficada.

El método del codo es una manera de validación de clusters con la que contamos, considerando que, a diferencia de en el aprendizaje supervisado donde contamos con las etiquetas de los grupos, acá no tenemos esta comparación directa que nos permite saber si el algoritmo está funcionando correctamente o no. Por eso los chequeos son más robustos si utilizamos otras técnicas para su evaluación.

Utilizar el Elbow o Silhouette es una forma de validación interna, porque estamos utilizando información que sale del proceso de clusterización para evaluar qué tan buena es la estructura que conseguimos. Otra forma de llevar a cabo esta validación es con la opinión de un experto en el tema que pueda validar si los resultados encontrados se encuadran dentro de lo posible. Si contamos con cualquier otra información externa también podemos utilizarla como método de validación.

Anexo

Tabla 1

Identidad de género autopercebida	Cantidad de inscriptos	Porcentaje sobre el total
Mujer	55,422	49.5%
Varón	53,697	47.9%
No binaria	1,317	1.2%
Prefiero no decirlo	1,066	1.0%
Otras	248	0.2%
Varón trans	164	0.1%
Mujer trans	141	0.1%

Tabla 2

Nacionalidad	Cantidad de inscriptos	Porcentaje sobre el total
Argentina	109,516	97.7%
Chile	357	0.3%
Venezuela	346	0.3%
Uruguay	318	0.3%
Colombia	306	0.3%
Paraguay	250	0.2%
Perú	233	0.2%
Bolivia	189	0.2%
Brasil	134	0.1%
España	63	0.1%
Estados Unidos	40	0.0%
Italia	39	0.0%
Ecuador	37	0.0%
Francia	34	0.0%
México	34	0.0%
Cuba	24	0.0%
Alemania	13	0.0%
República Dominicana	8	0.0%

Honduras	8	0.0%
Canadá	8	0.0%

Tabla 3

Provincia de residencia	Cantidad de inscriptos	Porcentaje sobre el total
BUENOS AIRES	28,541	25.47%
Capital Federal (C.A.B.A.)	17,131	15.29%
CÓRDOBA	9,145	8.16%
SANTA FE	6,366	5.68%
SANTIAGO DEL ESTERO	5,362	4.79%
CHACO	5,127	4.58%
TUCUMAN	4,742	4.23%
MENDOZA	4,002	3.57%
SALTA	3,884	3.47%
SAN LUIS	3,603	3.22%
MISIONES	3,235	2.89%
JUJUY	3,216	2.87%
CORRIENTES	2,738	2.44%
ENTRE RÍOS	2,211	1.97%
CATAMARCA	2,193	1.96%
FORMOSA	1,994	1.78%
SAN JUAN	1,931	1.72%
RIO NEGRO	1,778	1.59%
CHUBUT	1,196	1.07%
NEUQUÉN	1,036	0.92%
LA RIOJA	933	0.83%
SANTA CRUZ	617	0.55%
TIERRA DEL FUEGO	567	0.51%
LA PAMPA	507	0.45%

Gráfico 1

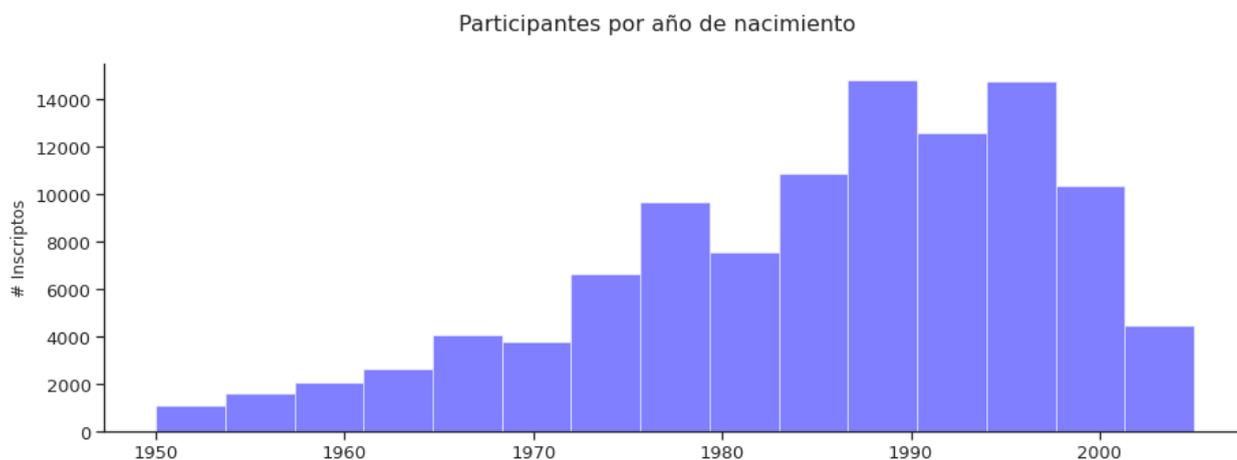


Tabla 4

Situación ocupacional en la actividad cultural	Cantidad de inscriptos	Porcentaje sobre el total
Trabajador/a informal	44,120	39.4%
Desocupado/a	21,000	18.7%
Monotributista	20,140	18.0%
Voluntario/a	14,316	12.8%
Relación de dependencia	9,450	8.4%
Responsable Inscripto / Autónomo	2,164	1.9%
Becario/a de estudio/investigación	859	0.8%

Tabla 5

Ingresos provenientes de la actividad cultural	Cantidad de inscriptos	Porcentaje sobre el total
0 a 25 %	40,388	36.0%
Más del 70%	25,639	22.9%
No corresponde	19,264	17.2%
26 a 50%	15,760	14.1%
51 a 70%	10,992	9.8%

Tabla 6

Situación ocupacional en la actividad cultural	Ingresos provenientes de la actividad cultural	Cantidad de inscriptos	Porcentaje sobre el total
Desocupado	No corresponde	19,264	17.2%
Trabajador/a informal	0 a 25 %	17,517	15.6%
Trabajador/a informal	Más del 70%	11,954	10.7%
Voluntario/a	0 a 25 %	10,998	9.8%
Monotributista	Más del 70%	9,168	8.2%
Trabajador/a informal	26 a 50%	8,564	7.6%
Trabajador/a informal	51 a 70%	6,079	5.4%
Monotributista	0 a 25 %	5,436	4.9%
Dependencia	0 a 25 %	3,921	3.5%
Monotributista	26 a 50%	2,914	2.6%
Dependencia	Más del 70%	2,754	2.5%
Monotributista	51 a 70%	2,620	2.3%
Voluntario/a	26 a 50%	1,876	1.7%
Dependencia	26 a 50%	1,700	1.5%
Desocupado	0 a 25 %	1,169	1.0%

Tabla 7

Inscrito a registros	Cantidad de inscripción es a registros	Porcentaje sobre el total de inscriptos
No	68,350	61.0%
Registro Nacional del Fondo Nacional de las Artes	25,859	23.1%
Registro Nacional de la Música	13,749	12.3%
Registro Nacional del Teatro Independiente	6,425	5.7%
Registro Nacional de trabajadores y trabajadoras de la Economía Popular	5,459	4.9%
Otro	3,925	3.5%
Registro Público de la Actividad Cinematográfica y	2,006	1.8%

Audiovisual (INCAA)		
Registro Nacional del MATRA	1,524	1.4%
MiPyme	1,118	1.0%
Registro Nacional de Promotorxs Territoriales de Género y Diversidad	778	0.7%

Gráfico 2

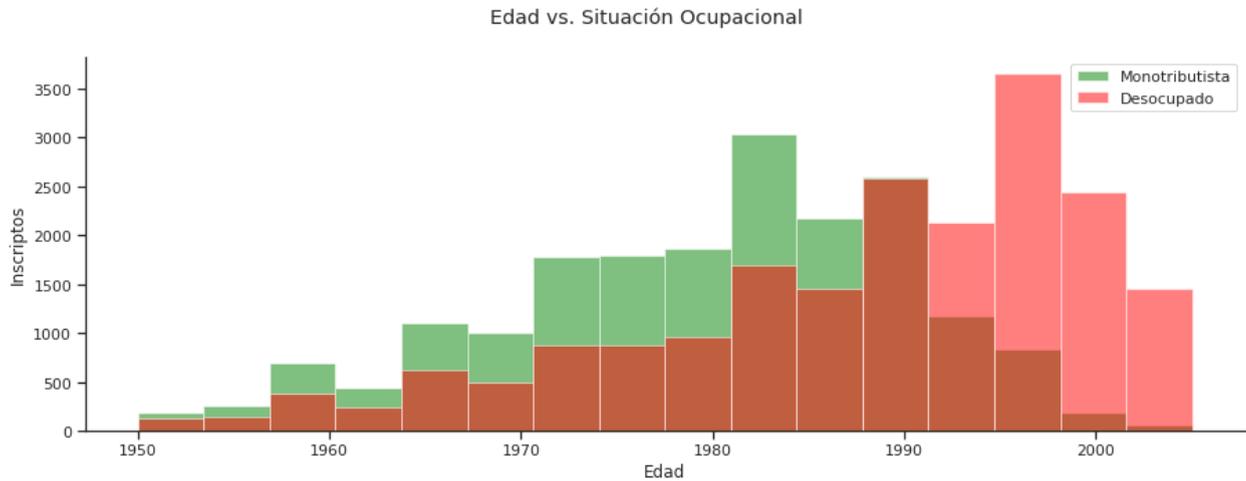
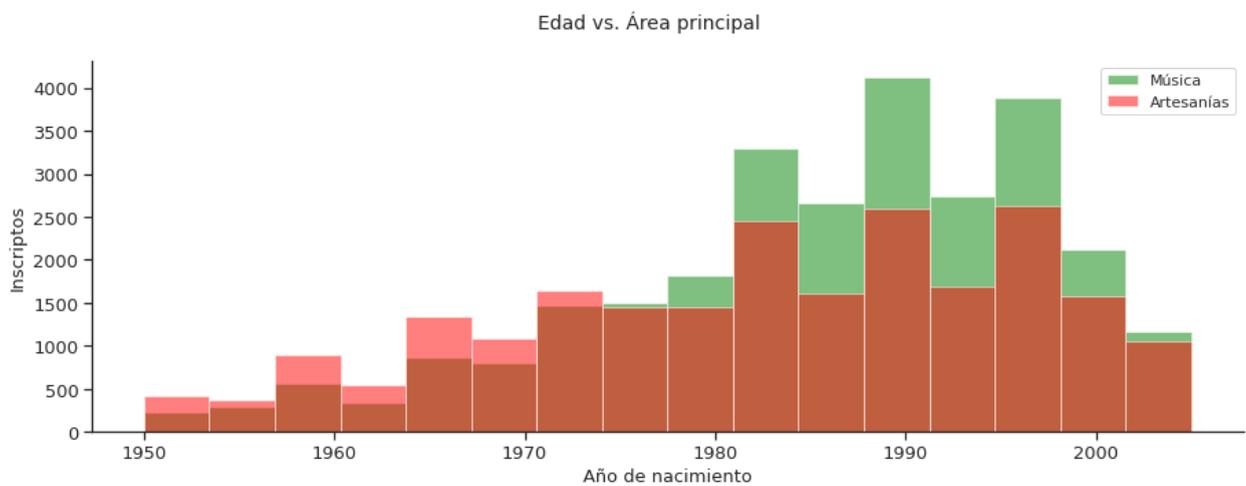


Gráfico 3



Bibliografía

Beecroft, E., y W. S. Lee. 2000. Looking beyond mean impacts to see who gains and who loses with time-limited welfare: Evidence from the Indiana Welfare Reform Evaluation. Trabajo presentado en el congreso anual de la *National Association for Welfare Research and Statistics*, Scottsdale, Arizona.

Huang, Zhexue. "Extensions to the k-means algorithm for clustering large data sets with categorical values." *Data mining and knowledge discovery* 2, no. 3 (1998): 283-304.