Recopilación y análisis de datos, proyecto Animupa: Creación Musical, Catastro FOJI y Fondos de Cultura

Ignacio Rivera
Doctor en Sociología
Equipo de Datos Anillo de Música de Arte como Patrimonio (Animupa)
animupadatos.ignacio@gmail.com

Louis Leclerc
Ingeniero en datos
Equipo de Datos Anillo de Música de Arte como Patrimonio (Animupa)
contacto.louisleclerc@gmail.com

Introducción

1.1 Animupa y el Equipo de Análisis de Datos

En abril del año 2023 se inició el Anillo de Música de Arte como Patrimonio (Animupa, Proyecto Anillo ATE220041), un proyecto de tres años financiado por la Agencia Nacional de Investigación y Desarrollo (ANID) para el estudio de la música clásica en Chile como práctica cultural y forma de patrimonio vivo. En su diseño, el proyecto contempló crear un área de trabajo sobre investigación estadística y recopilación de datos relacionados con la música de arte en Chile. Los autores de este documento formamos el Equipo de Análisis de Datos encargados de liderar esta área de trabajo.

1.2 Plan de gestión de datos

La primera etapa del proceso de recopilación y análisis de datos consistió en establecer los criterios de búsqueda de datos. Se dispuso la elaboración de una estrategia de gestión de datos que consistió en establecer por escrito, usando la plataforma DMP Tool,[1] el plan de recopilación y trabajo de datos de todo el proyecto. Para esto último, se especificó una serie de criterios, incluyendo los datos de interés, el método y estrategia de búsqueda, el procesamiento, el almacenamiento en archivos, nombramiento de documentos, etapas de análisis, metadatos y asuntos éticos.

El Equipo se propuso como objetivo caracterizar aspectos relacionados con la música de arte en Chile a partir de los siguientes datos:

- Eventos musicales (tiempo y espacio, agentes involucrados, entre otros).
- Financiamiento (eventos, instituciones y proyectos).
- Información proveniente de agentes diversos tales como compositoras/es, intérpretes, audiencias, *managers*, instituciones culturales, académicas, musicales, conservatorios, asociaciones musicales, coros y orquestas.

1. https://dmptool.org/.

Con respecto a los principales métodos de recopilación se establecieron los siguientes:

- Solicitud de base de datos a organizaciones públicas de Chile a través de la Ley de Transparencia.
- Recopilación de bases de datos descargables directamente de la web.
- Recopilación de datos de manera manual en páginas web.
- Solicitud de datos a personas claves (personas directoras de programas, encargadas de archivos, etc.), a través de contacto por correo electrónico y llamadas telefónicas.

Otros métodos fueron, dependiendo de su razón de costo/beneficio:

- *Scraping*,[2] cuando la fuente está bien identificada y con una estructura HTML/CSS conocida.
- *Crawling*,[3] cuando la fuente es diversa y la estructura es variable (ej. sitios web de prensa).

Respecto a la estrategia de recopilación de datos, se estableció priorizar la búsqueda de bases de datos de acceso público que contuvieran un número alto de información y que estuvieran estructuradas. Una vez que las bases de datos más grandes fueran recopiladas, se consideró comenzar a recopilar información de fuentes con un número de casos menor. Se decidió que se dejarían para el final las bases de datos que contuvieran información sensible y que, por lo tanto, requirieran más etapas para su procesamiento. En adición, se planteó como estrategia preguntar a las y los actores claves si pudieran compartir información y bases de datos propias con el equipo.

Una vez recopilados y almacenados los datos en bruto, se estableció realizar varias operaciones de procesamiento, como la creación de un identificador único para cada fila de datos, limpieza de información (por ejemplo, borrado de puntos), conversión del tipo de datos (por ejemplo, fechas), creación de nuevas variables, o completar valores faltantes. En el plan de gestión de datos se definió que los archivos trabajados se almacenarían en una carpeta de Google Drive creada especialmente para el proyecto Animupa. Se establecieron cuatro carpetas para el almacenamiento:

1. Datos brutos: corresponde a los datos crudos recolectados (a) manualmente, (b) a través de bases de datos disponibles directamente en la web, (c) a través de bases de datos disponibles a partir de solicitudes.

2. Datos transformados: son los datos modificados respecto a su estado original para facilitar su análisis. Estas modificaciones pueden incluir limpieza de datos, estandarización, descomposición, etc.

3. Análisis de datos: son los datos de la etapa anterior analizados, por ejemplo, en forma de gráficos, tablas o visualizaciones.

---

2. Por *scraping* entendemos el proceso de extraer datos o información de un sitio web de manera automatizada, generalmente a través de un *software* o conjunto de instrucciones (o *script*).

3. Por *crawling* entendemos el proceso de recorrer de manera sistemática las páginas de un sitio web para recopilar datos o indexar su contenido.

4. Resultados de la investigación: es la contextualización de los datos analizados, que generalmente se realiza por medio de un documento escrito. Los resultados de la investigación se consideran un análisis intermedio, que tiene como objetivo guiar las investigaciones en curso realizadas por los investigadores del equipo.

Otro aspecto que se consideró fue la inclusión de un documento de descripción de los archivos almacenados. A partir de reuniones con el equipo ampliado de Animupa,[4] se tomó la decisión de recopilar datos con información reciente (últimos diez años aproximadamente). Una de las razones por las que se tomó esta decisión fue para focalizarse en un periodo específico y, en concordancia con los fundamentos del proyecto, estudiar el rol y contribución actual de diversos agentes vinculados al mundo de la música de arte.

1.3 Datos recopilados y analizados a la fecha

Considerando los criterios de recopilación establecidos en el plan de gestión de datos, además de las sugerencias de las y los investigadores principales del proyecto, comenzamos la recopilación de datos a partir de tres fuentes sobre música de arte en Chile. En primer lugar, recopilamos información de la sección "Creación Musical Chilena" de la *Revista Musical Chilena*, correspondiente a los últimos diez años. Esta sección otorga una visión general de las obras de compositoras y compositores chilenos interpretadas en un periodo de tiempo, e incluye una serie de datos por cada obra interpretada, a saber, el nombre de el/la compositor/a, título y medio de la obra, fecha, ocasión y lugar, e intérpretes, entre otras características. La decisión de seleccionar esta sección con los datos que aporta en lugar de aquellos disponibles en otras revistas especializadas del medio local se justifica pues dan cuenta de un trabajo frecuente y sostenido de recopilación de información por parte del equipo editorial de la *Revista Musical Chilena*, con variables precisas, lo que hace posible obtener un gran volumen de datos estructurados.

En segundo lugar, se recopilaron datos del directorio de la red de orquestas registradas en la Fundación de Orquestas Juveniles e Infantiles de Chile (FOJI), disponibles en la página web de dicha institución.[5] Este directorio corresponde a un registro en línea de las orquestas que son parte de la comunidad FOJI.[6] Incluye información como el nombre de la orquesta, región y comuna de procedencia, número de instructores/as y elencos y datos de contacto del equipo.

En tercer lugar, se recopilaron datos del Fondo de la Música a través de solicitudes al Portal de Transparencia de Chile.[7] Algunas de las variables que se obtuvieron de esta base de

4. Estas reuniones de toma de decisión se realizaron el año 2023, correspondiente al primer año del proyecto. En estas reuniones participaron de manera regular las y los investigadores Rodrigo Cádiz, José Manuel Izquierdo, Daniela Fugellie, Eileen Karmy, René Silva, Nayivé Ananías y Pablo Soto. El año 2024 se incorporó el investigador Federico Eisner.

5. https://catastro.foji.cl/orquestas/ver.

6. Este catastro no incluye información sobre las dieciocho orquestas sinfónicas regionales de FOJI. El catastro abarca el registro de las orquestas de la comunidad FOJI, que son más numerosas (aproximadamente trescientas) y generalmente cuentan con elencos más pequeños que las orquestas sinfónicas regionales. En la primera etapa del proceso de recopilación de información se priorizó la obtención de los datos de las orquestas de la comunidad FOJI, ya que su número es mayor que el de las orquestas sinfónicas regionales de FOJI, lo que se alinea con la estrategia de buscar bases de datos con mayor cantidad de casos.

7. Ver https://www.consejotransparencia.cl/portal-de-transparencia/.

datos fueron: nombre del proyecto, línea, modalidad, estado de selección, nombre del o la responsable del proyecto, montos solicitados y montos asignados, entre otras. Uno de los principales motivos por los cuales seleccionamos las bases de datos de estas instituciones es porque contenían información en alto volumen, accesible, estructurada y acorde al tema de estudio de nuestro proyecto de investigación.

Una vez recopilados los datos en bruto, se trabajó en las etapas subsiguientes de limpieza y transformación de datos, análisis y comunicación de los resultados, para cada una de las fuentes recientemente expuestas.[8] La comunicación de los resultados se estableció a partir de resúmenes de hallazgos relevantes escritos en informes, los cuales son de acceso abierto, y disponibles a través de la plataforma GitHub.[9]

A continuación, compartimos la descripción, algunos hallazgos y reflexiones de cada una de estas bases de datos.

2. Bases de datos, hallazgos y reflexiones

2.1 *Revista Musical Chilena* - Sección Creación Musical Chilena

2.1.1 Descripción proceso de recopilación y procesamiento de datos

En primera instancia, nos comunicamos con el director y con el asistente de redacción de la *Revista Musical Chilena* para informar nuestro interés de armar una base de datos con la información de la sección, y para consultar si existía una versión en bruto de los datos en Excel u otro formato que nos pudieran compartir. La respuesta fue que no existía un documento con esas características, por lo tanto, procedimos a obtener la información directamente desde la página web de la revista.

Para ello, se descargaron los documentos correspondientes a la sección Creación Musical de los últimos diez años, en formato PDF. Se transformaron los documentos de formato PDF a Word, y luego se ingresaron los datos del documento Word en un Excel. Una vez ingresados en Excel, se trabajó en el ordenamiento de los datos de manera manual para que quedaran en formato de matriz de datos, con columnas y filas ordenadas, donde cada columna corresponde a una variable y cada fila a un caso. A cada caso se agregó un identificador (ID), por ejemplo, para el caso correspondiente a la interpretación de la obra "Sirilla del nuevo día" (2010), del compositor Juan Sánchez, por parte del dúo Orellana-Orlandini el 10 de agosto del 2013, se asignó el ID RMCh220_245, donde RMCh corresponde a *Revista Musical Chilena*, 220 al número de la revista, y 245 al número de caso dentro de ese número de la revista. El resultado fue la obtención de una base de datos con 4.841 casos y 9 variables.

Una vez que se finalizó el proceso de agregar un identificador a cada caso, se ingresó la base de datos a la carpeta de datos del proyecto. Luego se realizaron las primeras operaciones de

8. Se estableció el acrónimo RMCh como identificador de los datos de la sección Creación Musical Chilena, FOJI como identificador de los datos de la Fundación de Orquestas Juveniles e Infantiles, y FMu como identificador de los datos del Fondo de la Música.

9. GitHub es un repositorio en línea utilizado para gestionar proyectos y compartir códigos. En el enlace https://github.com/animupa se puede acceder a estos informes.

procesamiento de los datos. Una primera operación fue separar en dos celdas distintas las informaciones concentradas en una celda, como, por ejemplo "título y medio", que aparecen dentro de un mismo parámetro en la sección Creación Musical, pero que en nuestra base de datos dejamos como información diferenciada ("título" y "medio"). Otra operación fue modificar el formato de la fecha de cada caso.[10] Posteriormente se establecieron los primeros análisis de los datos, principalmente a través de gráficos y tablas, los cuales fueron escritos en un informe. A continuación, compartimos algunos de los hallazgos de este primer informe, que se centra en el periodo 2018-2022.

2.1.2 Principales hallazgos: periodo abril del 2018 hasta septiembre del 2022

Número de eventos por año. Durante el año 2018 (contabilizado desde abril), el número de eventos de obras de compositoras y compositores chilenos registrados por la sección Creación Musical Chilena fue de 245, mientras que para el año 2019 el número fue de 345. Hay una baja considerable durante el año de la pandemia, contabilizando 75 eventos, de los cuales más de la mitad (53%) fueron realizados en línea. El año 2021 sube la cantidad de eventos a 179, de los cuales la gran mayoría se realizó en formato presencial. En el año 2022 sigue subiendo el número de eventos y llegan a 196, considerando hasta septiembre de ese año.

Año	Cantidad de eventos	Facebook Live	% Online
2018	245	0	0%
2019	345	0	0%
2020	75	40	53%
2021	179	2	1%
2022	196	6	3%

Tabla 1 / Número ocurrencias eventos por año (Creación Musical Chilena, 2018-2022).

Compositores y compositoras más interpretadas. En este mismo periodo es posible contabilizar las ocurrencias de los/as compositores/as chilenas más interpretadas, según los datos proporcionados por la *Revista Musical Chilena*. Entre los primeros cinco encontramos a Violeta Parra, Enrique Soro, Pedro Humberto Allende, Rolando Alarcón y Alfonso Leng.

Nombre del compositor/a	Ocurrencias en la RMCh
Parra Sandoval, Violeta	34
Soro Barriga, Enrique	34
Allende Sarón, Pedro Humberto	30
Alarcón Soto, Rolando	28
Leng Haygus, Alfonso	24
Carvallo Pinto, Antonio	22

10. En el proceso de transformación de datos, es común realizar modificaciones del formato de las fechas con el fin de mejorar consistencia, comparabilidad y seguir estándares internacionales.

García Arancibia, Fernando	19
Sepúlveda Maira, María Luisa	19
Rifo Suárez, Guillermo	18
Matamoros Figueroa, Ximena	17

Tabla 2 / Número de ocurrencias por compositor/a (Creación Musical Chilena, 2018-2022).

Intérpretes más frecuentes. Los/as intérpretes con más ocurrencias en el periodo 2018-2022, fueron la Orquesta Marga Marga, Luis José Recart, Patricia Castro, Luis Orlandini, Karem Ruiz y José Luis Urquieta. Existen casos particulares como el de los "medios electrónicos", ya que no corresponde a una persona en particular, sino a un medio no humano de interpretación, lo que ilustra la evolución del rol del músico en función de nuevos avances tecnológicos.

Intérpretes	**Ocurrencias en la RMCh**
Orquesta Marga Marga	78
Luis José Recart	71
Patricia Castro	57
Luis Orlandini	23
Karem Ruiz	21
José Luis Urquieta	21
Medios electrónicos	19
Wilson Padilla	19
Trío Violeta	17
Paulette L'Huisser	15
Enrique Schadenberg	14
Orquesta de Cámara de Chile	13

Tabla 3 / Número de ocurrencias por intérprete (Creación Musical Chilena, 2018-2022).

2.1.3 Reflexiones

Respecto a los datos exhibidos es posible realizar una serie de análisis. Por ejemplo, al agregar variables como sexo/género y fecha de nacimiento de compositoras/es e intérpretes es posible examinar las distribuciones de sexo/género, la época de las compositoras/es más frecuentes y a qué generación pertenecen las y los intérpretes. También, si tomamos un periodo más largo para el análisis, por ejemplo, desde el año 2012 al 2022, es posible observar el efecto de ciertas conmemoraciones –por ejemplo, los cien años del natalicio de Violeta Parra en 2017 y los noventa años de Rolando Alarcón en 2019– en el aumento de la interpretación de las obras de esos autores. Por otro lado, es posible analizar la frecuencia de los eventos según mes del año, o bien, los efectos de la pandemia en la cantidad de presentaciones.

Sobre los potenciales análisis. Aparte de ocurrencias de eventos, fechas, compositoras/es e intérpretes, la base de datos nos entrega una vasta variedad de temas de análisis, como, por ejemplo, los títulos de las obras, así como los medios, la ocasión y lugares en los que se

interpretaron. También es posible recabar información valiosa con respecto a cuáles son las obras más interpretadas, los instrumentos o familias de instrumentos más frecuentes y los lugares de presentación más comunes, considerando espacios como teatros o universidades. También se puede indagar si la ocasión corresponde a algún festival o evento de temporada.

Valor de la base de datos. Es importante destacar el valor del trabajo de levantamiento de información realizado por la *Revista Musical Chilena*, considerando la labor de Julio Garrido Letelier, Karen Sánchez Guimai y Nancy Sattler Jiménez, responsables de la sección en los últimos años. El valor de esta base se sustenta en varios aspectos. El primero es su condición de ser periódica, lo que permite obtener información sobre periodos largos de tiempo. El segundo es su nivel de estructura, ya que la información levantada responde a parámetros claros como fecha, intérprete, ocasión y lugar, entre otros. Un tercer valor es la continuidad y responsabilidad institucional de que exista y se mantenga esta sección en la revista. Finalmente, la sección pone énfasis en las obras de compositoras y compositores chilenos, dando visibilidad a la creación musical nacional que, en ciertos espacios de la música de arte, tiende a tener un rol secundario respecto a la de compositores extranjeros.

Limitaciones. Es posible identificar algunas limitaciones de la base de datos recopilada. La primera es que el universo estudiado no es necesariamente representativo de toda la escena de composiciones de música de arte chileno, ya que es posible que existan eventos que no hayan sido registrados por diversos motivos. La segunda limitación es que la variable "ocasión y lugar" es particularmente trabajosa de analizar, porque es la que contiene información más desestructurada. En esta variable se concentra tanta información del evento (ej. concierto, festival), de la institución organizadora (ej. Universidad de Chile), y del lugar físico en que se llevó a cabo (ej. Sala Isidora Zegers), que el desafío del Equipo de Datos es encontrar fórmulas de análisis de esta variable que contrarresten esta dificultad.

2.2 Catastro Fundación de Orquestas Juveniles e Infantiles de Chile

2.2.1 Descripción de la base de datos

En paralelo al proceso de levantamiento de datos de la *Revista Musical Chilena* realizamos el levantamiento de datos del Directorio de Orquestas que son parte de la comunidad de la Fundación de Orquestas Juveniles e Infantiles (FOJI). La primera etapa del proceso de recolección fue consultar directamente a través de correo electrónico al área encargada del catastro de FOJI si existía alguna base de datos en formato Excel u otro que nos pudieran compartir. La respuesta fue que no era posible compartir esos datos ya que había información sensible de menores de edad. Considerando esa situación, tomamos la decisión de recopilar la información directamente de la página web, la cual, dada su condición de página pública, no contiene información sensible.

Evaluamos si era conveniente levantar la información mediante la técnica de *web scraping*, pero llegamos a la conclusión de que, dada la estructura de la página, era mejor invertir el tiempo en recopilar la información manualmente. Se procedió a ingresar a través de la página web de FOJI a cada una de las fichas de las 296 orquestas catastradas en ese momento, y a copiar la información a una pestaña de Excel. Una vez copiada la información, se armó una matriz de datos con columnas y filas con la información levantada. Una vez construida la

matriz de datos, se estableció el ID de cada uno de los casos. El resultado fue la obtención de una base de datos con 296 casos y 15 variables.

Las siguientes operaciones fueron similares a las descritas anteriormente para la base de la *Revista Musical Chilen*a. El archivo se ingresó a la carpeta de datos del proyecto y se realizaron operaciones de procesamiento de datos como, por ejemplo, llenar los valores nulos. Luego se establecieron los primeros análisis de los datos y se escribieron dos informes de resultados. El primero de estos informes se enfocó en el análisis de la distribución geográfica de las orquestas y el segundo analizó otras variables, como la cantidad de instructores por orquesta, el tamaño de los elencos y los instrumentos interpretados. A continuación, compartimos algunos de los hallazgos de los dos informes mencionados, que incluyen información de las orquestas de la comunidad FOJI catastradas el año 2023.

2.2.2 Principales hallazgos

Distribución de orquestas por región y comuna. De acuerdo a los datos del catastro, las regiones que tienen un mayor número de orquestas son la Metropolitana (69), la de Valparaíso (35) y la del Biobío (30); mientras que las que tienen un menor número son las regiones de Aisén (8), Tarapacá (6), Arica y Parinacota (4) y Magallanes (1).

Región	**N**
Región Metropolitana	69
Región de Valparaíso	35
Región del Biobío	30
Región del Maule	26
Región de los Lagos	24
Región de Coquimbo	20
Región de Ñuble	16
Región de Antofagasta	14
Región de Atacama	11
Región del Libertador Gral. Bernardo O'Higgins	11
Región de la Araucanía	11
Región de los Ríos	10
Región de Aisén del Gral. Carlos Ibáñez del Campo	8
Región de Tarapacá	6
Región de Arica y Parinacota	4
Región de Magallanes y de la Antártica Chilena	1
TOTAL	296

Tabla 4 / Distribución de orquestas por región (Catastro FOJI, 2023).

En cuanto a las comunas, las tres con más orquestas que forman parte de la comunidad FOJI de Chile son Antofagasta (13), La Serena (9) y Santiago (9). Luego le siguen cuatro comunas con siete orquestas cada una, que son Coquimbo, Valparaíso, Puente Alto y Copiapó. En el siguiente gráfico se puede visualizar la cantidad de orquestas por comuna. En el norte de

Chile se observa una mayor dispersión geográfica de las orquestas en comparación con la zona centro-sur, donde existe una alta densidad de orquestas distribuidas en el territorio. En los extremos de Chile, sobre todo el austral, la presencia de orquestas es mínima o nula.

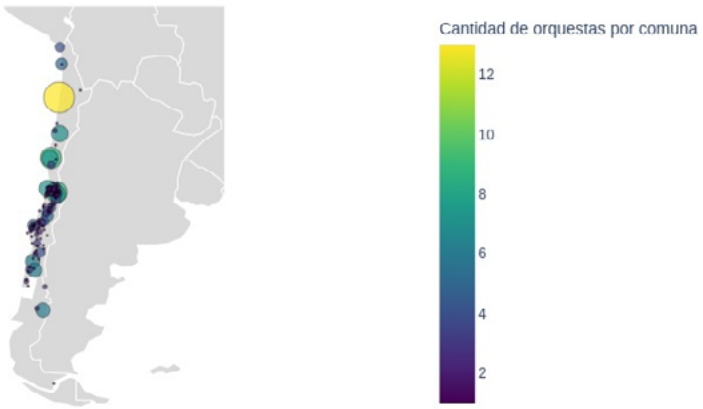

Gráfico 1 / Cantidad de orquestas por comuna (Catastro FOJI, 2023).

Cantidad de instructores/as por orquesta y tamaño de los elencos. De las 296 orquestas catastradas, 172 tienen registro del número de instructores/as. Entre las orquestas con registro del número de instructoras/es, el promedio de instructores/as por orquesta es de 5.5, con un mínimo de un instructor/a y un máximo de veinte instructores/as. En tanto, la diferencia entre el tamaño del elenco de las orquestas es significativa, ya que la orquesta con el elenco más pequeño tiene solo tres integrantes, mientras que la orquesta con el elenco más grande tiene 96 integrantes. El promedio del elenco por orquesta es de veintiocho integrantes.

En el Gráfico 2 podemos visualizar la razón de estudiantes por instructor/a según región. Regiones extremas como las de Arica y Parinacota y la de Tarapacá en el norte, y la de Aysén en el sur, presentan casos de orquestas con una razón de instructores por estudiantes baja. Una de las comunas no extremas en la cual ocurre un fenómeno similar es la de la Araucanía, donde existen pocas orquestas y la razón de instructores/as por estudiante es baja.

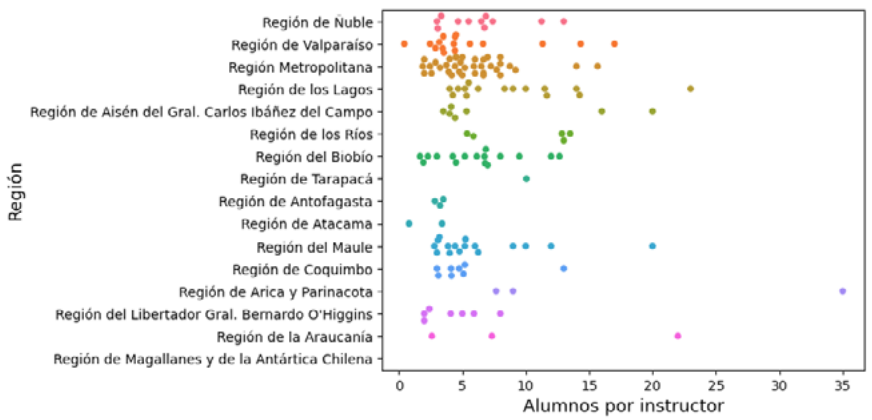

Gráfico 2 / Razón de estudiantes por instructor/a en cada región (Catastro FOJI, 2023).

2.3.3 Reflexiones

Los datos sobre la distribución de las orquestas en Chile nos otorgan luces sobre los problemas de centralismo en el país, ya que hay una concentración muy alta de orquestas en la zona centro-sur, y en las zonas extremas la presencia es mucho menor o nula. También dan cuenta de la existencia de comunas específicas como Antofagasta o La Serena que cuentan con un número importante de orquestas. Ahora bien, estos datos se podrían comparar con los datos poblacionales para ver si la distribución de orquestas obedece a tendencias demográficas u otros factores. También es posible mencionar la desigualdad existente entre estudiantes por instructor/a, donde se identifica una orquesta con una razón de un instructor/a por cada 35 estudiantes, lo que difiere considerablemente del promedio.

Sobre los potenciales análisis. La base de datos también nos entrega información sobre los tipos de instrumentos tocados por las y los estudiantes, si la orquesta es estudiantil o juvenil y si ocupan o no redes sociales. Existiría más potencial de análisis si agregáramos la información relativa a las orquestas sinfónicas juveniles que son parte de FOJI.

Limitaciones. Una primera limitación de esta base de datos es que tiene mucha información en blanco. Una segunda limitación es que hay algunos datos que no son confiables o de dudosa procedencia, por ejemplo, cuando se señala que hay una orquesta que tiene como elenco a una sola persona. La otra gran limitación tiene relación con el proceso de recopilación de datos, ya que es plausible que existan datos relevantes sobre las orquestas que no estén disponibles por contener información sensible de menores de edad.

2.3 Fondo de la Música 2012-2023

2.3.1 Descripción de la base de datos

La tercera base de datos recopilada y analizada fue la del Fondo de la Música en el periodo 2012-2023. Se solicitó esta base de datos con el propósito de seleccionar aquellos proyectos asociados a música de arte y analizar sus características, ya sea dentro de este mismo grupo, o bien en comparación con proyectos asociados a otros géneros musicales. La base de datos se obtuvo a partir de una solicitud realizada al Ministerio de las Culturas, las Artes y el Patrimonio a través del Portal de Transparencia. Recibimos dos bases de datos, una correspondiente a los proyectos postulados entre los años 2012-2016, y otra a los proyectos postulados entre los años 2017-2023. Cada una de estas bases de datos tenía información disgregada por año, es decir, una pestaña de Excel para los proyectos del año 2012, otra para los del 2013, etc. La siguiente operación fue agregar un ID a cada caso. El resultado fue la obtención de una base de datos con 21.198 casos y 15 variables. Cada caso corresponde a una postulación.

Una vez agregados los identificadores, la siguiente operación fue ingresar la base de datos a la carpeta de datos del proyecto. La siguiente etapa fue la limpieza y transformación de datos, tal como en los casos anteriores. Un proceso importante fue agregar todos los datos disgregados en distintas pestañas por año, en una única matriz de datos que contuviera todos los años y casos. Este trabajo presentó el desafío metodológico de homologar las diversas variables existentes entre años distintos en una misma matriz. Luego se procedió a mejorar el sistema de identificación asociando a cada proyecto un ID que contuviera información del número

de folio del proyecto. Otra etapa importante de la transformación de los datos fue el proceso de identificar los proyectos que estaban asociados a la música de arte dentro del Fondo de la Música.[11]

Una vez que la base de datos estuvo lo suficientemente limpia, se realizaron los primeros análisis a partir de los cuales se escribieron un par de informes. Después de compartir algunos resultados con el equipo ampliado de Animupa, se planteó la posibilidad de realizar una solicitud al Mincap para obtener una base de datos que contuviera más variables aún, tales como las características de las/los postulantes y más información sobre el género musical de cada proyecto individual. Recibimos la nueva base de datos del Mincap con las características de las/los postulantes. Se agregaron los ID a los casos de los datos obtenidos, se limpiaron y estructuraron, para luego generar una visualización y gráficos incluidos en un tercer informe. A continuación, compartimos algunos hallazgos de este último informe.

2.3.2 Principales hallazgos

Montos y adjudicaciones. Durante el periodo 2012-2023 postularon 28.040 proyectos, con un promedio de 2.337 proyectos postulados cada año. Se aprobaron 5.804 proyectos, con un promedio de 483 proyectos aprobados por año. La tasa de aprobación promedio del periodo fue de un 21%, es decir, alrededor de uno de cada cinco proyectos fue aprobado. El monto total adjudicado del periodo 2012-2023 fue superior a los 36 mil millones de pesos, el promedio fue de un poco más de 3 mil millones por año. Alrededor de mil millones se han descontado por ajustes presupuestarios. El ajuste promedio por año ha sido de 86 millones de pesos. En todos los años se observa un ajuste negativo, a excepción del año 2020 donde hay un ajuste positivo de alrededor de 73 millones.

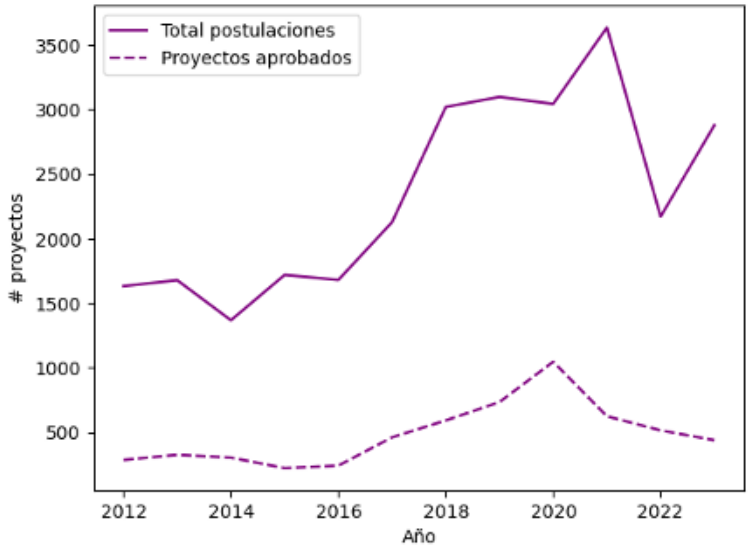

Gráfico 3 / Cantidad de proyectos por año (Fondo de la Música 2012-2023).

11. Este proceso en particular ha requerido de muchas etapas y operaciones por la dificultad que conlleva asignar un género musical a proyectos que no necesaria ni explícitamente hacen referencia a su género musical de pertenencia. En las reflexiones dedicaremos un espacio a este tema.

Características de las y los postulantes. Alrededor del 80% de los proyectos contiene información sobre el sexo de la persona responsable. Dentro del grupo de proyectos que contiene información del sexo de la persona responsable, alrededor de un 24% corresponde a responsables femeninos, y un 76% a responsables masculinos. El rango de edad donde se concentra la mayor cantidad de responsables de proyectos aprobados es entre los 25 a 40 años, con una baja considerable entre menores de 25 y mayores de 45. Casi la totalidad de las personas responsables de los proyectos aprobados son chilenas (98%). La Región Metropolitana es por gran margen la región de donde proviene la mayor cantidad de responsables de proyectos aprobados y donde se realiza la mayor cantidad de proyectos, con alrededor de 2.500 responsables de proyectos provenientes de aquella región, correspondiente al 50% del total de los proyectos adjudicados. Le sigue la Región de Valparaíso con 804 proyectos, correspondiente al 16% del total de los proyectos adjudicados.

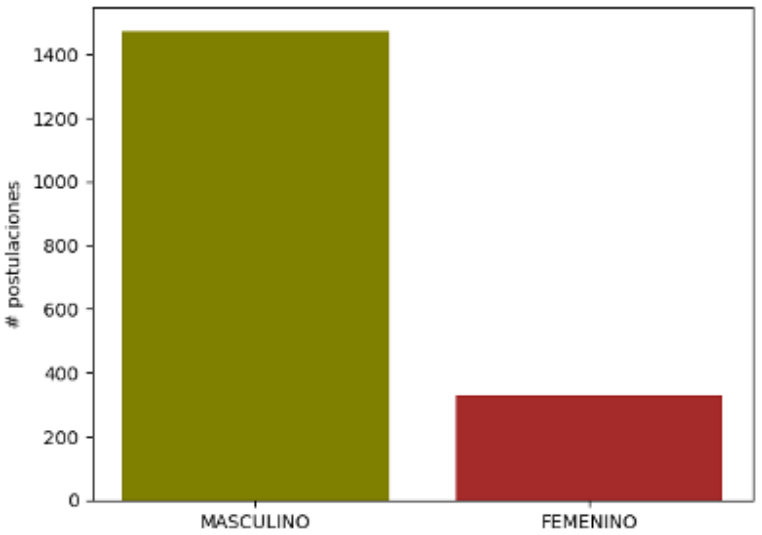

Gráfico 4 / Sexo declarado del responsable proyecto (Fondo de la Música 2012-2023).

Construcción de categorías. La identificación del género musical de cada uno de los proyectos postulados es una tarea que implica un trabajo de procesamiento y transformación de los datos obtenidos directamente desde el Mincap. Esto se debe a que en la base entregada no existe una variable que identifique a cada proyecto individualmente con un género musical específico. Sin embargo, existen líneas, modalidades y submodalidades que sí tienen un rótulo asociado a géneros musicales. Por este motivo, el proceso de tratamiento de datos implicó crear una variable en que se identifican los géneros musicales según la información recientemente señalada. Con ello, se incluyeron cuatro categorías dentro de la variable género musical creada:

- Categoría Música de Arte: se creó a partir de proyectos asociados a Líneas, Modalidades o Submodalidades que contengan las palabras "Clásica" y "Coros, Orquestas y Bandas Instrumentales (C.O.B.I)".

- Categoría Música Popular: se creó a partir de proyectos asociados a Líneas, Modalidades o Submodalidades que contengan la palabra "popular".

● Categoría Música Folklórica: se creó a partir de proyectos asociados a Líneas, Modalidades o Submodalidades que contengan la palabra “folklórica” (y derivados).

● Categoría Otros: son todos los proyectos restantes.

Resultados categoría música de arte. Entre las postulaciones de música de arte, el porcentaje es similar al general. Alrededor de un 81% son proyectos de postulantes masculinos, versus 19% de proyectos con postulantes femeninas. El promedio de edad de los responsables de proyectos de música de arte es de 39 años y la concentración de casos se da entre los 30-35 años. Dentro de este género, la región que tiene más proyectos aceptados por millón de habitantes es la de Los Ríos, con una cifra que casi triplica a las regiones de Coquimbo y Aysén, que son las que les siguen con más proyectos aceptados por millón de habitantes.

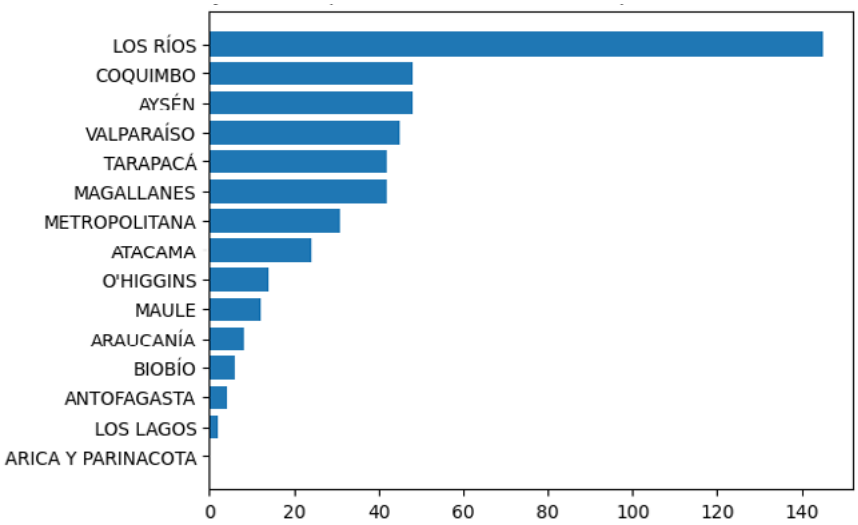

Gráfico 5 / Proyectos aceptados en música de arte, por millón de habitantes (Fondo de la Música 2012-2023).

2.3.3 Reflexiones

Los datos exhibidos nos dan luces de tendencias generales del Fondo de la Música. Aparte de las ya señaladas, es posible disgregar las tendencias generales por año o conjunto de años y analizar cambios en el tiempo. Es posible analizar las diferencias del año 2020 de la pandemia por Covid-19 con respecto a otros años. Siguiendo las categorías de género musical expuestas anteriormente, también es posible analizar tendencias por género musical, y comparar un género musical, por ejemplo, música de arte, con relación al promedio general o con relación a otros géneros musicales.

Sobre los potenciales análisis. Una de las dificultades principales del trabajo con esta base de datos ha sido la diferenciación de proyectos por género musical de aquellos proyectos que no pertenecen a una línea, modalidad o submodalidad asociada a un género musical (clásico, folclor, o popular). El Equipo de Datos está indagando si es posible que algún algoritmo de inteligencia artificial pueda identificar géneros musicales de cada proyecto de manera

individual. Este trabajo piloto se encuentra en curso. Si logramos obtener una diferenciación de géneros musicales por proyecto, será posible realizar comparaciones con un número de casos más elevados según diversas variables (montos, tasas de adjudicación, características de responsables) por género musical.

Limitaciones. Como ya se ha mencionado, una de las principales limitaciones de esta base de datos es la de no tener una diferenciación de proyectos por género musical, lo que ha conllevado la realización de una serie de operaciones. Otro aspecto complejo del procesamiento de esta base de datos ha sido su estructuración, ya que aquellas entregadas directamente por Mincap estaban separadas por año, lo que dificultaba la comparabilidad. Frente a esta limitación, el Equipo de Datos ha trabajado en la homologación y estructuración de las bases, llegando a un resultado satisfactorio en cuanto a posibilidades de análisis, pero de todas maneras consideramos importante plantear la pregunta de por qué Mincap no cuenta con una base estructurada, ya que esto podría generar varias posibilidades de análisis para dicha institución.

3. Conclusiones

Una primera reflexión tiene relación con el uso de la ciencia de datos y la estadística para los estudios musicales. En muchas áreas de conocimiento y profesionales, tanto la ciencia de datos como las estadísticas se utilizan de manera regular con provechosos resultados para los análisis. También es posible identificar un desarrollo importante de la ciencia de datos en los últimos años, especialmente con el desarrollo de la inteligencia artificial y técnicas específicas de recopilación de información como el *web scraping*. Por lo tanto, consideramos importante aprovechar los avances de estas metodologías y ciencias para el estudio de la música.

Una segunda reflexión tiene relación con el contenido más sustantivo de los hallazgos. Es posible identificar ciertos temas transversales en las tres fuentes de datos recientemente analizadas. Uno es la desigualdad de género, notoria en la distribución de postulantes del Fondo de la Música y de los/as compositores/as de obras registradas en la sección Creación Musical Chilena, ya que, en ambos casos, existe una participación masculina notoriamente mayor respecto a la femenina. Otro tema importante es la centralización, ya que, tanto en los datos del Catastro FOJI como de la *Revista Musical Chilena*, es posible percibir una centralización en la Región Metropolitana y en la Región de Valparaíso, lo cual se explica también por la centralización demográfica del país.

Finalmente, quisiéramos mencionar que actualmente hay diversas bases de datos que están en construcción y que se presentan como potencial recurso de análisis para el futuro. Algunas bases de las que estamos recopilando datos son las de: (a) las orquestas sinfónicas juveniles de la FOJI; (b) resultados del concurso de la División de las Culturas, las Artes, el Patrimonio y Diplomacia Pública (Dirac); (c) base de datos de instituciones de la Sociedad Internacional por la Música Chilena (Simuc); y (d) listado de elencos de orquestas grandes de Chile (elaboración propia realizada principalmente con datos de las páginas web de las orquestas).

Esperamos que la narración de esta experiencia sirva para abrir nuevas perspectivas de investigación en el campo de los estudios musicales.

R