



**DEPARTAMENTO
DE COMPUTACION**

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 1: “Análisis de la relación entre el flujo migratorio y la cantidad de sedes argentinas en un país”.

Laboratorio de Datos.

Grupo “Pescado Rabioso”

Integrante	LU	Correo electrónico
Souto, Sebastian Manuel	43/23	soutosebastianma@gmail.com
Sanza, Gian Lucca	149/23	gianluccalord723@gmail.com
Goldfarb, Bruno	1164/23	bgoldfarb2003@gmail.com



1. Resumen

En este trabajo se pretende analizar la relación entre el flujo migratorio de cada país y la cantidad de sedes argentinas establecidas en él. Se partió de bases de datos sobre migraciones y sobre sedes argentinas. Luego se organizaron los datos utilizando métodos adecuados para el análisis de datos. Una vez que se obtuvieron los datos limpios se procedió a analizar sus relaciones utilizando SQL y herramientas de visualización para una mejor comprensión.

Una vez finalizado el análisis se notó que no hay una relación estrecha entre el flujo y la cantidad de sedes. Sin embargo, los resultados obtenidos nos ayudan a comprender la lógica utilizada para distribuir las sedes argentinas en los distintos países, y cómo varía esta distribución alrededor de las regiones geográficas y de los países cercanos a Argentina.

2. Introducción

El flujo migratorio es un fenómeno que refleja el movimiento de personas entre diferentes regiones y países. Se ve impulsado por diferentes motivaciones, como la búsqueda de mejores oportunidades económicas o el deseo de conocer nuevas culturas a través de viajes. Se calcula a partir de la resta entre el número de inmigraciones y emigraciones.

Las sedes argentinas alrededor del mundo son las fuentes principales de promoción del interés nacional. Tienen un significado comercial, diplomático, así como de ayuda al argentino que se encuentra en el exterior. Al ser el lugar a donde se acude cuando uno desea inmigrar o emigrar, analizaremos la importancia de estas, y su relación con el flujo migratorio de los países donde se encuentran las sedes hacia Argentina y el flujo neto de estos países.

En este trabajo se utilizarán herramientas y métodos adecuados para analizar la relación entre el flujo migratorio de cada país y la cantidad de sedes en el exterior que tiene Argentina en dicho país a partir de datos de flujo migratorio entre países y datos de las sedes Argentinas obtenidos del Ministerio de Relaciones Exteriores, Comercio Internacional y Culto.

Este análisis resulta relevante ya que se podrá dar cuenta si las sedes Argentinas están distribuidas en función del flujo migratorio y si su distribución es eficiente respecto al flujo.

Por eficiente al flujo vamos a intentar entender a que se traduce un flujo positivo o negativo. Nos invita a preguntar si es que el flujo es la razón más importante para tener la sede, o si hay otros factores, deducibles en los datos, que nos den un motivo de por el cual Argentina posee más sedes en un lugar u otro. Investigaremos si una sede más compleja, es decir, una sede con mayor separación en secciones y sitios web que la representen, está relacionada con un flujo migratorio mayor o menor, o si es que no están relacionados de ninguna manera.

Resulta importante efectuar el análisis ya que las migraciones de los argentinos hacia otros países, así como sus retiradas, pueden verse afectadas de muchas maneras a lo largo de un viaje. Es por eso que entender la distribución de las sedes y relacionarlas con el flujo podría solucionar problemas de muchos argentinos en un futuro. Si por ejemplo, la cantidad de sedes

tiene una relación con la inmigración de ese país hacia Argentina, se podría promover un aumento de sedes en países donde conocemos que hay un foco de inmigración.

Para poder responder las preguntas que impulsan el desarrollo de este trabajo, se van a importar los datos necesarios y armar un DER y modelo relacional con estos datos. Luego se realizarán diferentes consultas SQL y se utilizarán herramientas de visualización para realizar el análisis correspondiente.

3. Procesamiento de datos

Para comenzar a trabajar se importaron las siguientes bases de datos:

Migraciones: Información de flujos migratorios entre países.

[/https://drive.google.com/file/d/1zYjV6YN3ZPGpGkWWW_UqXp0-wnB0HMcXt/view](https://drive.google.com/file/d/1zYjV6YN3ZPGpGkWWW_UqXp0-wnB0HMcXt/view)

Representaciones Argentinas

<https://datos.gob.ar/dataset/exterior-representaciones-argentinas>

En dicho sitio se importaron datos completos de las sedes y datos completos de las secciones de las sedes.

Luego se determinó la forma normal de las fuentes de datos originales:

Respecto a la tabla de migraciones, se encuentra en 1FN, pues no tiene atributos multivaluados. Está en 2FN, ya que ningún dato depende parcialmente de alguna ck. Además está en 3FN, ya que las transitividades de los datos pertenecen a claves candidatas. Es decir, no hay transitividad entre claves y atributos.

Respecto a la tabla de datos completos de las sedes, no se encuentra en 1FN pues tiene columnas con atributos multivaluados, por ejemplo “Redes Sociales”, que enumera todos los sitios web de la sede en diferentes redes sociales. Como no está en 1FN tampoco está en 2FN ni en 3FN.

Por último en la tabla de datos completos de las secciones de las sedes sucede lo mismo que en la tabla anterior, la columna “Números de Teléfonos Adicionales” tiene atributos multivaluados, por ende no está en 1FN.

A la hora de importar los datos, la tabla de datos completos presentaba un error que no permitía abrir el archivo utilizando pandas. La línea de datos correspondiente a la República De Chile utilizaba un formato para una de sus celdas que terminaba agregando columnas de datos inexistentes. Para eso modificamos dicha celda, y replicamos el mismo formato que utilizan todas las demás celdas.

Utilizando el sistema de métricas GQM, se apuntó a mejorar la calidad de los datos.

Comenzando con la tabla datos completos, se observó que muchas de las filas correspondientes a redes_sociales están vacías. Poniendo en práctica la métrica resulta que redes sociales vacías / cantidad total de sedes = $38/164 \sim 23.2\%$. El atributo de calidad afectado es la completitud y corresponde a instancia del modelo.

En la tabla de datos básicos, los nombres de los países contenían irregularidades en la forma en la que estaban escritos. Utilizando la métrica, se obtiene: Datos escritos irregularmente (en

minúscula) / cantidad total de nombres de países = 3/164 ~ 1.82% de los países de la tabla. El atributo de calidad afectado es la consistencia y corresponde a instancia del modelo. Para la tabla de migraciones se planteó como objetivo que el dato correspondiente a la columna “2000 [2000]” sea distinto de ‘..’, pues con estos valores no se puede operar como si fueran números. Utilizando la métrica se obtiene: Filas con valores igual a ‘..’/ cantidad de filas = 7548/160781 ~ 4.69%. Aunque parezca una proporción baja es información que se pierde o que no se podrá utilizar para realizar los análisis. El atributo de la calidad afectado en este caso es la completitud, ya que se puede asumir que tener como valor ‘..’ donde debería haber un número significa que no se tiene presente el valor de ese dato. Y el problema corresponde a instancia del modelo.

Luego se desarrolló el DER y el modelo relacional teniendo en cuenta que información va ser útil para el análisis que se quiere realizar.

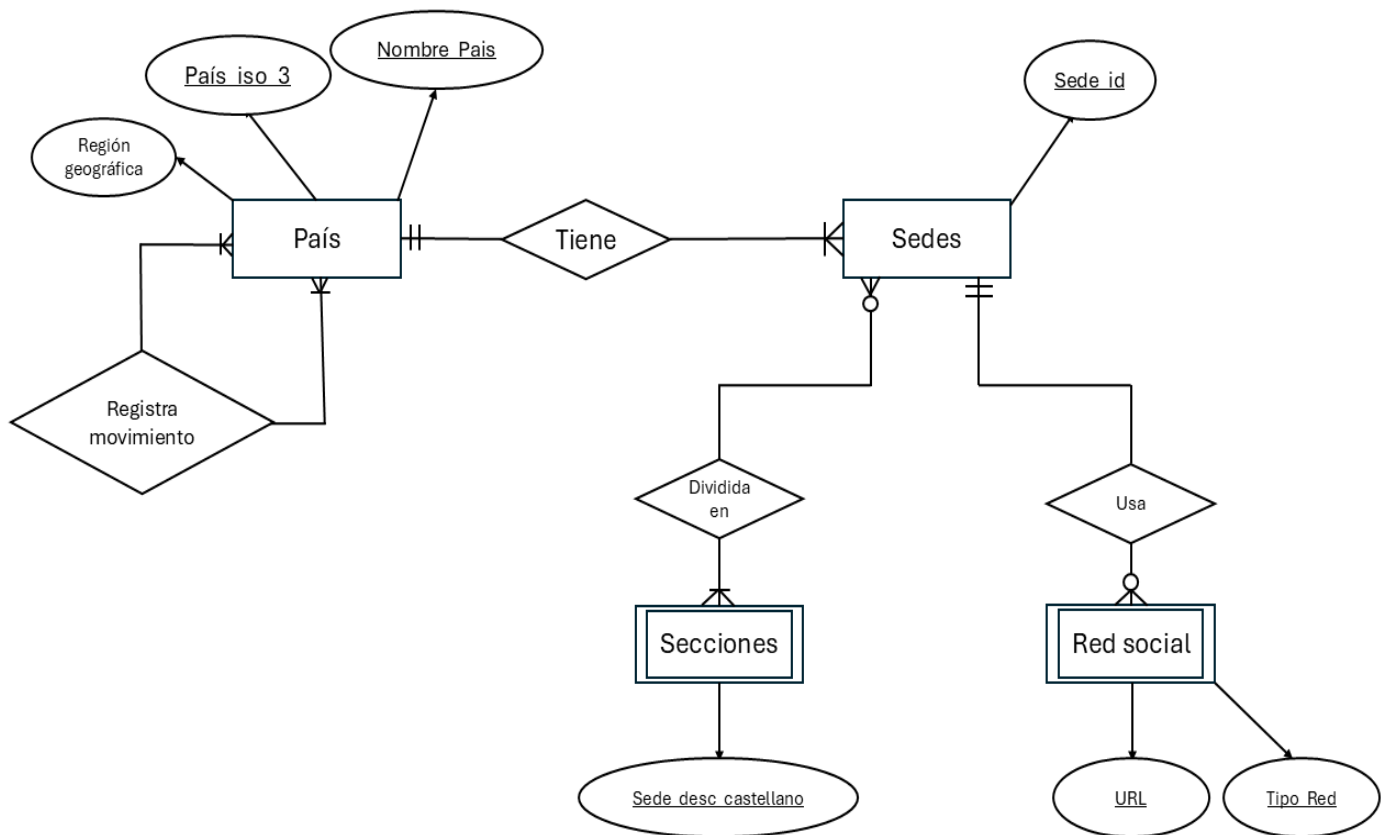


Figura 1. Diagrama Entidad Relación del modelo de datos elaborado.



Dependencias funcionales

Pais:

$\text{pais_iso_3} \longrightarrow \{\text{región_geográfica}, \text{nombre_pais}\}$

Sede:

$\text{Sede_id} \longrightarrow \text{Pais}$

Redes sociales:

$\text{URL} \longrightarrow \{\text{Sede_id}, \text{Tipo_red}\}$

Registra movimiento:

$\{\text{Pais_origen}, \text{Pais_destino}, \text{Década}\} \longrightarrow \text{Cantidad}$

Figura 2. Esquema del DER en el modelo relacional.

El esquema mostrado en la Figura 2 demuestra que las tablas están en 3FN, no hay atributos multivaluados, se cumple la dependencia completa con las claves primarias y no hay dependencias transitivas.

4. Decisiones Tomadas

Las fuentes de datos originales fueron transformadas para obtener un modelo de datos óptimo para poder trabajar con el mismo y lograr el análisis deseado. De la tabla de datos completos de las sedes sólo se conservó “sede id”, información del país de dicha sede (“nombre_pais” y “región geográfica”) y la columna de “Redes Sociales”. Respecto a los datos completos de las secciones de las sedes, se mantuvo la columna de “sed_desc castellano” y “sede id”, ya que solo resulta relevante la cantidad de secciones.

Para construir la identidad País se utilizaron los países que tienen sedes argentinas, de la tabla de datos completos se utilizaron “pais_iso_3”, “nombre_pais” y “región_geográfica” y se optó por tomar como clave primaria al atributo “pais_iso_3”.

Para construir la tabla de ‘registra_movimiento’ se utilizó la tabla de migraciones. Se descartaron aquellas filas que tenían como valor ‘.’ para columnas donde debían ir datos numéricos, ya que con estos datos no se puede trabajar. Y también las filas que estaban divididas por género, quedándonos con las filas que tenían ‘total’ como valor en la columna de género.

Los datos de “Redes Sociales” extraídos de la tabla de datos completos de las sedes, fueron transformados para que las tablas resultantes cumplan con las formas normales. Originalmente en dicha columna los valores eran una lista, donde cada red social estaba separada por ‘\’. Mediante consultas SQL se logró dividir estos valores en atributos atómicos para cumplir con la 1FN. Además respecto a redes sociales se decidió no considerar los URL que no estaban en un formato correcto, por ejemplo ‘@ArgEnRoma’ pues no es posible saber a qué red social pertenece esta cuenta.

5. Análisis de Datos

Primero se observó la cantidad de sedes, el promedio de secciones por sedes y el flujo migratorio neto en el año 2000 por cada país:

nombre_pais	cant_sedes	secciones_pro medio	flujo_mundo_2000
REPÚBLICA FEDERATIVA DEL BRASIL	11	1.6363	-226752.0
ESTADOS UNIDOS DE AMÉRICA	9	3.33333	3.26312e+07
REPÚBLICA ORIENTAL DEL URUGUAY	8	0.5	-147137.0
ESTADO PLURINACIONAL DE BOLIVIA	7	2.14286	-272782.0
REINO DE ESPAÑA	7	2.28571	641977.0

...
-----	-----	-----	-----

Tabla 1. Relación entre cantidad de sedes, secciones promedio y flujo migratorio.

Analizando los datos de la Tabla 1 , se observa que Brasil tiene 2 sedes más que Estados Unidos pero su flujo migratorio es menor. Bolivia por su lado tiene la misma cantidad de sedes que España, sin embargo, el flujo de estos dos países no es similar.

Se observa que la cantidad de secciones no necesariamente tiene una relación con el flujo tampoco. Un buen ejemplo para respaldar este análisis es Mozambique:

nombre_pais	cant_sedes	secciones_promedio	flujo_mundo_2000
...
REPÚBLICA DE MOZAMBIQUE	1	8.0	-390061
...

Tabla 2. Muestra de República de Mozambique extraído de la Tabla 1.

Mientras que en Rusia...

nombre_pais	cant_sedes	secciones_promedio	flujo mundo _2000
...
FEDERACIÓN DE RUSIA	1	13.0	1.67538e+06
...

Tabla 3. Muestra de Rusia extraído de la Tabla 1.

Aquí se ve como teniendo un número similar de secciones y el mismo número de sedes, el flujo no tiene relación entre Rusia y Mozambique.

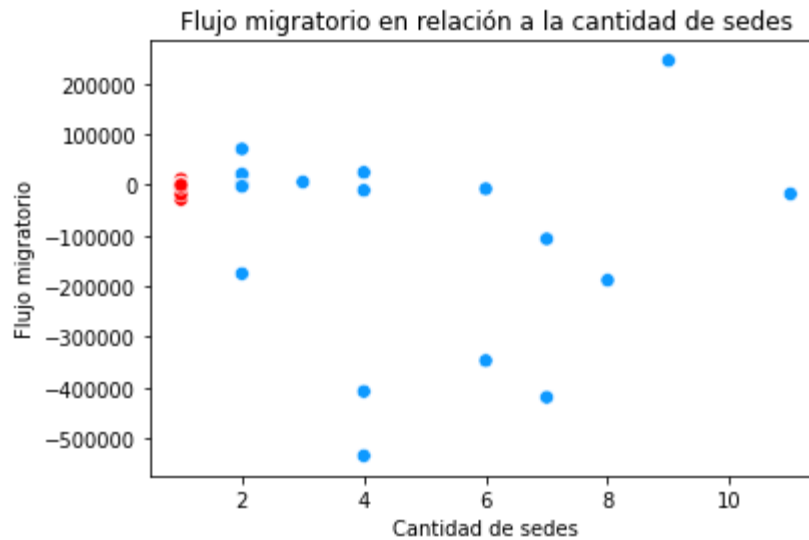
Luego, se agruparon los países según su región geográfica y se determinó la cantidad de países con sedes argentinas en dicha región y el promedio del flujo migratorio con Argentina de estos países en el año 2000.

región_geográfica	países_con_sedes	flujo_promedio_arg_2000
AMÉRICA DEL NORTE	3	46072.3
OCEANÍA	2	5202.5
ASIA	23	1175.22
ÁFRICA SUBSAHARIANA	7	260.286
AMÉRICA CENTRAL Y CARIBE	14	92.0714
...

Tabla 4. División de los países por región geográfica, cantidad de países con sedes y flujo promedio con Argentina.

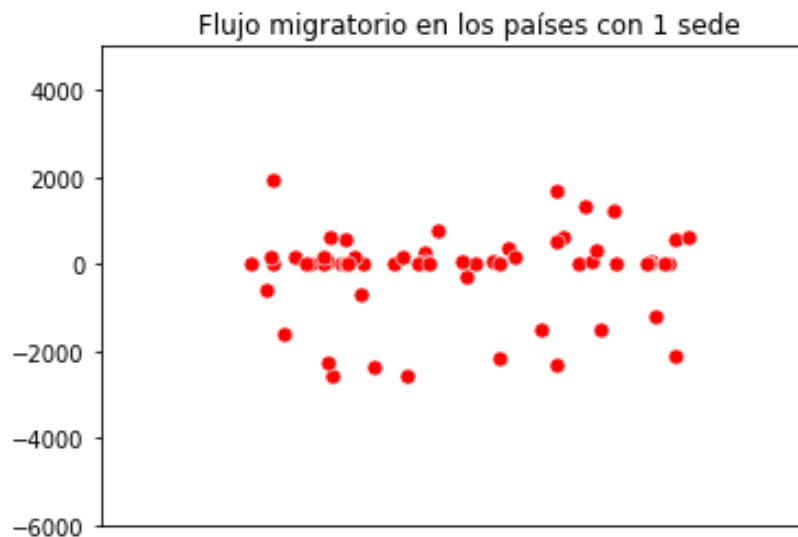
A partir de la Tabla 4, se puede descartar la idea que las regiones con más sedes son aquellas donde el flujo migratorio promedio es mayor. La región con mayor flujo migratorio promedio es América del Norte, esto se debe a que Estados Unidos aumenta este valor y la cantidad de países en dicha región es menor que en otras.

Con el objetivo de profundizar lo que conocemos sobre la relación entre el flujo migratorio de los países hacia Argentina en el año 2000 y la cantidad de sedes se realizó el siguiente gráfico de dispersión:



En la Figura 3 no parece que exista una relación de proporcionalidad entre el flujo migratorio y la cantidad de sedes. Se puede ver una ligera tendencia a un flujo negativo, lo cuál implica mayor cantidad de emigrantes que inmigrantes, pero no vemos relación con la cantidad de sedes.

Si nos enfocamos en los países que tienen 1 sede y hacemos zoom en el gráfico nos encontramos con lo siguiente:



Además de poder ver que la mayoría de países se encuentran en esta categoría, podemos notar que el flujo migratorio es relativamente bajo, lo que no necesariamente implica baja cantidad de migrantes sino que también puede significar un equilibrio entre inmigrantes y emigrantes.

Siguiendo por la línea del análisis respecto a la región geográfica, se realizó el siguiente gráfico de barras.

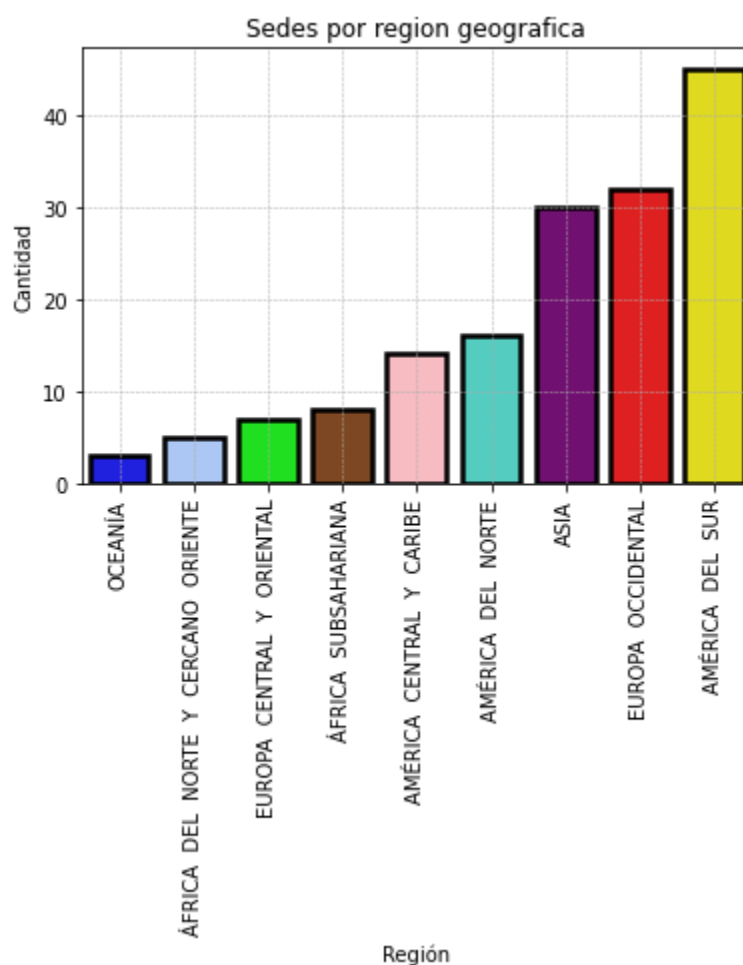


Figura 5. Gráfico de barras de cantidad de sedes por región geográfica

Se observa que América del Sur es la región que más sedes argentinas tiene, seguido por Europa occidental, Asia y América del Norte.

Para entender mejor el flujo migratorio analizaremos cómo varía este a lo largo de los años. Para eso, se desarrolló un gráfico de boxplot separado por regiones con el flujo migratorio de los países hacia Argentina de la década del 60 hasta la década del 2000.

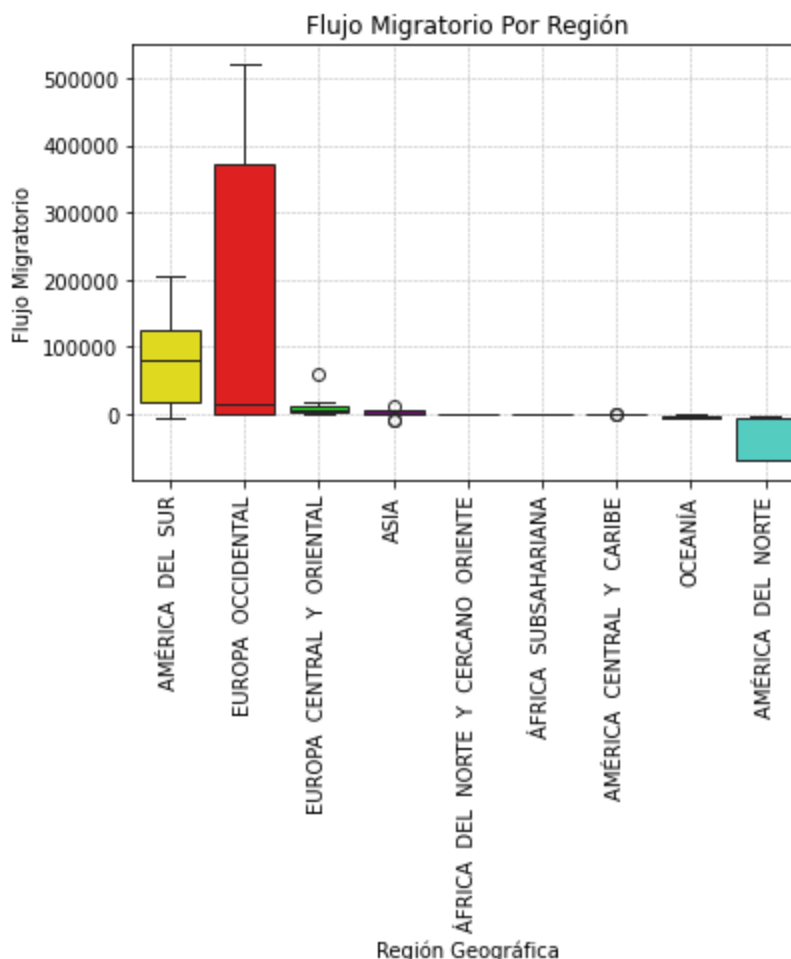


Figura 6. Boxplot, por cada región geográfica, el promedio de flujo migratorio con Argentina desde 1960 hasta la década del 2000.

En un primer lugar salta a la vista que las dos regiones con mayor variación (y mayor flujo promedio) son también aquellas que mayor cantidad de sedes argentinas tienen. Sin embargo, podemos observar en el gráfico de cantidad de sedes (figura 5) que Asia y Europa Occidental tienen una cantidad similar de sedes (30 y 32 respectivamente), pero cuando observamos el gráfico de flujo migratorio, (figura 6) mientras que Europa Occidental tiene la mayor variación y los países con mayor flujo, el otro es apenas distinguible en el gráfico por prácticamente carecer de ambas. Otro caso a comparar es el de Oceanía y Asia, en el que ambas regiones tienen una variación de flujo migratorio similar pero el segundo tiene 10 veces la cantidad de sedes del primero.

Podríamos pensar que Asia es un caso particular, tal vez por ser el continente más lejano tanto en cultura como en distancia, pero algo similar a la comparación entre sedes y variación de Asia sucede entre América del Norte y América Central y Caribe, que tienen respectivamente 16 y 14 sedes pero la diferencia en variación y cantidad de flujo es significativa.

Estas observaciones nos llevan a creer que no existe una relación entre la cantidad de sedes y el flujo migratorio, ni la variación del mismo.

También se quiere analizar los medios de comunicación que utiliza cada sede. Para esto, se confeccionaron las siguientes tablas:

nombre_pais	cant_redes
AUSTRALIA	2
BARBADOS	1
CANADÁ	5
Confederación Suiza	4
EMIRATOS ÁRABES UNIDOS	1
ESTADO DE ISRAEL	2
...	...

Tabla 5. Cantidad de Redes Sociales utilizadas por país.

A partir de la Tabla 5, se nota que en general las sedes optan por utilizar las mismas redes sociales. El valor máximo alcanzado por un país es 6 y lo alcanzan Bélgica y Estados Unidos.

País	Sede	red_social	URL
AUSTRALIA	EAUST	Twitter	https://twitter.com/ARGinAustralia
AUSTRALIA	EAUST	Facebook	https://www.facebook.com/ArgentinaEnAustralia/
AUSTRALIA	CSIDN	Facebook	https://www.facebook.com/ArgentinaEnSidney/
BARBADOS	EBARB	Facebook	https://www.facebook.com/ArgentinaEnBarbados/
CANADÁ	ECANA	Twitter	https://twitter.com/ArginCanada

Tabla 6. Red social con URL de cada sede.

6. Conclusiones

A partir del análisis realizado se puede llegar a una conclusión final sobre el objetivo planteado al comienzo del trabajo.

Se quería analizar la relación entre el flujo migratorio de cada país y la cantidad de sedes en el exterior que tiene Argentina en dicho país. Tras elaborar las tablas, gráficos y análisis a partir del modelo de datos desarrollado, se notó que no hay una relación entre el flujo migratorio de cada país y la cantidad de sedes argentinas. Esto se evidencia en la Tabla 1 y Figura 3. A continuación se decidió analizar los datos agrupando los países según su región geográfica. Aquí se notó que las regiones con más sedes no son necesariamente aquellas en las que el flujo promedio es mayor que en las demás (Tabla 4). En la Figura 5 se observan las regiones con más sedes y en la Figura 6 el promedio de flujo migratorio por cada región y la variación de los valores del flujo en cada región. Comparándolas, se observa que la cantidad de sedes por región no depende del flujo migratorio. Se podría pensar que el número de sedes se ve influenciado por la cantidad de países de dicha región y la cantidad de inmigraciones y emigraciones en cada país, más que por el flujo migratorio. Un ejemplo que motiva esto es América del sur. Siendo estos los países más cercanos a Argentina, se puede ver como también son los países que más sedes argentinas tienen.

Notamos que, al analizar el flujo migratorio (inmigraciones - emigraciones), estamos ignorando el caudal de inmigraciones y emigraciones. Estos números podrían ser “grandes” pero parecidos, lo cual implica que el flujo migratorio tiende a 0. Esto podría ser otro causante de la cantidad de sedes que tienen países con flujo bajo.

7. Anexo

Archivos csv de las tablas elaboradas:

[Tabla1](#)

[Tabla4](#)

[Tabla5](#)

[Tabla6](#)