



Universidad de Buenos Aires

Facultad de Ciencias Exactas y Naturales

Laboratorio de Datos

Trabajo Práctico

*Relación entre el flujo migratorio entre Argentina y países del mundo
y la cantidad de sedes argentinas en dichos países*

Equipo. Serendípicos

Integrantes - LU.

- | | |
|---------------------------------------------------------|--------|
| • Alvarez, Matías
matyy.alvarez@gmail.com | 90/12 |
| • Dumas, Román
romanimanga@gmail.com | 650/16 |
| • Nogueroles, Patricio
patricio.nogueroles@gmail.com | 653/19 |

Cursada. 2° Cuatrimestre 2024

Fecha de entrega. Miércoles 23 de Octubre de 2024

Resumen

En la presente investigación se realizó un análisis exhaustivo de los datos de las representaciones argentinas en el exterior y los flujos migratorios entre países, con el fin de evaluar el manejo de datos y su visualización. Se exploraron fuentes de datos oficiales de migraciones y sedes argentinas en el exterior y se descubrió que no existe una relación directa entre la cantidad de estas y el flujo migratorio neto. Por otro lado, se identificaron varios problemas de calidad de datos en las fuentes utilizadas, como inconsistencias en los registros, datos nulos, una organización ineficiente de las tablas, entre otros. El proceso de análisis incluyó la normalización de los datos, la creación de un Diagrama Entidad-Relación (DER) y el diseño de esquemas relacionales en tercera forma normal (3FN). Posteriormente, se utilizó el lenguaje de programación Python para filtrar las bases de datos crudas provenientes de los sitios Databank y de datos proveídos por el Estado Argentino. Además, se realizaron consultas SQL para generar reportes con información detallada sobre la cantidad de sedes por región geográfica, el promedio de flujo migratorio y la diversidad de redes sociales utilizadas por las sedes en cada país. A su vez, se desarrollaron visualizaciones utilizando herramientas de visualización de Matplotlib y Seaborn, como boxplots y gráficos de dispersión, para analizar la relación entre el flujo migratorio y la cantidad de sedes en el exterior de Argentina. Para concluir, pudo explorarse en profundidad, análisis de datos mediante, la relación entre las sedes de representación de Argentina en el exterior y los flujos migratorios. Los resultados mostraron una falta de relación directa entre estas dos variables. Se sugiere explorar con mayor profundidad otras variables más finas que pueden explicar el flujo migratorio en futuras investigaciones. El trabajo, por su parte, demuestra la importancia de una sólida comprensión de los datos y su correcta manipulación para obtener conclusiones significativas en análisis de datos complejos.

1. Introducción, objetivos y metodologías

1.1 Contexto del problema

Se entiende por “migración” al movimiento de personas o grupos de personas de un lugar a otro, generalmente cruzando fronteras geográficas o políticas. Este fenómeno puede ocurrir por una variedad de razones, incluyendo la búsqueda de mejores oportunidades laborales, la reunificación familiar, la huida de conflictos o persecuciones, desastres naturales, o cambios en las condiciones económicas o políticas. Por esto es que hablar de migración es hablar de la historia de la humanidad misma. Hablar de la historia de las relaciones internacionales, la diplomacia, y en particular de la existencia de embajadas o representaciones de un país en otro corresponde ya a otro fenómeno. El concepto de “embajada” que se usa hoy en día se remonta a los Siglos XIII-XIV en Italia, siendo su motivo principal el de la diplomacia, aunque en tiempos modernos, dada la gran importancia del turismo para las economías, los alcances se extendieron en gran medida a brindar ayuda y servicios a los ciudadanos que representan.

Cuantificar la representación de un país en otro es un tema de gran importancia, porque es una manera de informar al mundo cuánto valora ese país la relación con el otro. Se esperaría que esto sea acorde al volumen de flujo migratorio entre esos países, y por eso, en cierta medida sea proporcional a la cantidad de personas que requerirían de servicios consulares. Surge entonces la pregunta de si existe en la práctica alguna relación entre estos dos fenómenos en el caso de Argentina para con el resto del mundo. Esto puede resultar relevante, por ejemplo, en el caso de identificación de outliers: si se detectara que un país se encuentra subrepresentado en relación a sus migrantes desde o hacia Argentina, esto podría servir como motivación para un estudio específico de las causas de este fenómeno, para poder actuar acordemente en caso de considerarse anómalo. Debe notarse que no se pretende establecer relaciones de causalidad entre ambas variables en este trabajo, sino más bien establecer una relación observacional que funcione como puntapié para futuras investigaciones.

1.2 Objetivo

El objetivo principal de este trabajo es entonces determinar si existe una relación entre la cantidad de sedes diplomáticas que tiene Argentina en otros países y el flujo migratorio neto entre ambos territorios. En este contexto, se entenderá por “migrante” a aquella persona donde su país de residencia no sea el mismo que su país de nacimiento. Se intentará responder también si existe cierta relación entre flujos migratorios y la lengua de un país desde/al que se viaja desde/hacia Argentina, si la cercanía influye (si hay más o menos migración desde/hacia países limítrofes), o si hay más migración con países con mayor población.

1.3 Materiales y metodología

Basándonos en las fuentes de datos y sus características (Sección 2.2), se confeccionó un modelo relacional y un diagrama de entidad-relación (DER) para modelar la estructura de los datos (Sección 2.3). Previo a hacer los análisis, se realizó un proceso de

limpieza y preprocesamiento para garantizar la sanidad y consistencia de los datos, así como para eliminar datos espurios. Algunos de los problemas de calidad de datos encontrados se describen en la Sección 3.4, con métricas que evalúan su magnitud y discusiones en cuanto a la forma normal de las tablas originales. Asimismo, se documentan las decisiones de procesamiento tomadas, así como los métodos elegidos para el análisis en la Sección 3.5.

1.3.1. Herramientas

Se utilizaron tanto Python como SQL como lenguajes de programación principales. Por una parte, se utilizó la biblioteca Pandas para el procesamiento y filtrado de las tablas de datos, y por otra, se empleó SQL desde Python mediante la biblioteca sqlite para realizar las consultas necesarias a la base de datos.

Para la visualización, se utilizaron las bibliotecas Matplotlib y Seaborn para la creación de gráficos y reportes visuales. Los anexos con el código empleado se encuentran en formato .py, correspondiente a scripts de Python, y sirven como guía para el interesado. Estos versan sobre los procesos descritos en la sección de decisiones tomadas (Sección 3), y a la confección de los gráficos y las tablas que se observan en la sección de análisis de datos (Sección 4).

1.3.2. Fuentes

Para este estudio se hizo uso de dos fuentes de datos: una que, dado un cierto año, reúne la cantidad de migrantes entre cada par de países del mundo y otra que reúne la lista de sedes de representación argentina que existen actualmente en otros países. A continuación, se detallan las características de ambas fuentes de datos.

Dataset N° 1: Global Bilateral Migration¹

La base de datos fue recolectada y provista por el Grupo Banco Mundial, es de libre acceso y representa la cantidad de personas que migraron a otro país desde su país de origen, agrupados por género y por décadas. Se encuentran agrupados según la información de los años 1960, 1970, 1980, 1990 y 2000. Dentro del presente informe se llamará *Migrantes* a este dataset (no confundir con la tabla *Migrantes*, generada luego de la limpieza y preprocesamiento) y se utilizará directamente el ya provisto por la cátedra.

Dataset N° 2: Representaciones argentinas en el exterior²

La fuente de datos contiene información relacionada a las representaciones argentinas en el exterior (consulados, embajadas y demás delegaciones argentinas en el extranjero). Los datos están divididos en tres archivos, con el nombre como se detalla en ², cada uno siendo una tabla que contiene las siguientes vistas.

- *Datos básicos Sedes*. Lista de sedes, con su descripción, códigos de identificación únicos y ubicación (ciudad y país, e identificadores). Excepto un campo, sus

¹ Dataset original extraíble desde <https://databank.worldbank.org/source/global-bilateral-migration#>
El dataset utilizado fue el [provisto por la cátedra](#) en Google drive — *datos_migraciones.csv*

² Datasets originales extraídos desde <https://datos.gob.ar/dataset/exterior-representaciones-argentinas>
Los tres datasets en cuestión son *lista-sedes.csv*, *lista-sedes-datos.csv* y *lista-secciones.csv*

columnas son un subconjunto de la tabla siguiente, *Datos completos Sedes*, es por esto que decidimos no utilizarlos durante la generación del modelo y tablas finales.

- *Datos completos Sedes*. Información más completa de las sedes. Se agregan datos de contacto (redes sociales, teléfonos adicionales), así como direcciones, códigos de región geográfica, entre otros.
- *Datos secciones Sedes*. Información sobre las secciones constitutivas de cada sede (por ejemplo, administrativas, de turismo, culturales). Cuenta con información de contacto, horarios de atención, entre otros.

A lo largo del informe se referenciará a *Sedes_Básico*, *Sedes* y *Secciones* como a estos tres últimos datasets, respectivamente. Debe notarse que la información de estos últimos datos es actual, actualizada durante 2018, y no necesariamente se corresponde a la realidad de los momentos históricos bajo estudio. No obstante, por simplicidad, en este trabajo se asume que esto es así.

1.4. Organización del trabajo y resolución

Las siguientes secciones de este informe referirán en primer lugar al manejo y análisis de los datos desde el punto de vista de su calidad y organización partiendo de las fuentes de datos (Sección 2); esto es, analizar las formas normales de las tablas, entender qué atributos serán relevantes, determinar una métrica de calidad y organizar la limpieza acorde a los requerimientos, planteo de un DER y su representación en un modelo relacional y cómo se resolvió la preparación de los datos finales. Luego se procederá a plantear las decisiones tomadas en base a todo lo previamente analizado (Sección 3) para pasar al análisis de los datos (Sección 4), lo que permitirá responder las preguntas planteadas. Por último, se cerrará el informe con conclusiones finales (Sección 5).

2. Procesamiento de Datos

2.1. Formas normales de las fuentes de datos

Las tablas proporcionadas por las fuentes de datos originales no siempre cumplían los requerimientos mínimos de regularidad.

La tabla de *Sedes_Básico* se encontraba en 2FN pero no en 3FN, dado que el valor de muchos campos se deducen de otros por medio de una relación transitiva (por ejemplo, el código ISO-2 en función del código ISO-3, o el nombre en inglés en función del nombre en español). Se decidió no trabajar con esta tabla por ser un subconjunto de *Sedes* a los fines de trabajo y por lo tanto no proseguiremos con su análisis.

La tabla de *Sedes* no se encontraba en 1FN, pues existían atributos multivaluados, aunque justamente algunos contenían información relevante para este trabajo, como es el caso de las URL del atributo *redes_sociales* (fue por esto, por ejemplo, que la decisión tomada fue procesar ese atributo y representarlo como una nueva entidad/tabla).

La tabla *Secciones* tampoco se encontraba en 1FN por contener instancias con valores multivaluados (por ejemplo, en el atributo *teléfonos_adicionales*).

La tabla *Migrantes* se encontraba en 1FN pero no en 2FN; esto es porque si se asume la clave primaria como { 'Country Origin Code', 'Country Desc Code' }, que representan códigos ISO-3 de país de origen y destino, sucede que los otros atributos que

indican el nombre de país dependen sólo parcialmente de esta clave (por ejemplo, 'Country Origin Name' depende funcionalmente sólo de 'Country Origin Code').

2.2. Análisis y mejora de la calidad de los datos

Se detallan a continuación algunos problemas observados sobre la calidad de datos de las tablas (indicando el atributo afectado y si el problema corresponde al modelo o la instancia) y se definen métricas utilizando el método *GQM* para cuantificar la magnitud del problema en esa dimensión.

a) En la tabla *Secciones*, se encontraron instancias inválidas en los atributos de la clave { sede_id, nombre }, ya que no representaban información real (típicamente entradas de prueba, por ejemplo, valores 'test') o estaban incompletas. El problema en este caso se corresponde a las instancias. Se definió la métrica M_1 dada por la Ec. 1. Se tiene que $M_1 \in [0; 1]$, siendo $M_1 = 1$ el valor óptimo.

$$M_1 = \frac{\text{Instancias válidas}}{\text{Instancias Totales}} \quad (1)$$

En la fuente original, esta métrica arrojó $M_1 = 99,42\%$. Al corregir la información y descartar las instancias inválidas, se logró tener $M_1 = 100\%$. Se utilizaron dos métodos para verificar la integridad de los datos: por un lado, un JOIN con la tabla de *Sedes*, para preservar sólo aquellas instancias que correspondieran a sedes existentes. Por otro lado, la inspección manual de las claves primarias de la tabla. Ambos métodos dieron con las mismas instancias inválidas. La inspección manual determinó que no se perdió información en la tabla corregida.

b) En la tabla *Sedes*, se encontraron instancias con múltiples entradas multivaluadas, nulas y/o inválidas para el atributo *redes_sociales*. El problema en este caso se corresponde al modelo, ya que se asume que la relación de Sede a Red Social debe vincular "0 a Muchos". Se definió la métrica M_2 dada por la Ec. 2, con el objetivo de evaluar la proporción de información válida en la tabla, donde N es la cantidad de instancias de la tabla, N_{Null} es la cantidad de instancias con un valor NULL (o equivalente) en el atributo correspondiente al url, $url_{inválidos}$ y $url_{totales}$ son la cantidad de URLs no nulos que son inválidos y la cantidad total de URLs no nulos respectivamente. Los términos cuyo denominador es 0 se computan como 1. Una URL se determinó como inválida cuando la información asociada no era suficiente para determinar su red social³. En este caso, $M_2 \in [0; 1]$. Se considera que $M_2 = 1$ es el valor óptimo de esta métrica, que corresponde a ausencia de NULLs y de urls inválidos.

$$M_2 = 1 - \frac{1}{2} \cdot \left(\frac{N_{Null}}{N} + \frac{url_{inválidos}}{url_{totales}} \right) \quad (2)$$

Aplicado a la tabla original, se obtuvo un valor de $M_2(\text{Sedes}) = 80,51\%$. Para corregirlo, se creó la entidad *Red Social*, como se describe en el Modelo Relacional

³ Por ejemplo: "@prueba" puede hacer referencia tanto a la red "Instagram", como a "Twitter/X". Se consideraron inválidos estos casos por más que haya presencia de varios "@nombre" con el mismo nombre. También se tomaron inválidos a urls propiamente dichas con espacios en su cadena de caracteres, o sin presencia de ".com". La única diferencia entre el conjunto de urls "b" es que no se contabilizaron como inválidos ni como parte del total aquellas cadenas que no remitían a una red social (por ejemplo, un email), puesto que en esta métrica se quería cuantificar la pérdida de información relacionada a las redes sociales.

(Sección 3.3), con la metodología detallada en la Sección 3.5. Al aplicar la métrica a esta relación, se obtuvo un valor óptimo de $M_2(\text{Red_Social}) = 100\%$. No obstante, por más que la métrica arroje un valor óptimo para la calidad de datos de la tabla, no toma en cuenta cuánta información se perdió al crearla, ya que si bien se descartaron datos con información incompleta, es información parcialmente válida y quizás reconstruible manualmente. Para cuantificar la proporción de esta información que se descartó, se definió la métrica M_{2b} dada por la Ec. 3, con $M_{2b} \in [0; 1]$, siendo 0 el valor óptimo.

$$M_{2b} = \frac{url_{invalidos_b}}{url_{Totales_b}} \quad (3)$$

En este caso, M_{2b} se interpreta sólo aplicada a la relación *Sedes* original. El caso $\frac{0}{0}$ no se considera relevante. Se obtuvo un valor de $M_{2b} = 13,28\%$.

c) En la tabla *Migrantes*, se encontraron instancias con información inválida en los atributos de cantidad de migrantes. En vez de un valor numérico, se encontraron strings '..', que por falta de documentación, no aportaron información. Estos datos se consideraron inválidos. Se definió la métrica M_3 dada por la Ec. 4, con $M_3 \in [0; 1]$, siendo 1 el valor óptimo. N_{arg} hace referencia a la cantidad de datos donde Argentina figura como país de origen o destino.

$$M_3 = \frac{N_{Arg\ validas}}{N_{Arg\ Totales}} \quad (4)$$

Se obtuvo un valor de $M_3 = 97,84\%$. para la tabla original y $M_3 = 100\%$ para la tabla procesada obtenida filtrando las instancias con información inválida. Inspeccionando los datos filtrados, se consideró que no se perdió información relevante para este trabajo.

2.3. DER y Modelo Relacional

En la *Figura 1* se presenta el Diagrama Entidad-Relación (DER), que representa de forma gráfica las entidades involucradas en este trabajo, así como las relaciones que las vinculan. En el caso de las entidades *Migrantes*, *Sede* y *Sección*, se decidió que representen a los propios datasets de igual nombre, mientras que con las entidades *País*, *Región* y *Red Social*, se decidió darles carácter de entidad por ser atributos importantes para los presentes objetivos y que no podían ser derivados funcionalmente por otros.

Por otro lado, en la *Figura 2* se muestra el modelo relacional, que es la traducción del DER a un formato más formal y estructurado. El modelo relacional especifica cómo se organizan los datos en tablas y cómo se relacionan entre ellas a través de claves primarias y foráneas, así como los atributos que caracterizan a cada entidad, lo que constituye un paso esencial para la implementación física de la base de datos en un sistema de gestión de bases de datos (SGBD).

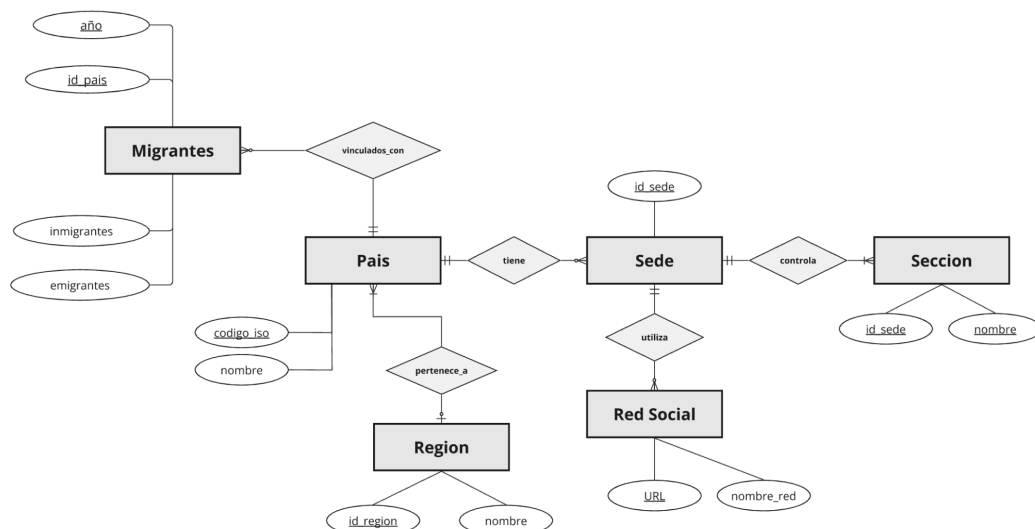


Figura 1. Diagrama Entidad-Relación.

MIGRANTES(id_pais, año, inmigrantes, emigrantes)

PAIS(codigo_iso, nombre, id_region)

REGION(id_region, nombre)

SEDE(id_sede, codigo_iso_pais)

RED_SOCIAL(URL, id_sede, nombre_red)

SECCION(nombre, id_sede)

Figura 2. Modelo relacional.

A continuación se transcriben las claves primarias elegidas para las relaciones del modelo y se da un conjunto minimal de sus dependencias funcionales. Estas claves primarias garantizan la unicidad de los registros en cada tabla, mientras que las dependencias funcionales describen las relaciones entre los atributos.

Claves primarias (PKs):

- ❖ PK(MIGRANTES) = { id_pais, año }
- ❖ PK(PAIS) = { codigo_iso }
- ❖ PK(REGION) = { id_region }
- ❖ PK(SEDE) = { id_sede }
- ❖ PK(RED_SOCIAL) = { URL }
- ❖ PK(SECCION) = { nombre, id_sede }

Conjunto minimal de DFs:

- ❖ DF_MIGRANTES : { id_pais, año } → { inmigrantes, emigrantes }
- ❖ DF_PAIS : { codigo_iso } → { nombre, id_region }
- ❖ DF_REGION : { id_region } → { nombre }
- ❖ DF_SEDE : { id_sede } → { codigo_iso_pais }
- ❖ DF_RED_SOCIAL : { URL } → { nombre_red, id_sede }

3. Proceso de importación y decisiones tomadas

Una vez analizadas las fuentes de datos, presentados los pasos requeridos de limpieza y calidad de datos, y la generación de un modelo que facilite la manipulación y

análisis de los datos, es necesario definir cómo se pasará desde los datos originales a los que representen el modelo.

Se desarrolló un script en python que, tomando los contenidos de los tres datasets importantes (Migrantes, Sedes y Secciones), remueve los registros inválidos o los cuales no interesan a los fines del trabajo. Los pasos se describen en detalle en el archivo *Script - Creacion de tablas.py*, en donde hay seis etapas, una para cada generación de las seis entidades/tablas del DER adoptado.

Las precondiciones para correr este script son que se cuente con los tres datasets requeridos en la carpeta Materiales/ y que se mueva manualmente una celda en particular (de lo contrario pandas no puede leer un csv). Esto está todo detallado en los comentarios.

3.1. Migrantes

Toda la información de esta tabla se extrajo del dataset *Migrantes* (*datos_migrantes.csv*). De estos datos, se exportó sólo un subconjunto de las instancias y atributos de la relación. Se descartaron las instancias divididas por género, conservando sólo aquellas que reportan la cantidad total de migrantes (suma de ambos géneros), dado que no estaba en los objetivos esta diferenciación. Luego, se conservaron únicamente las instancias donde Argentina figuraba como país de origen o destino. Se descartaron las instancias que tuvieran a “Refugiados” como país. Aquellas entradas con valores numéricos inválidos (problema de calidad de datos descrito en la Sección 2.2) se asumieron y reemplazaron por “0”. Se consolidaron los datos en un único DataFrame, con tests para asegurar que todos los valores se integraran correctamente.

3.2 Regiones

Se obtuvo la información de la región de cada país a partir de la tabla *Sedes* (*lista-sedes-datos.csv*). Por esto, solo se asignó una región geográfica a aquellos países que tuvieran al menos una sede, que son los que figuran en esa tabla.

3.3 Países

Se combinaron los datos de los datasets *Migrantes* (*datos_migrantes.csv*) y *Sedes* (*lista-sedes-datos.csv*) para obtener la lista completa de nombre de países y sus códigos en formato ISO-3. Se corrigió un error en el código de Gran Bretaña, que figuraba como “GRB” (en lugar de “GBR”), para mantener uniformidad en los códigos. Se mantuvieron los nombres en inglés, ya que es más común encontrar información en este idioma. Se aseguró de incluir todos los países que figuraran en al menos una de las tablas para priorizar la completitud de códigos y nombres de países. Por medio de un LEFT JOIN, se asignó donde fue posible un código de región a cada país. Como se aprecia en la Sección 2.3, se tomó la decisión de permitir que no todos los países tuvieran una región asignada (clave foránea *id_región*), por lo que surgieron muchos valores NULL, pero sin embargo se respetó esta decisión a los efectos de priorizar que estuvieran todos países. En caso de haber sido posible, se habría descargado alguna base de datos completa y estandarizada.

3.4 Sedes

Se importó el conjunto de datos *Sedes* (*lista-sedes-datos.csv*), conservando solo los atributos relevantes a los efectos de los ejercicios: id de la sede y código ISO 3 del país.

3.5 Secciones

Se importó el conjunto de datos *Secciones* (*lista-secciones.csv*) y se filtraron las columnas necesarias dejando sólo atributos de clave primaria: código de la sede a la cuál pertenece y nombre de la sección. Se filtraron las instancias inválidas (descrito en la Sección 2.2).

3.6 Redes Sociales

Se extrajo la información de la tabla *Sedes* (*lista-sedes-datos.csv*). Como se detalla en el pie de la Sección 2.2, se mantuvieron sólo aquellos enlaces que resultaran no ambiguos o bien formados y se identificó la red social correspondiente a partir del texto del enlace (el universo de redes permitidas se redujo a Facebook, Twitter, Instagram, LinkedIn, Flickr, Youtube).

4. Análisis de datos, reportes, resultados y discusión

Se presentan a continuación los resultados y su discusión. Aquellas tablas que, por espacio, no se encuentran completas en el cuerpo del texto, pueden encontrarse en el Anexo.

4.1 Flujo de migrantes en el año 2000 y cantidad de sedes.

Para responder a la pregunta de investigación, se buscó determinar si existe una relación entre el flujo migratorio entre Argentina y los demás países y su cantidad de sedes en el extranjero. En todos los casos se estudió la situación del año 2000.

En la Tabla 1 se reúnen por país la cantidad de sedes argentinas, la cantidad de secciones promedio que tiene cada sede, y la diferencia neta de migrantes entre Argentina y ese país (calculado como *Migrantes extranjeros en Argentina – Migrantes argentinos en el exterior*).

Tabla 1. Cantidad de sedes, promedio de secciones por sede y flujo migratorio neto entre cada país con Argentina para el año 2000.

País	Cantidad_Sedes	Secciones_Promedio	Flujo Migratorio Neto
REPÚBLICA FEDERATIVA DEL BRASIL	11	1.6	8871
Estados Unidos de América	9	3.3	-122872
REPÚBLICA ORIENTAL DEL URUGUAY	8	0.5	94023
ESTADO PLURINACIONAL DE BOLIVIA	7	2.3	209601
REINO DE ESPAÑA	7	2.3	53281
...

No se logra observar, por los resultados obtenidos en la Tabla 1, una relación lineal entre ambas variables. Para dar más sustento a esta hipótesis, se realizó un análisis de correlación lineal entre ambas variables. El coeficiente de correlación lineal obtenido fue $R^2 = 0,31$, lo que se considera bajo en este contexto. Este valor sugiere que, aunque exista alguna relación lineal entre ellas, es débil y no indica una dependencia fuerte, por lo que puede sugerirse que los cambios en la cantidad de sedes no están asociados de manera significativa, por lo menos linealmente, con los cambios en el número del flujo migratorio neto. Los datos y el modelo de regresión ajustado pueden observarse gráficamente en la Figura 3. Sin embargo, no hay que descartar la posibilidad de que exista otro tipo de dependencia no lineal que no haya sido explorada en este informe.

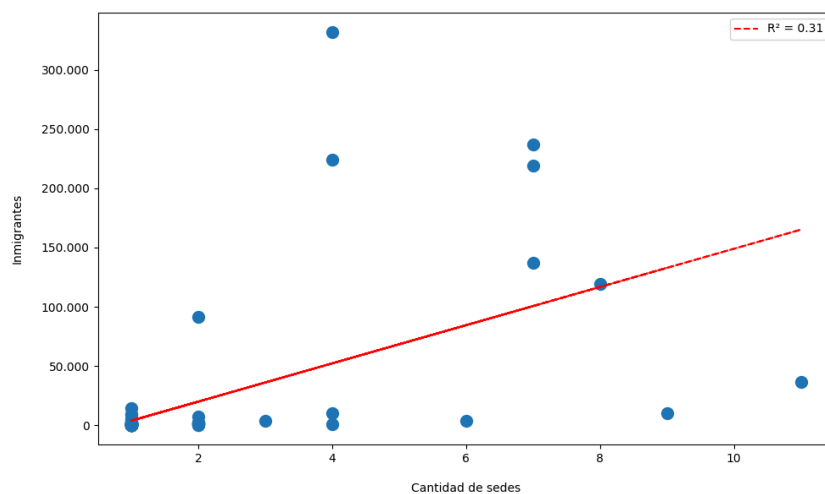


Figura 3. Relación entre migrantes en Argentina y la cantidad de establecimientos que representan a Argentina para cada país.

Se observa que, a grandes rasgos, los países pueden dividirse en dos grupos: aquellos de “bajo flujo migratorio”, y aquellos de “alto flujo”. Por otra parte, no resulta evidente que el flujo neto es una variable representativa de lo que se quiere estudiar, dado que el flujo entrante y saliente pueden ser similares. Es por esto que, además del flujo neto, se hicieron también dos análisis adicionales: El del flujo migratorio entrante por un lado, y el del flujo migratorio saliente por otro (definidos respectivamente como la cantidad de migrantes extranjeros en Argentina, y la cantidad de migrantes argentinos en el exterior). Esta información se presenta en las Figuras 4 y 5.

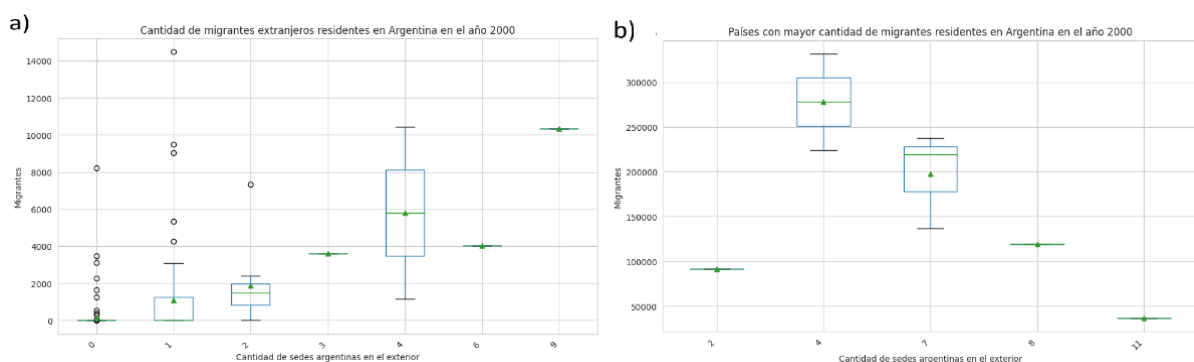


Figura 4. Relación entre la cantidad de migrantes extranjeros en Argentina y la cantidad de establecimientos que representan a Argentina en ese país, para el año 2000. a) Flujos menores a 35.000 b) Flujos mayores a 35.000

En la Figura 4 se estudió el caso del flujo entrante. En este contexto, se estableció como criterio para dividir a los países el superar un flujo umbral de 35.000 migrantes. Se obtuvo así ocho países con alto volumen de inmigración en Argentina, de los cuales la mayoría son países de la región (en orden decreciente: 1-Paraguay, 2-Bolivia, 4-Chile, 6-Uruguay, 7-Perú y 8-Brasil). Los restantes pertenecen a Europa (3-Italia y 5-España respectivamente). Para estos, la relación entre cantidad de sedes e inmigración parece ser inversamente proporcional, como se ve a partir de la distribución de medianas en la figura 4.b. En cambio, para los países con baja cantidad de inmigración, la relación parece ser directamente proporcional, como se aprecia en la Figura 4.a a partir de las medias.

Se consideró relevante el caso del flujo saliente y la cantidad de sedes en el extranjero, dado que es esperado que las sedes en el exterior estén en relación con la cantidad de argentinos en el exterior. Esta relación se presenta en la Figura 5, donde la distribución de las medianas sugiere una tendencia directamente proporcional entre las variables.

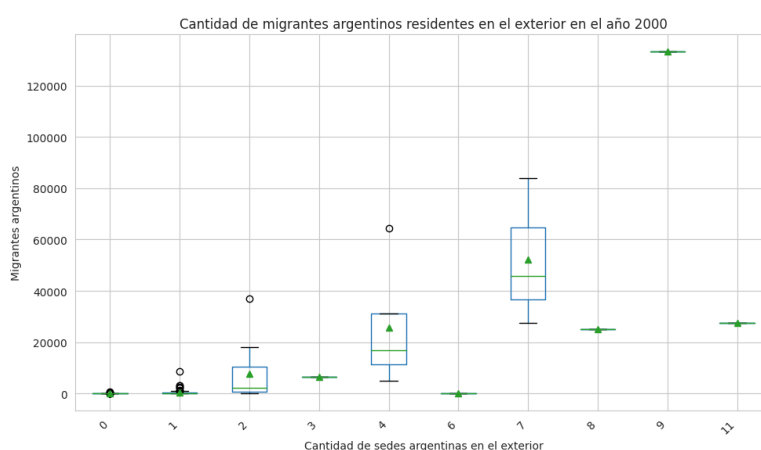


Figura 5. Relación entre la cantidad de migrantes argentinos en el extranjero y la cantidad de establecimientos que representan a Argentina en ese país, para el año 2000.

A modo de comentario, debe notarse que estas medidas son representativas del volumen del flujo migratorio, pero no dan cuenta de que tanto flujo “relativo” hay de un país a otro. Para poder evaluar esto, debería también tenerse en cuenta la población total, para poder así estimar la proporción del flujo en relación a este número. No se dispuso de acceso a esa información en esta instancia, por lo que es algo que queda como mejora futura.

4.2 Flujo migratorio y cantidad de sedes según región geográfica

En la Tabla 2 se reúne el promedio por país de migrantes argentinos en el exterior, agrupado según la región geográfica, para aquellos países que cuentan con al menos una sede argentina y durante el año 2000. Esto se calculó como la cantidad de migrantes en esa región, sobre la cantidad de países de la región.

Tabla 2. Cantidad de migrantes argentinos en el exterior, agrupados por región geográfica y ordenados de manera decreciente

Región Geográfica	Países con sedes argentinas	Flujo saliente de ARG promedio
AMÉRICA DEL NORTE	3	51103.33
AMÉRICA DEL SUR	11	18889.82
EUROPA OCCIDENTAL	18	7648.22
OCEANÍA	2	5565.50
ASIA	23	1917.96
AMÉRICA CENTRAL Y CARIBE	14	446.36
ÁFRICA SUBSAHARIANA	7	260.29
EUROPA CENTRAL Y ORIENTAL	8	90.25
ÁFRICA DEL NORTE Y CERCAÑO ORIENTE	5	68.00

En la Tabla 2 se observa la información geográfica de diferentes regiones del mundo y el flujo migratorio saliente de Argentina promedio.

Por otra parte, se exploró la distribución de las sedes argentinas en el exterior según las regiones geográficas. Se halló que esta es desigual. Las regiones más representadas son América del Sur, Europa Occidental y Asia, lo que indica una concentración significativa en estas áreas (Figura 4).

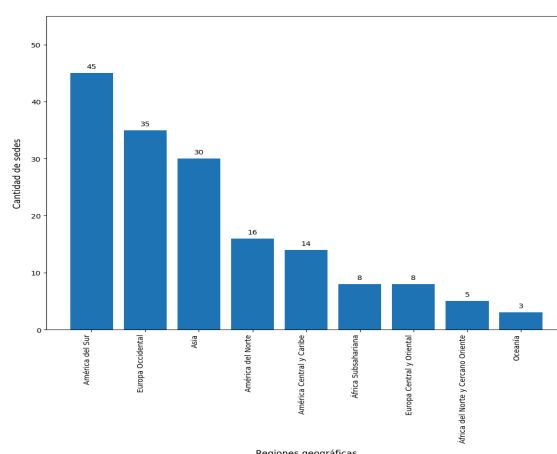


Figura 6. Cantidad de sedes (representaciones argentinas en el exterior) por región geográfica

América del Sur lidera con 45 sedes, Europa Occidental (35) y Asia (30) siguen de cerca, mientras que, en el otro extremo, Oceanía tiene solo 3, y regiones como África del Norte y Cercano Oriente apenas 5.

Se analizó por región el promedio del flujo migratorio neto a lo largo de los años 1960, 1970, 1980, 1990 y 2000. El flujo migratorio neto se define como en la Sección 4.1. Se segmentó la información según el volumen de migrantes neto, como se ve en la Figura 7. En el panel izquierdo, que representa las regiones con mayor flujo migratorio, se observa que América del Sur presenta un flujo migratorio neto positivo y América del Norte uno negativo, mayormente acaparado por Estados Unidos. En el panel derecho se observa cómo, a su vez, varían los flujos netos, siendo el de mayor valor la mediana de Europa Central y Oriental.

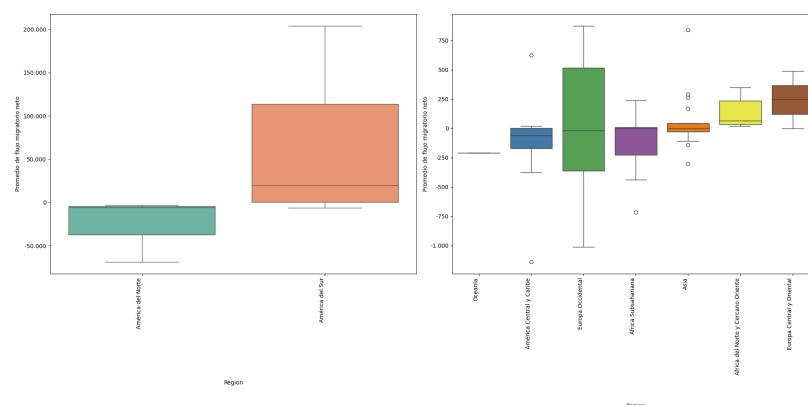


Figura 7. Flujo migratorio neto entre Argentina y los demás países, según región geográfica. Se dividió en dos paneles para una mejor exposición de los boxplots. Cada boxplot fue ordenado por las medianas del flujo migratorio neto de menor a mayor, y se utilizaron colores diferenciados para cada boxplot con el fin de mejorar la claridad visual.

4.3 Flujo Migratorio y Redes Sociales

Se estudió también si existe algún tipo de relación entre la presencia de las sedes argentinas en redes sociales y el flujo migratorio entre esos países. Para eso, se contabilizó la cantidad de tipos de redes sociales diferentes con que cuentan las sedes argentinas en ese país. Esto se ve parcialmente en la Tabla 3. Para esto, se recopiló la información de redes de cada sede (según lo exhibido en la Sección 2.2.b), como se muestra en la Tabla 4.

Tabla 3. Cantidad de tipos de redes sociales con que cuentan las sedes argentinas en cada país, ordenadas de manera decreciente

País	Cantidad_Tipo_Red
Estados Unidos de América	6
Reino de Bélgica	6
CANADÁ	5
REINO UNIDO DE GRAN BRETAÑA E IRLANDA D...	5
Confederación Suiza	4
...	...

Tabla 4. Cantidad de migrantes argentinos en el exterior, agrupados por región geográfica y ordenados de manera decreciente.

País	Sede	Red Social	URL
AUSTRALIA	CSIDN	Facebook	https://www.facebook.com/ArgentinaEnSidney/
AUSTRALIA	EAUST	Facebook	https://www.facebook.com/ArgentinaEnAustralia/
AUSTRALIA	EAUST	Twitter	https://twitter.com/ARGinAustralia
BARBADOS	EBARB	Facebook	https://www.facebook.com/ArgentinaEnBarbados/
CANADÁ	CTORO	Facebook	https://www.facebook.com/ArgentinaEnToronto/
...

Si bien el país con más cantidad de redes coincide con aquel con mayor cantidad de migrantes argentinos, se concluyó que esto se trata de un caso especial y que no es representativo de la relación. En general, estas variables se determinaron independientes.

5. Conclusiones

No se observó una relación lineal directa entre la cantidad de sedes argentinas en los países extranjeros y el flujo migratorio neto con dichos países. Tanto el análisis por país como por regiones no permitió confirmar la hipótesis. Sin embargo, no hay que descartar la posibilidad de que exista otro tipo de dependencia no lineal no considerada en esta instancia. No obstante, sí se observó una relación entre la cantidad de sedes y los flujos al dividirlos en flujos entrantes y salientes respectivamente. Las tendencias se observaron en función del volumen del flujo migratorio. Es posible, por otro lado, que la función de las sedes influya en las relaciones. Se sugiere explorar en futuras investigaciones si las secciones de las sedes (por ejemplo, asistencia directa, fines diplomáticos, o comerciales) son un factor relevante.

Se concluyó que la mayor cantidad de flujo migrante entrante es de países limítrofes o geográficamente cercanos, a la vez que de países de Europa cercanos en la lengua. Estos forman un conjunto de ocho países que sobresalen en relación a los otros. Para el flujo saliente, estos son mayormente los mismos (ocho de nueve).

La concentración de sedes argentinas en América del Sur, Europa Occidental y Asia puede reflejar vínculos históricos, económicos y culturales más estrechos entre Argentina y estos continentes, así como la presencia de comunidades argentinas en el extranjero que facilitan la creación de sedes. En contraste, se observa que regiones como África del Norte y el Cercano Oriente, así como Oceanía, tienen una representación notablemente baja en cuanto a la cantidad de sedes. Esta disparidad puede deberse a varios factores, como limitaciones económicas, menores flujos migratorios hacia Argentina desde estas áreas, o una falta de intereses estratégicos en establecer sedes en estas regiones. Se sugiere que la estrategia de expansión internacional de Argentina se ha centrado principalmente en regiones donde hay una mayor interacción y demanda; sin embargo, sacar conclusiones de este tipo puede ser apresurado. Se puede concluir que, en lugar de extrapolar conclusiones de todo el conjunto de datos, se debe segmentarlo en grupos en relación al volumen de sus flujos, para luego determinar posibles causas de esas tendencias.

6. Referencias

[1] **World Bank Group. (2011, June 28).** Global Bilateral Migration. DataBank. <https://databank.worldbank.org/source/global-bilateral-migration#>

[2] **Ministerio de Relaciones Exteriores y Culto. (2018, August 29).** Representaciones argentinas en el exterior. Representaciones Argentinas. <https://datos.gob.ar/dataset/exterior-representaciones-argentinas>