

# **Trabajo Práctico N°1**

Instituto Industrial Luis A. Huergo — 2025

Materia: Desarrollo de Sistemas 6°AO

Integrantes: Federico Greco y Rocío Penalva

Fecha de entrega: 10/09/2025

## Resumen

En este trabajo se busca analizar si existe relación entre el Producto Bruto Interno (PBI) per cápita del año 2023 y la cantidad de sedes argentinas en el exterior por país. Para ello, se combinaron dos fuentes de información: los datos del Banco Mundial sobre PBI per cápita y el conjunto de datos de las representaciones argentinas disponibles en el portal oficial de datos abiertos.

Se llevó a cabo un proceso de normalización, limpieza y homogeneización de datos, resolviendo problemas de calidad y de consistencia. A partir de ello, se elaboraron reportes y visualizaciones que permitieron responder las preguntas planteadas.

Las conclusiones indican que no se observa una correlación directa entre el nivel de riqueza de un país y la cantidad de sedes argentinas, sino que intervienen otros factores como la región geográfica, vínculos históricos, comercio bilateral y presencia de comunidades argentinas en el exterior..

## Introducción

### Problema a resolver

Partimos de dos bases de datos abiertas: una sobre las representaciones diplomáticas de Argentina en distintos países y otra sobre el PBI per cápita (año 2023) publicada por el Banco Mundial. Lo que buscamos es ver si existe alguna relación entre el nivel económico de cada país y la cantidad de sedes diplomáticas que Argentina mantiene allí.

### Objetivo general

El objetivo del trabajo es analizar si el PBI per cápita de los países tiene algún vínculo con la distribución de embajadas, consulados u oficinas argentinas en el exterior. Para eso procesamos la información, armamos reportes estadísticos y generamos gráficos que nos ayuden a entender mejor los patrones.

### Actividades realizadas

- **Limpieza de datos:** corregimos errores, unificamos nombres de países y preparamos los datasets para poder compararlos.
- **Organización:** armamos diagramas y cargamos la información en data frames para trabajar de forma más ordenada.
- **Reportes:** realizamos cuatro reportes con estadísticas clave sobre el tema.
- **Gráficos:** hicimos visualizaciones para explorar los datos y encontrar tendencias.
- **Correlación:** analizamos si había relación entre las variables económicas y la presencia diplomática.

### Resumen de la resolución

El proyecto quedó dividido en 4 carpetas:

“TablasOriginales”, donde están guardados los datasets tal cual vinieron descargados de sus fuentes.

“TablasLimpias” que tiene todos los datasets creados por nosotros, tanto los datos limpios como los datasets para resolver la consigna. También se incluyen uno de datos descartados del PBI, debido a su no utilidad (datos faltantes), y uno de problemas que no se detectaron en la primera limpieza pero sí a la hora de migrar dichos datos. Finalmente, incluye los datasets creados por los reportes que se piden en la consigna.

“Graficos” que guarda los gráficos generados para el punto 8 de la consigna.

“codigo”, que contiene todo el código utilizado en el proyecto, tanto para limpiar los datos como para generar los reportes y los gráficos.

# Procesamiento de datos

## Limpieza

Luego de ser descargados, los datos contaban con varios problemas e inconsistencias que afectaban la calidad de estos, por lo que se tomaron distintas medidas para aumentar su utilidad y coherencia a la hora de generar reportes. Cada uno de los dataframes tenía problemas distintos:

### **- Dataset PBI mundial:**

Nos encontramos con el problema de valores faltantes en la columna del valor del PBI del año 2023 para algunos países.

- **G (Goal):** Asegurar que el PBI per cápita 2023 esté disponible para la mayor cantidad de países posible.
- **Q (Question):** ¿Cuántos países no tienen un valor de PBI en 2023?

¿Es posible obtener un valor de años anteriores?

- **M (Metric):** Valores faltantes en la columna de 2023.

Valores que se pueden recuperar de años anteriores

- **Corrección:** Rellenar con el último valor válido (forward fill desde 1960–2022) y en caso de no existir, eliminar el país del dataset.

### **- Dataset Sedes:**

Los nombres de países no están normalizados (ej. "Estados Unidos" vs "United States").

- **G (Goal):** Garantizar que los nombres de países se puedan unir con los del Banco Mundial.
- **Q (Questions):** ¿Cuántos nombres de país en sedes no coinciden con los de la columna "Country Name" en el dataset del PBI?
- **M (Metric):** El número de nombres de países en sedes que no tienen correspondencia exacta con la columna "Country Name" del PBI.
- **Corrección:** Mapear los códigos en la columna "pais\_iso\_3" y compararlos con "Country Code" en el dataset del PBI.

### **- Dataset Secciones:**

Se encontraron secciones repetidas (con distintos empleados en el registro).

- **G (Goal):** Mantener consistencia en el número de secciones por sede.
- **Q (Question):** ¿Cuántas secciones tienen nombre duplicado dentro de la misma sede?

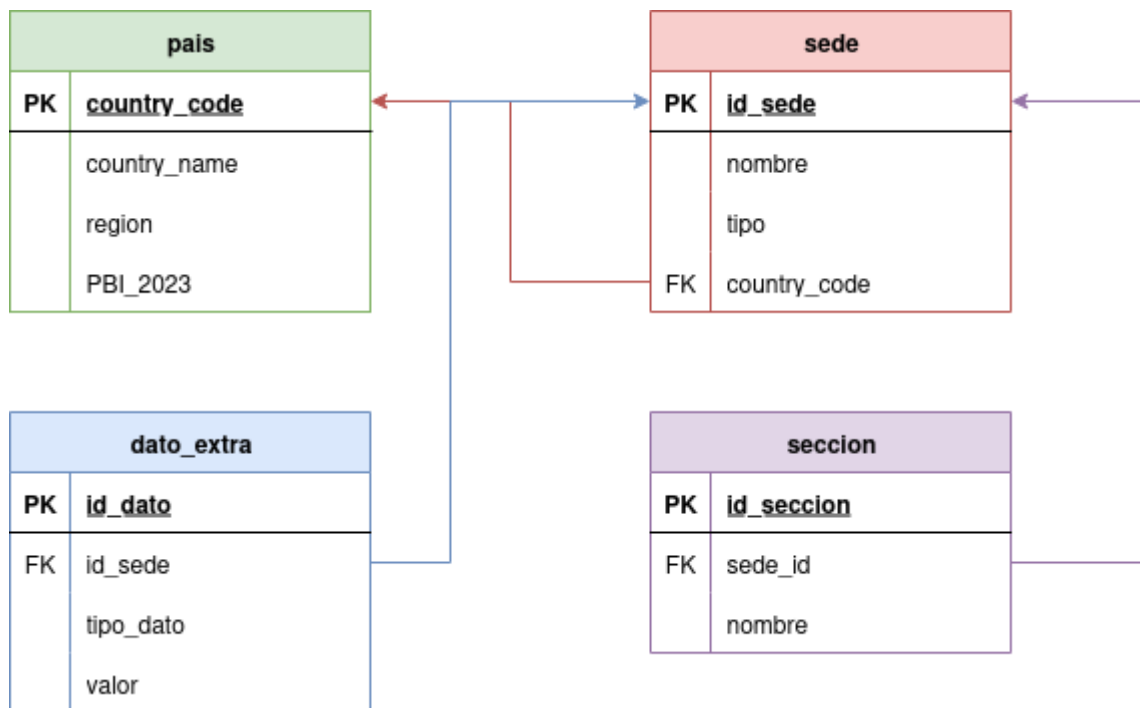
- **M (Metric):** Cantidad de secciones con “sede\_id” y “tipos\_seccion” repetidos
- **Corrección:** Eliminar duplicados.

#### - Dataset Sedes-Datos:

**Problema:** URLs de redes sociales inválidas o mal formadas.

- **G (Goal):** Asegurar que las URLs de redes sociales sean válidas para análisis.
- **Q (Question):** ¿Cuántos registros de red social tienen una URL inválida?
- **M (Metric):** Cantidad de registros que tienen una URL inválida.
- **Corrección:** Normalizar las URLs (agregar https:// si falta) o marcarlas como inválidas.

### Proceso de importación



- **Tabla “pais”:** Para completar los datos de esta tabla se importaron todos desde el dataframe del PBI mundial excepto la región. La región se importó desde el dataframe de “sedes-datos”, y para ello hizo falta ver a qué región pertenecía la sede de ese registro, y luego mediante el dataframe de sedes ver a qué país pertenecía dicha sede.
- **Tabla “sede”:** Los datos de esta tabla vienen todos del dataframe de sedes, y para el “country\_code” se utilizó la columna “pais\_iso\_3” para que estos códigos coincidieran con los de la tabla pais.

- **Tabla “seccion”:** Para esta tabla fue necesario crear la columna “id\_seccion” y completarla nosotros mismos, ya que en el dataset de secciones, que es de donde importamos el resto de datos, no existía tal campo.
- **Tabla “dato\_extra”:** Lo primero a mencionar es que el nombre de esta tabla fue una mala elección, ya que no representa correctamente sus datos, basada en el nombre del dataset del que importamos sus datos, “sedes-datos”. Esta tabla contiene los datos sobre las redes sociales, es decir que red social se usa en qué sede, y su respectiva URL, por lo que quizás un nombre como “redes\_sociales” habría sido más acertado.

## Decisiones tomadas

- Una decisión que tomamos fue incluir entre las tablas limpias un .csv de los datos que se eliminaron del dataframe del PBI por ser datos incompletos, por lo que se puede ver el registro de los 4 países cuyos PBI no estaba disponible
- La otra decisión fue crear un .csv de registros descartados durante la migración por errores que no se encontraron en la limpieza, en el que se aclara qué registros son y el problemas con estos.

## Análisis de datos

### Punto 7

Los .csv generados para esta consigna se encuentran en “TablasLimpias/reportes”, y no se incluyen debido a su gran extensión.

**A)**

	País	sedes	secciones promedio	PBI per cápita 2023 (US\$)
1	Brazil	11	0.72727275	10377.5892792557
2	United States	9	0.7777778	82304.6204272866
3	Uruguay	8	0.125	23019.4221560046
4	Bolivia	7	0.42857143	3686.27996442176
5	Chile	7	0.42857143	17067.0364799154
6	Spain	7	0.5714286	33509.0127982139

Los países que cuentan con más sedes argentinas son los países limítrofes o países con un alto PBI.

**B)**

	Región Geográfica	Países con Sedes Argentinas	Promedio PBI per cápita 2023 (US\$)
1	AMÉRICA DEL NORTE	3	333033.76327752206
2	EUROPA OCCIDENTAL	17	98285.0131489682
3	OCEANÍA	2	89163.3720891925
4	AMÉRICA DEL SUR	11	49150.80184743896

En este reporte están las regiones con menor cantidad de sedes Argentinas.

**C)**

	País	Cantidad Redes Sociales
1	Algeria	4
2	Angola	3
3	Argentina	1
4	Armenia	4
5	Australia	4

Este reporte no proporciona información de utilidad para la consigna, ya que el PBI de un país no guarda relación con las redes sociales que se utilizan en las sedes de dicho país.



D)

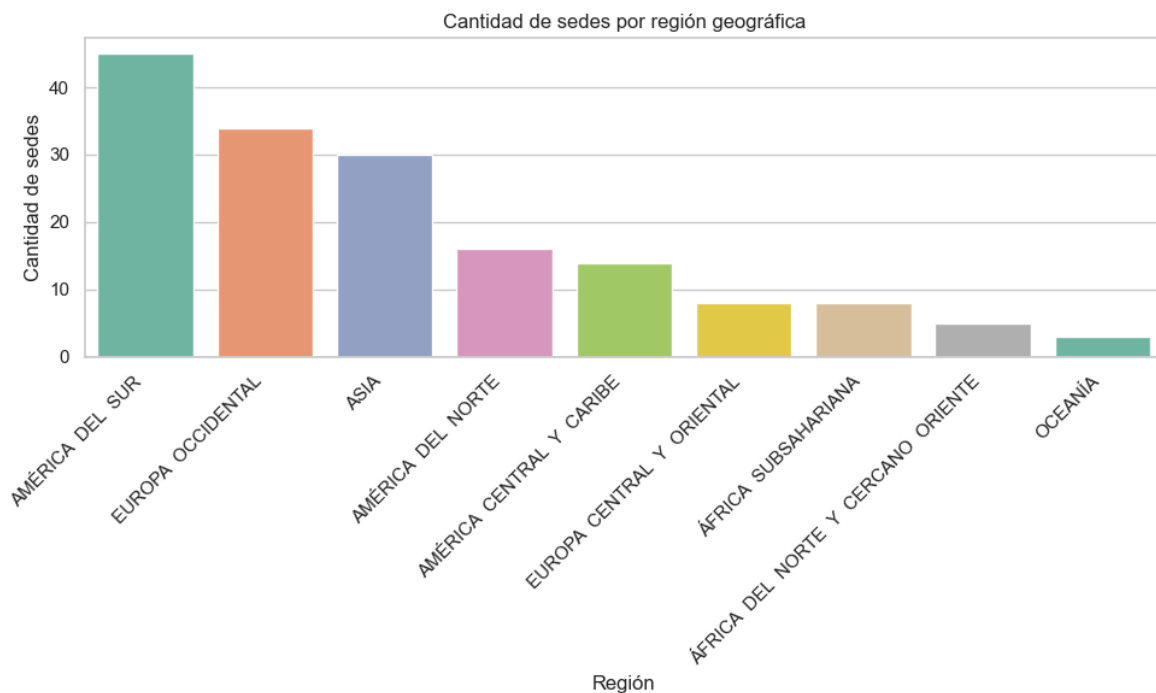
1	País	Sede	Red Social	URL
2	Algeria	EARGE	sitio_web	<a href="https://earge.cancilleria.gob.ar/">https://earge.cancilleria.gob.ar/</a>
3	Algeria	EARGE	facebook	<a href="https://facebook.com/ArgentinaEnArgelia">https://facebook.com/ArgentinaEnArgelia</a>
4	Algeria	EARGE	instagram	<a href="https://instagram.com/argenargelia">https://instagram.com/argenargelia</a>
5	Algeria	EARGE	twitter	<a href="https://twitter.com/ARGenArgelia">https://twitter.com/ARGenArgelia</a>

Este reporte no proporciona información de utilidad para la consigna, ya que el PBI de un país no guarda relación con las redes sociales que se utilizan en las sedes de dicho país.

## Punto 8

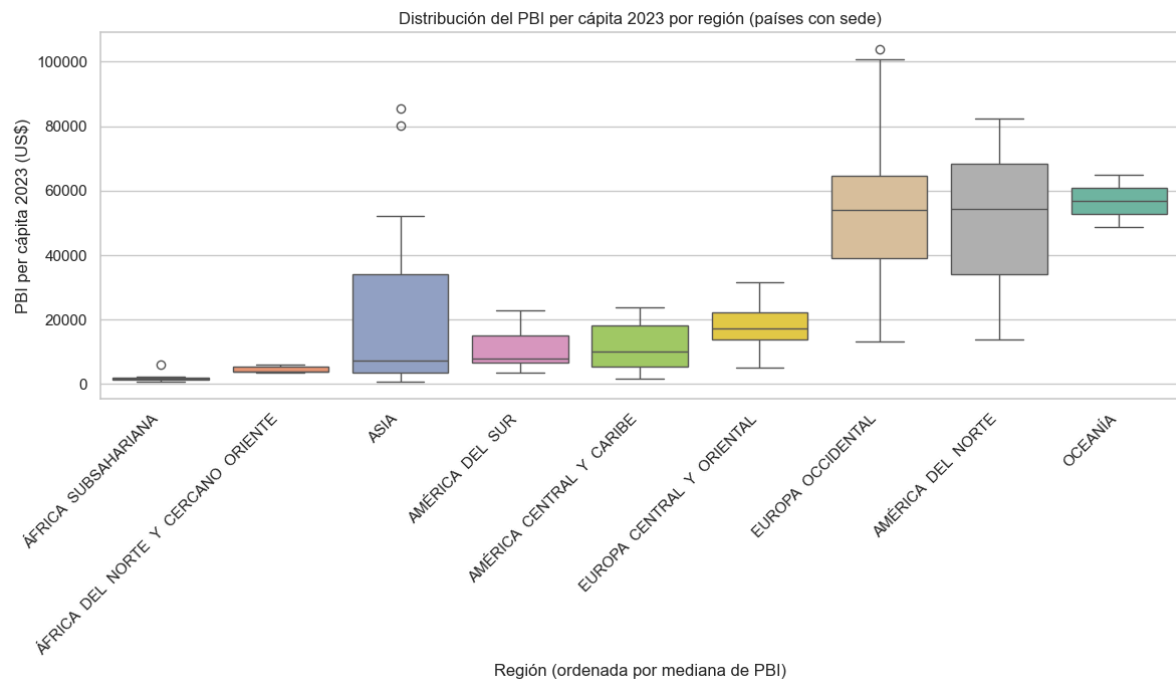
Los gráficos que se muestran también se encuentran adjuntos en el proyecto, en la carpeta “Graficos”

A)



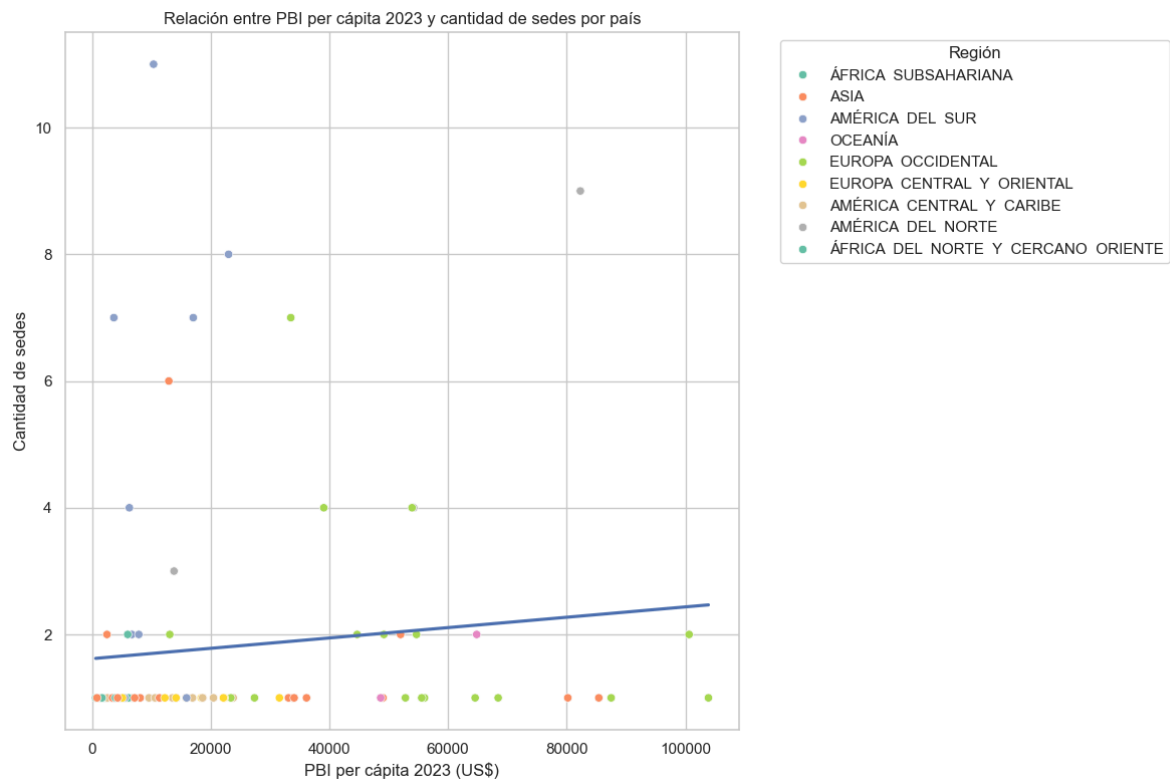
La distribución espacial sigue una lógica geográfica y de densidad de países: Sudamérica concentra muchas sedes (estrategia territorial y vecindad), Europa y Asia también tienen presencia notable (muchos países / vínculos históricos y económicos). Regiones con pocos países desarrollados o lejanía (Oceanía, África del Norte) muestran pocas sedes.

**B)**



Los datos muestran una clara separación: los países más ricos tienden a agruparse en Oceanía, Norteamérica y Europa Occidental, mientras que América Latina y numerosos países africanos se sitúan en la parte baja del espectro de ingreso. Las regiones con mayor cantidad de países presentan saltos más grandes entre sus valores extremos y sus medianas, lo que indica heterogeneidad interna. Asia, en particular, combina economías muy distintas dentro de la misma región. En síntesis, los boxplots ponen de relieve la fuerte desigualdad económica mundial; aún con presencia argentina en todos los continentes, las diferencias entre regiones son notorias.

c)



Existe una **ligera tendencia positiva** (a más PBI, *ligeramente* más sedes en promedio), pero visualmente la relación es **débil**: muchos países ricos tienen 1–2 sedes y algunos países con PBI bajo/medio tienen más sedes por motivos geopolíticos o de cercanía.

## Conclusiones

En conclusión, los resultados no apuntan a una relación importante o significativa entre el PBI de un país y las representaciones con que Argentina cuenta allí. La ubicación de las sedes responde en mayor medida a factores de proximidad geográfica, lazos históricos y afinidades políticas o culturales. La presencia diplomática se concentra especialmente en regiones vecinas y con relaciones estratégicas o sociales, mientras que sólo en casos puntuales, como los de China, Estados Unidos o Canadá, puede advertirse una coincidencia más marcada entre poder económico y número de representaciones.