



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Practico N°1

Evaluar el manejo de datos y su visualización

12 de Mayo de 2024

Laboratorio De Datos

Grupo 4

Integrante	LU	Correo electrónico
Suarez Ines	890/22	ine.suarez22@gmail.com
Ramirez Ana	931/23	correodeanar@gmail.com
Wittmund Montero, Lourdes	1103/22	lourdesmonterochiara@gmail.com



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (+54 +11) 4576-3300

<http://www.exactas.uba.ar>

1. Resumen

Este trabajo trata sobre saber la existencia de cierta relación entre los flujos monetarios netos de Inversión Extranjera Directa (IED) de cada país y la cantidad de sedes en el exterior que tiene Argentina en dicho país. Para realizar esto se nos fueron proporcionados distintas fuentes de datos incluyendo información de flujos monetarios netos de Inversión Extranjera Directa (IED), Datos de las Representaciones Argentinas en el exterior y la tabla de países que vincula los datos de las fuentes IED y Representaciones Argentinas Datos anuales.

Nuestro objetivo es limpiar y organizar la información para facilitar la comprensión y manejo de los datos.

2. Introduccion

El presente trabajo tiene como objetivo principal evaluar el manejo de datos y su visualización por parte de cada uno de los alumnos en el contexto del análisis de las Representaciones Argentinas en el exterior de la República Argentina y las inversiones extranjeras. Esta tarea surge a partir del interés de los docentes de la materia Laboratorio de Datos por explorar una fuente de datos abiertos relacionada con los flujos monetarios netos de Inversión Extranjera Directa (IED) de diferentes países y su posible correlación con la cantidad de sedes en el exterior que tiene Argentina en dichos países.

Antes de proceder con el análisis y la interpretación de los datos, es fundamental comprender la naturaleza y la procedencia de la información con la que trabajaremos. Por ende, los docentes han identificado la necesidad de conocer cierta información adicional de las fuentes de datos utilizadas.

En este contexto, el presente trabajo se propone como una oportunidad para profundizar en el conocimiento sobre el manejo de fuentes de datos abiertos y su aplicación.

Para poder llevarlo a cabo en primer lugar leímos el enunciado completo y definimos qué actividades realizar y qué datos eran necesarios para resolverlas (consultas, visualizaciones, etc.). En base a eso armamos un diagrama conceptual de los datos (DER) que refleje la estructura necesaria para resolver el problema. Tuvimos que decidir de donde obtener los datos, diseñar los esquemas correspondientes al modelo relacional y corregirlos con el método GQM para poder cargar los datos limpios.

Finalmente mediante herramientas de visualización facilitamos la comprensión de los datos incluyendo gráficos que ayudan a identificar patrones.

3. Procesamiento de datos

Para este trabajo se nos proporcionaron 5 bases de datos:

- *flujos-monetarios-netos-inversion-extranjera-directa.csv* que cuenta con información de los flujos monetarios netos anuales de Inversión Extranjera Directa.
- *lista-sedes.csv* que cuenta con la información básica sobre las sedes argentinas en países extranjeros.
- *lista-sedes-datos.csv* que cuenta con la información completa sobre las mismas sedes.
- *lista-secciones.csv* que cuenta con la información completa sobre las secciones de las sedes.
- *países.csv* que cuenta con datos básicos sobre cada país.

Los primeros 4 se obtuvieron de la página oficial de datos del gobierno, en cambio, la última, de una fuente no oficial.

Empezando nuestro proceso de limpieza de datos podemos reconocer que *lista-sedes-datos.csv* y *lista-secciones.csv* no se encuentran en **PRIMERA FORMA NORMAL (1FN)** debido a que tienen más de un atributo que no son atómicos, en algunas celdas hay más de un valor. Por ejemplo en las columnas:

- Telefonos adicionales (en ambas bases de datos)
- Correos adicionales (en ambas bases de datos)
- Redes sociales (en *lista-sedes.csv*)

entre otras.

Por otro lado, *lista-sedes.csv* se encuentra en **SEGUNDA FORMA NORMAL (2FN)**. Esto pasa porque al tener todos atributos atómicos ya podemos confirmar que, a diferencia de las otras, se encuentra en **PRIMERA FORMA NORMAL (1FN)**, luego, definimos como **primary key (PK)** al atributo *sede_id* y vimos que había dos **dependencias funcionales (DF)**. La primera **DF** que encontramos era que todos los atributos dependían directamente de *sede_id*, nuestra **PK**, y la segunda era que los atributos *nombre-pais-castellano* y *nombre-pais-ingles* dependían de *iso3* (o *iso2*, la conclusión es la

misma). Como nuestra **PK** esta compuesta de un solo atributo, automaticamente nuestra base se encuentra en **SEGUNDA FORMA NORMAL (2FN)**, ya que la condicion para que no lo este es que ningun atributo puede depender parcialmente de la **PK**, cosa que es imposible por lo primero que mencione. Por ultimo nos fijamos si se encontraba ademas en **TERCERA FORMA NORMAL (3FN)** pero como la segunda **DF** mencionada anteriormente es **dependencia transitiva** de la **Pk**, siendo esta la condicion para no estar en **TERCERA FORMA NORMAL (3FN)**, concluimos que esta base esta en **SEGUNDA FORMA NORMAL (2FN)**.

Como nosotras queremos trabajar con bases de datos en **TERCERA FORMA NORMAL (3FN)** debemos buscar la(s) forma(s) de alcanzarla. Con este objetivo en mente, generamos un **DIAGRAMA ENTIDAD-RELACION (DER)** en el cual nos descartamos de la informacion que no era necesaria para la resolucion de nuestros problemas.

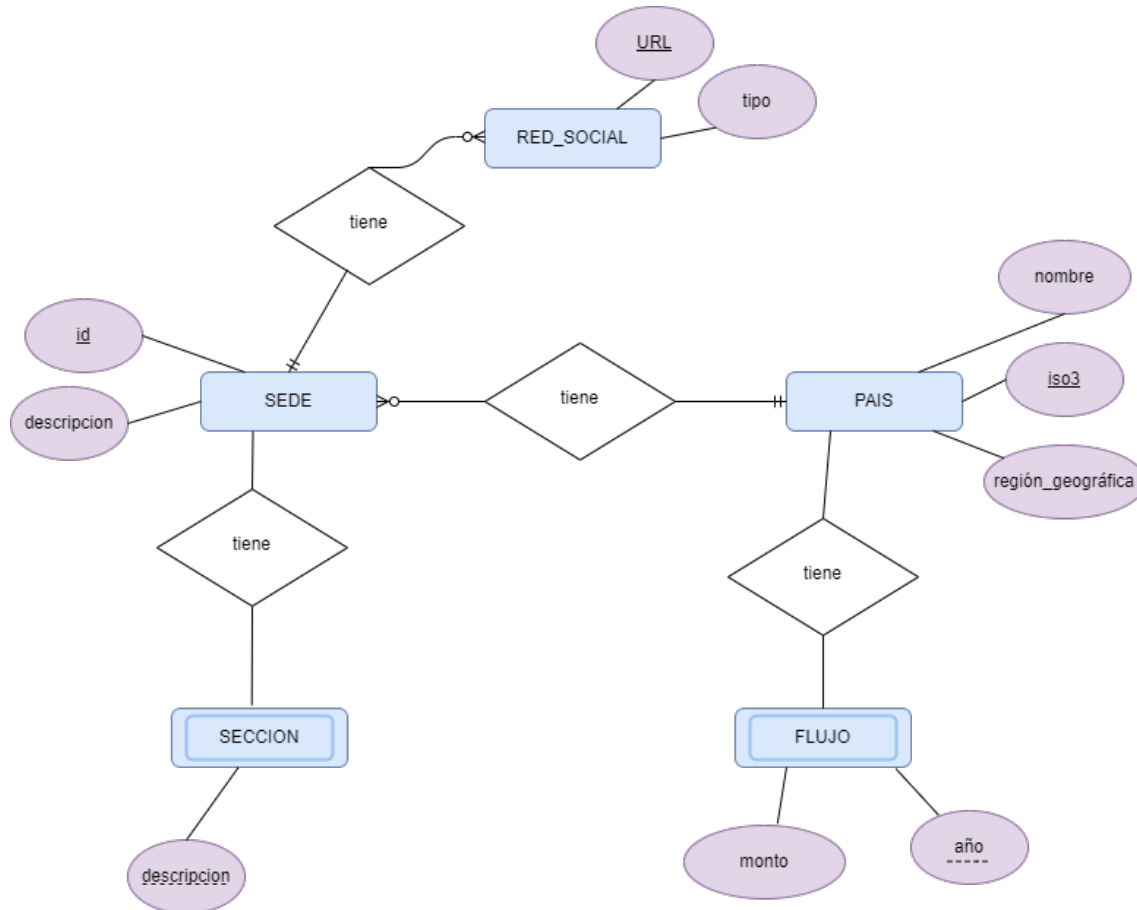


Figura 1: DER

En primer lugar identificamos las entidades fuertes **sede**, **pais** y **red social** y luego, en base a estas, pudimos definir ademas dos entidades debiles, **seccion** y **flujo**, las cuales no existen por si mismas y dependen de las entidades sede y pais respectivamente. Una vez que teniamos nuestras entidades, identificamos sus atributos.

- **sede**: *id* (el id de la sede), *descripcion* (la descripcion o nombre de la sede)
- **pais**: *iso3* (el codigo de identificacion del pais con tres letras), *nombre* (el nombre del pais), *region_geografica* (la region geografica en la cual se encuentra el pais)
- **red_social**: *url* (el link a una cuenta en la red social), *tipo* (el nombre de la red social)
- **seccion**: *descripcion* (la descripcion de la seccion de la sede)
- **flujo**: *monto* (el monto del flujo monetario neto anual en el pais), *año* (el año del monto)

Para cada una, identificamos los atributos clave.

- **sede**: *id*, el identificador unico de cada sede
- **pais**: *iso3*, el identificador unico de cada pais
- **red_social**: *url*, el link a una red social de una sede especifica por lo tanto, unica
- **seccion**: *descripcion*, la manera de identificar cada seccion (acompañada por el identificador de su sede por ser entidad debil)

- **flujo**: *año*, me va a definir el monto específico del año del que estoy hablando (acompañado por el identificador del país por ser entidad débil)

Luego identificamos las relaciones entre cada una de ellas.

Como ya mencionamos **seccion y flujo** son entidades débiles, por lo tanto siempre tienen una restricción de participación total respecto a su relación identificativa y están vinculadas de la siguiente forma:

- sede **tiene** seccion
- país **tiene** flujo

Las entidades fuertes por otro lado se relacionan así:

- sede **se ubica en** país. Es para ver en qué país se encuentra una sede, es una relación uno a muchos porque cada país puede tener muchas sedes o ninguna, mientras que cada sede tiene uno y solo un país en el cual está ubicada.
- sede **tiene** red_social. Es para ver las redes sociales que tiene una sede, es una relación uno a muchos porque cada sede puede tener muchas redes o ninguna, mientras que cada red social pertenece a uno y solo un país.

A partir del DER, hemos generado el **Modelo Relacional**, herramienta con la cual pudimos definir las **claves primarias (PK)**, **dependencias funcionales (DF)** y las **claves foráneas (Foreign keys)**.

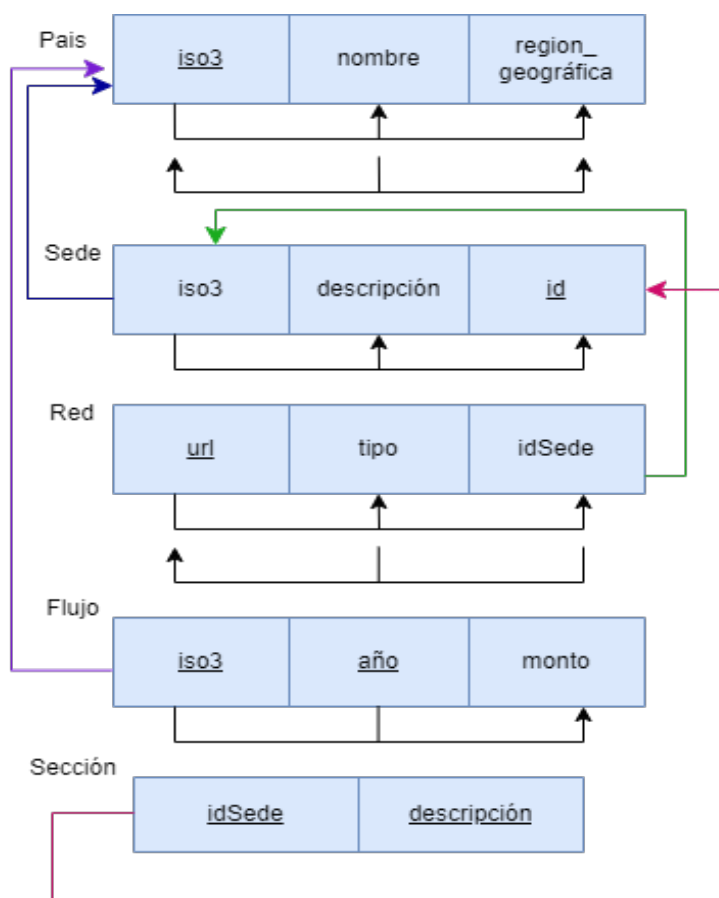


Figura 2: Modelo Relacional

Las tablas de relaciones que pudimos identificar fueron **Sedes(id, descripción, iso3)**, **Pais(iso3, nombre, region_geográfica)**, **Red(url, tipo, idSede)**, **Seccion(descripción, idSede)** y **Flujo(año, monto, iso3)**. Para cada uno definimos las siguientes claves primarias y claves foráneas:

- **Pais**: *iso3* es la clave primaria ya que lo era en el DER.
- **Sede**: *idSede* es la clave primaria ya que lo era en el DER, *iso3* es una clave foránea ya que es la clave primaria de **Pais** y como entre estas dos entidades hay una relación uno a muchos, de esta forma nos aseguramos que cada sede tenga su país y evitamos nulls.
- **Red**: *url* es la clave primaria ya que lo era en el DER, *idSede* es una clave foránea ya que es la clave primaria de **Pais** y como entre estas dos entidades hay una relación uno a muchos, de esta forma nos aseguramos que cada sede tenga su país y evitamos nulls.

- **Secciones:** *descripcion*, *idSede* es la clave primaria.*idSede* forma parte ya que al ser entidad debil en el DER, necesita de la clave primaria de su entidad fuerte asociada, en este caso **Sede**, para poder estar definida por completo. Ademas en este caso *idSede* es una clave foranea. Por otro lado *descripcion* forma parte por ser la clave primaria en el DER.
- **Flujo:** *año*, *iso3* es la clave primaria.*iso3* forma parte ya que al ser entidad debil en el DER, necesita de la clave primaria de su entidad fuerte asociada, en este caso **Pais**, para poder estar definida por completo. Ademas en este caso *iso3* es una clave foranea. Por otro lado *año* forma parte por ser la clave primaria en el DER.

Ahora logramos que todas nuestras tablas esten en **3FN**, a continuacion esta la justificacion:

- **1FN:** Nuestro modelo esta en 1FN si todos los atributos de la tabla son atómicos. En nuestro caso, teníamos el atributo *url* , *telefonos adicionales* , *correos adicionales* de las sedes como atributos no atómicos, pero luego de realizar la limpieza de **rede**, se puede decir que **sede**, **pais**, **flujo**, **seccion** y **red** no tienen atributos atómicos, por lo que están en Primera Forma Normal.
- **2FN:** Nuestro modelo esta en 2FN si esta en 1FN y además todo atributo (no clave) depende de manera completa a la clave primaria. En nuestro caso, podemos ver que todos los atributos de **pais** dependen completamente de *iso3*. Lo mismo ocurre en **redes** con el *url* , en **sede** con el *id* y en **flujo** con el *iso3* respectivamente. Entonces, se puede decir que están en Segunda Forma Normal.
- **3FN:** Nuestro modelo esta en 3FN si esta en 2FN y además ningún atributo depende transitivamente de la clave primaria, a menos que el atributo mensajero sea una clave candidata. En nuestro caso podemos ver que hay dependencias transitivas, por ejemplo en Pais, *nombre* depende de *iso3* mientras que *region_geografica* depende de *nombre*. Como *nombre* es una clave candidata en nuestra base de datos, podemos decir que *region_geografica* depende transitivamente de la clave primaria *iso3*. Vamos a utilizar la técnica **GQM** (Goal Question Metrics) , para evaluar la calidad de nuestras bases de datos. Con esta

metodología, primero definimos un objetivo claro. Luego, formulamos preguntas específicas cuyas respuestas nos indicarán si estamos alcanzando dicho objetivo. Finalmente, definimos métricas que nos ayudarán a responder esas preguntas, proporcionándonos datos concretos para evaluar el rendimiento de nuestras bases de datos.

- ***países.csv*** atributo de la calidad afectado : Relevancia- Problema correspondiente a Instancia
 - **1. Objetivo:** Todos los atributos sean relevantes y no haya informacion de más que no aporte informacion que no se pueda conseguir con los atributos ya disponibles.
 - **2. Pregunta:** ¿Cual es la proporcion de atributos que no son relevantes?
 - **3. Metrica:** Proporción de atributos irrelevantes (para hacer este calculo tomamos como atributos relevantes a los nombres de los países en castellano y el codigo iso3). Cantidad de atributos irrelevantes/cantidad total de atributos. Son 3 atributos irrelevantes (name, nom, iso2) de un total de 6, equivale a un 50 %.
- ***flujos-monetarios-netos-inversion-extranjera-directa.csv*** atributo de la calidad afectado : Consistencia-Problema correspondiente a Instancia
 - **1. Objetivo:** Todos los países tengan sus nombres sin _
 - **2. Pregunta:** ¿Cual es la proporcion de países con nombre con _?
 - **3. Metrica:** Proporción de países con nombre con _ . Cantidad de países con nombre con _/cantidad total de países. Son 43 países con _ en el nombre de un total de 172, equivale a un 25 %.
- ***lista-sedes-datos*** atributo de la calidad afectado : Completitud-Problema correspondiente a Instancia
 - **1. Objetivo:** Todos los atributos tengan sus datos completos.
 - **2. Pregunta:** ¿Cual es la proporcion de atributos incompletos?
 - **3. Metrica:** Proporción de atributos con algun registro vacio. Cantidad de atributos con algun registro vacio/cantidad total de atributos. Son 23 de un total de 37 atributos, equivale a un 62,16 %.

Para crear los DataFrames utilizados en el proyecto, se extrajeron todos los datos necesarios de las fuentes proporcionadas. A continuación se presenta una descripción detallada de las fuentes de datos utilizadas para cada DataFrame.

- **sedes:** Todos los datos se obtuvieron de *lista-sedes-datos.csv* luego de sacar los atributos que tenian algun dato vacio y quedarnos solo con los necesarios.
- **países:** Para obtener nombre y codigo usamos el archivo *países.csv* sacando los atributos irrelevantes, para region_geografica usamos *lista-sedes-datos.csv*.
- **redesSociales:** Los datos para este DataFrame se tomaron exclusivamente del archivo *lista-sedes-datos.csv*.
- **secciones:** Para armar este DataFrame todos los datos los extrajimos de *lista-secciones.csv*.
- **flujos:** Los datos de año y monto los obtuvimos de *flujos-monetarios-netos-inversion-extranjera-directa.csv*, luego le sacamos los _ a los países que contaban con ellos y esto nos ayudo a obtener codigo de una mezcla entre *países.csv* y *lista-sedes-datos.csv*.

4. Decisiones tomadas

Durante el avance del trabajo nos encontramos en distintas situaciones las cuales requerian una toma de decisiones:

1. Se toma *flujo* como entidad debil a pesar de que en un contexto general un flujo no depende de la existencia de un país, pero es necesario para nuestro trabajo.
2. Para utilizar la tecnica GQM en *países.csv* definimos como atributos relevantes al nombre del país en nuestro idioma y al código iso3 por ser el mas completo, inmediatamente esto nos definio como atributos irrelevantes a el nombre del país en ingles, frances, y al código iso2.
3. Sacamos los tildes, mayusculas y guiones bajos de los nombres de los países, sin embargo los que tienen un nombre distinto entre una tabla y otra se pierden.
4. Dejamos como tipo de redes sociales Facebook, Instagram, Twitter, Linkedin, Flickr, Youtube. En la base de datos proporcionada, ademas de las redes ya mencionadas, estaba como red un Gmail, decidimos sacarlo.
5. Sacamos los url que tenian solo un sin un link especifico.
6. Para obtener las regiones geograficas hubo que unir la tabla *países.csv* con *lista-sedes-datos.csv*, por lo tanto los países en los cuales no hay sedes argentinas se perdieron.
7. Hay sedes que no tienen secciones, y tambien se pierden al unir *lista-sedes-datos.csv* y *lista-secciones.csv*.

5. Análisis de datos

Para esta seccion generamos los siguientes reportes utilizando consultas de SQL.

1. **Para cada país informar cantidad de sedes, cantidad de secciones en promedio que poseen sus sedes y el flujo monetario neto de Inversión Extranjera Directa (IED) del país en el año 2022. El orden del reporte debe respetar la cantidad de sedes (de manera descendente). En caso de empate, ordenar alfabéticamente por nombre de país.**

Obtuvimos la siguiente tabla:

Índice	País	Sedes	seccionesPromedio	IED 2022(MU\$S)
0	BRASIL	11	1.63636	86050.4
1	URUGUAY	8	0.5	3838.75
2	BOLIVIA	7	2.28571	-26.3584
3	CHILE	6	2.33333	19786
4	CHINA	6	2.83333	189132
...
66	SUECIA	1	4	45963.1
67	TAILANDIA	1	6	10033.7
68	TUNEZ	1	7	713.365
69	UCRANIA	1	4	848
70	VENEZUELA	1	5	941.333

Tabla 1: Tabla resultante del ejercicio i

En este reporte se puede ver que los países que tienen más sedes tienen menos secciones en promedio. Por ejemplo, Uruguay que tiene 8 sedes, tiene en promedio 0,5 secciones por cada una. En cambio, Ucrania que tiene 1 sede tiene 4 secciones, por lo que el valor es mucho más alto. Esto se debe a que el promedio de secciones depende de la cantidad de sedes, por lo tanto, los países que tienen una sola sede, solo aportan al promedio sus secciones, así que el promedio sera igual a la cantidad de secciones en esa unica sede. Además vimos que el país con mayor cantidad de sedes es Brasil.

2. **Reportar agrupando por región geográfica: a) la cantidad de países en que Argentina tiene al menos una sede y b) el promedio del IED del año 2022 de esos países (promedio sobre países donde Argentina tiene sedes). Ordenar de manera descendente por este último campo**

Nos queda la siguiente tabla:

Indece	RegionGeografica	PaísesConSedesArgs	PromedioIEDMU\$ PaísesCon SedesArgs
0	AMÉRICA DEL NORTE	3	52633.2
1	OCEANÍA	2	34584.1
2	ASIA	23	29599.4
3	AMÉRICA DEL SUR	11	15564.2
4	EUROPA CENTRAL Y ORIENTAL	8	9687.43
5	EUROPA OCCIDENTAL	17	6305.73
6	ÁFRICA DEL NORTE Y CERCANO ORIENTE	5	3585.86
7	AMÉRICA CENTRAL Y CARIBE	14	1366.44
8	ÁFRICA SUBSAHARIANA	7	1340.52

Tabla 2: Tabla resultante del ejercicio ii

En este reporte se puede observar que no parece haber una relacion entre la cantidad de países con sedes argentinas por region geografica y el promedio de su IED en el año 2022, ya que la region con mas países con sedes es Asia y sin embargo no es la de mayor promedio. Lo mismo pasa con Oceania, es la region con menos sedes y sin embargo su promedio es el segundo mayor. Además, notamos que la hay una gran distancia entre los promedios de "las americas", por lo que la cercanía a Argentina tampoco parecería tener una gran influencia en su IED (al menos en el año 2022).

3. Para saber cuál es la vía de comunicación de las sedes en cada país, nos hacemos la siguiente pregunta: ¿Cuán variado es, en cada el país, el tipo de redes sociales que utilizan las sedes? Se espera como respuesta que para cada país se informe la cantidad de tipos de redes distintas utilizadas.

Esta es la tabla resultante:

Indice	País	cantDeRedes
0	BELGICA	6
1	ESTADOS UNIDOS DE AMERICA	6
2	CANADA	5
3	CIUDAD DEL VATICANO	4
4	ESPAÑA	4
...
66	POLONIA	1
67	REPUBLICA DOMINICANA	1
68	SIRIA	1
69	TAILANDIA	1
70	TRINIDAD Y TOBAGO	1

Tabla 3: Tabla resultante del ejercicio iii

En este reporte se puede observar que tan variado es el tipo de redes sociales que utilizan las sedes en cada país. Observamos que hay dos países que utilizan las 6 redes sociales que identificamos (BELGICA y ESTADOS UNIDOS DE AMERICA).

4. Confeccionar un reporte con la información de redes sociales, donde se indique para cada caso: el país, la sede, el tipo de red social y url utilizada. Ordenar de manera ascendente por nombre de país, sede, tipo de red y finalmente por url.

Índice	País	Sede	RedSocial	URL
0	ALEMANIA	CBONN	Facebook	.../ConsuladoArgentinoenBonn/
1	ALEMANIA	CBONN	Instagram	.../consulado.argentino.com/
2	ALEMANIA	EALM	Facebook	.../ArgEnAlemania/
3	ANGOLA	EANGO	Facebook	.../ArgentinaEnAngola/
4	ANGOLA	EANGO	Instagram	.../embargentinaenangola/
...
224	VENEZUELA	EVENE	Facebook	.../EmbajadaArgentinaEnVenezuela/
225	VENEZUELA	EVENE	Instagram	.../argenvenezuela/
226	VENEZUELA	EVENE	Twitter	.../argenvenezuela?lang=es/
227	VIETNAM	EVIET	Facebook	.../ArgentinaEnVietnam/
228	VIETNAM	EVIET	Instagram	.../argenvietnam/

Tabla 4: DataFrame del ejercicio iv

(a) Aclaración: Los puntos suspensivos de URL es la pagina de la Red social, por ejemplo si la red es Facebook el URL empezaria con <https://www.facebook.com>

En esta tabla pudimos ver todas las redes sociales utilizadas por cada país para mejorar el acceso a estas. Se encuentra el tipo de red, el país, y el link hacia la página de de cada una de las sedes.

Utilizando herramientas de visualización para poder analizar mejor los datos generamos los siguiente graficos en base a las consignas planteadas.

1. Cantidad de sedes por región geográfica. Mostrarlos ordenados de manera decreciente por dicha cantidad.

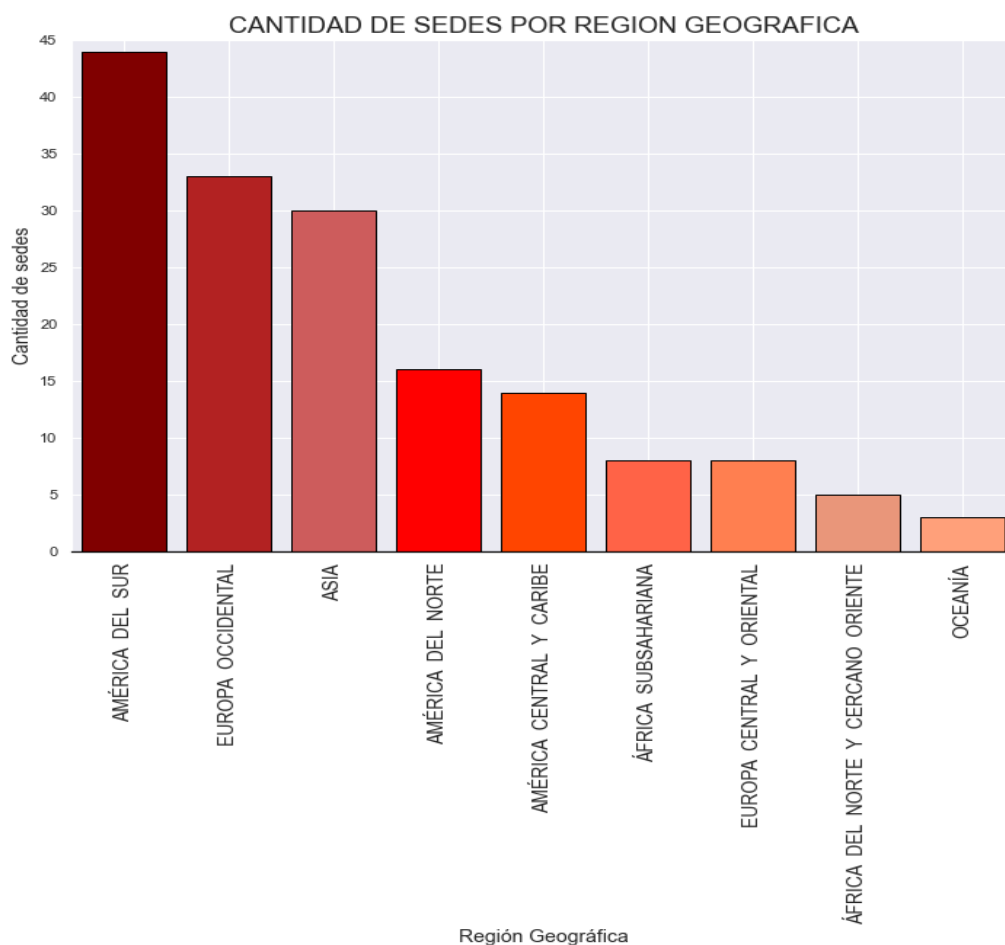


Figura 3: Cantidad de sedes por región

En la Figura 3 se presenta un grafico de barras donde podemos observar que la region geografica en la cual hay mayor cantidad de sedes es America del Sur. Esto se podria relacionar con que es la region en la cual se encuentra Argentina. Ademas, habiendo calculado en el ejercicio 2 la cantidad de paises con sedes por region geografica, notamos que aunque no sea la region con mayor cantidad de paises, cada uno de estos paises debe tener muchas sedes. Por otro lado, Oceania es la region con menor cantidad de sedes y tambien con menor cantidad de paises con sedes.

2. **Boxplot, por cada región geográfica, del promedio IED (2018-2022) de los países donde Argentina tiene una delegación. Mostrar todos los boxplots en una misma figura, ordenados por la mediana de cada región**

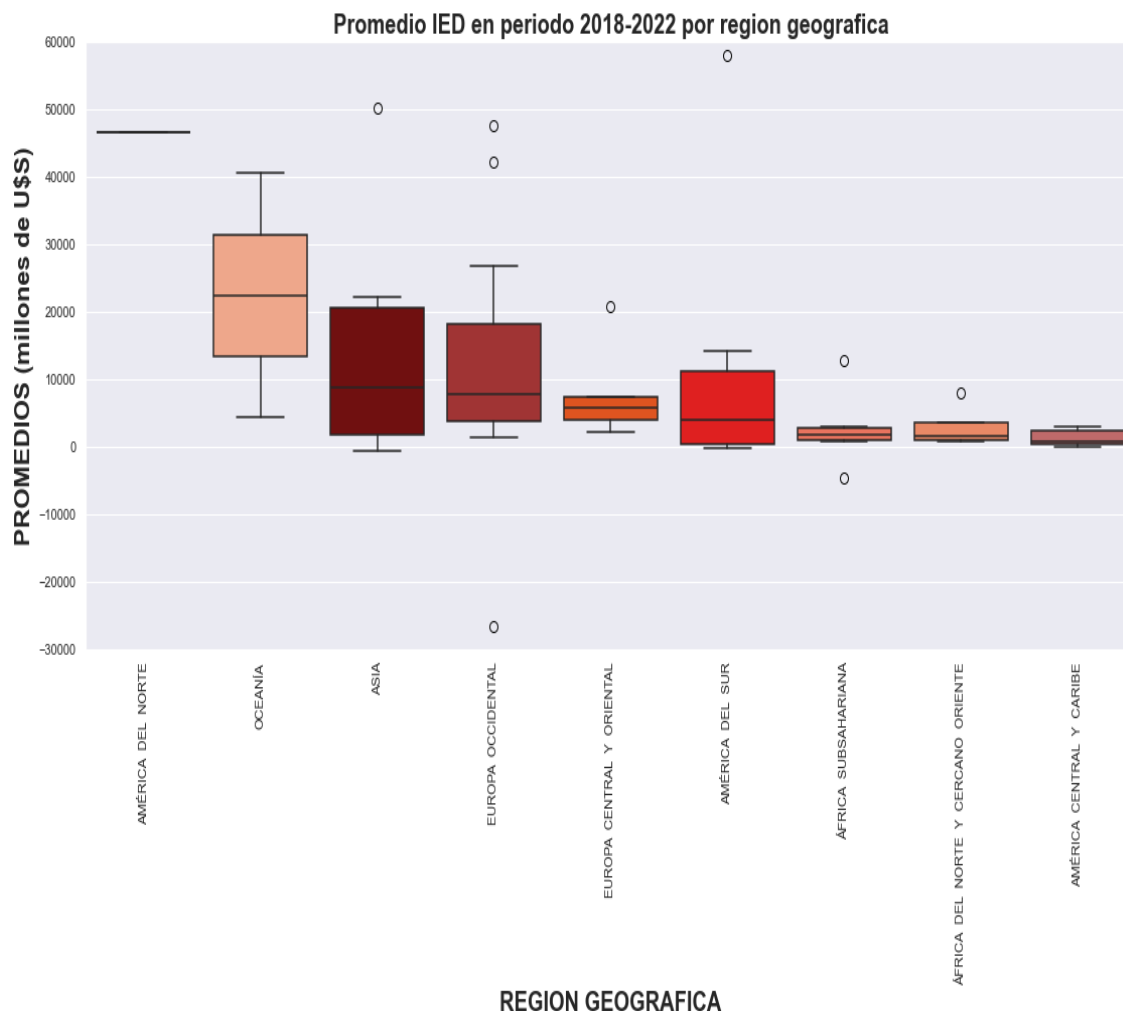


Figura 4: Promedio IED (2018-2022) por Region Geografica

En la Figura 4 se presenta un gráfico de BoxPlot donde se muestra el promedio IED en el periodo 2018-2022 y la variabilidad de este en diferentes regiones geográficas. Las regiones con el mayor valor promedio IED en el periodo 2018-1022 son América del Norte , Oceania y Asia. En America del Norte notamos que hay una sola linea porque solo nos quedo 1 pais en la tabla dado que los otros se perdieron por tener distinto nombre, por lo tanto el unico registro es el de ese pais y este mismo registro representa su mediana, razon directamente relacionada con que sea la mayor ya que no hay un rango amplio para calcularla. Por otro lado, las que presentan mayor variabilidad son Oceania, Asia y Europa Occidental.

3. Relación entre el de cada país (año 2022 y para todos los países que se tiene información) y la cantidad de sedes en el exterior que tiene Argentina en esos países.

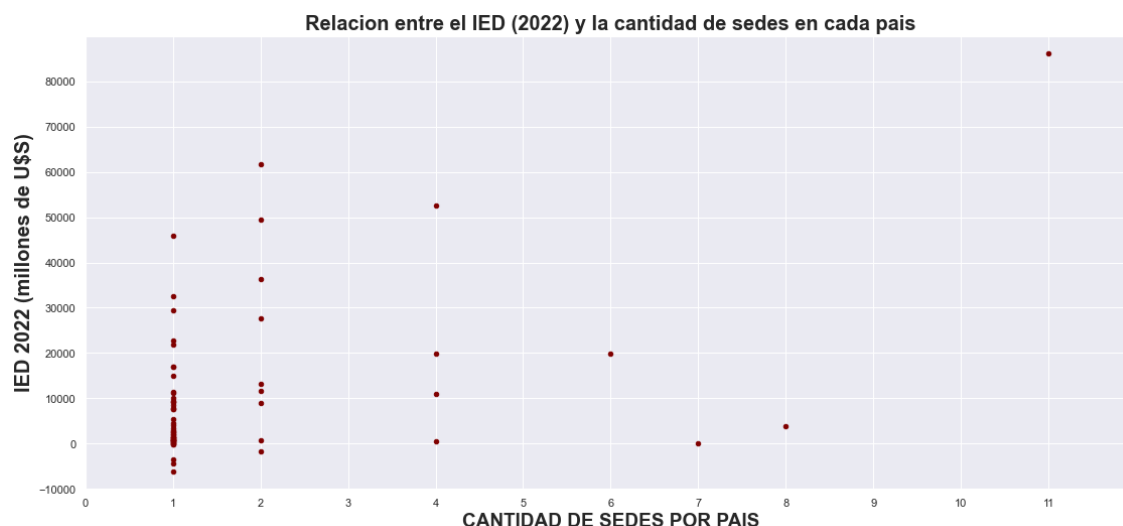


Figura 5: Relación entre IED 2022 y cantidad de sedes

En la Figura 5, se presenta un gráfico scatter plot que busca ilustrar la relación entre el IED (2022) y el número de sedes. Se destaca que la mayoría de los países cuentan con una única sede, mientras que Brasil sobrepasa notablemente las 10 sedes, destacándose como una excepción significativa en el conjunto de datos.

6. Conclusiones

En este proyecto, hemos llevado a cabo un análisis exhaustivo de los flujos monetarios netos de Inversión Extranjera Directa (IED) y la distribución de sedes argentinas en el extranjero. A través de la implementación de la técnica GQM (Goal Question Metrics), pudimos lograr una comprensión más profunda de la calidad y la relevancia de nuestras bases de datos. Además, examinamos la consistencia, completitud y relevancia de nuestros datos, aplicando técnicas de limpieza y normalización para garantizar su integridad y utilidad.

No identificamos relaciones significativas entre la cantidad de sedes en el exterior que tiene Argentina y los flujos de IED en diferentes países. En la primera tabla podemos observar esto ya que, por ejemplo, el tercer país con mayor cantidad de sedes, Bolivia, tiene un flujo monetario neto IED en el año 2022 negativo, aunque Brasil siendo el que tiene mayor cantidad, sí tiene un flujo considerablemente grande comparado con el resto. Lo que nos dice que no hay una proporción entre estas métricas. Tal vez para tener una mayor perspectiva sobre esto habría que agregar información sobre los años anteriores al 2022, pero analizando este caso específico esa es la conclusión a la que llegamos.

Igualmente, los hallazgos de nuestro análisis sugieren que esta relación puede variar significativamente según la región geográfica. Por ejemplo, en el gráfico 1, observamos una alta concentración de sedes en América del Sur, mientras que los flujos de IED muestran una mayor variabilidad en otras regiones como Oceanía y Europa Occidental, como se ve en el gráfico 2.

En resumen, este trabajo no solo ofrece información valiosa sobre la relación entre la expansión internacional y la inversión extranjera de Argentina, sino que también destaca la importancia de mantener bases de datos de alta calidad para realizar análisis precisos y fundamentados. Estos hallazgos pueden servir como punto de partida para futuras investigaciones y políticas relacionadas con la expansión global de Argentina y su impacto económico.