A yellow and blue logo

AI-generated content may be incorrect.

**Universidad Católica Boliviana “San Pablo”**

**Sede Santa Cruz**

**Ingeniería en Inteligencia Artificial**

A blue circle with white and black logo

AI-generated content may be incorrect.

**Proyecto: Exploración de Indicadores Socioeconómicos Mundiales**

**Materia:** Adquisición, Análisis y Procesamiento

**Tema:** recolección, limpieza y análisis exploratorio

**Docente:** Ing. MARIN SALAZAR CARMEN ROSA

**Estudiante:** José Alfredo Zambrana Cruz

**Correo electrónico:** jose.zambrana.c@ucb.edu.bo

**Fecha:** 29 / 05 / 2024

***Santa Cruz - Bolivia***

# Introducción

En el mundo actual, donde los datos se generan en cantidades masivas, su correcto aprovechamiento depende en gran medida de la calidad con la que son recolectados y preparados. Este proyecto tiene como finalidad demostrar la importancia de una recolección y limpieza adecuada de los datos como paso fundamental previo a cualquier análisis significativo.

A menudo se piensa que el análisis de datos comienza con la interpretación de resultados; sin embargo, la realidad es que una base de datos mal recolectada, incompleta o sucia puede conducir a conclusiones erróneas o poco útiles. Por ello, este trabajo se centra en mostrar cómo un proceso meticuloso de recolección, limpieza y transformación permite obtener un conjunto de datos fiable, listo para ser analizado de forma precisa.

El proyecto recorre todas las etapas clave del flujo de trabajo en ciencia de datos: desde la obtención de la información, pasando por la depuración de errores, hasta la preparación final de los datos para su análisis. Al documentar este proceso, se busca evidenciar que la calidad del análisis depende directamente de la calidad del tratamiento previo de los datos.

Para ello, recolectaremos datos de todos los países del mundo de los siguientes indicadores: PBI, Tasa de Desempleo, esperanza de vida, Accesos a internet, densidad poblacional por km^2, índice de alfabetización %, inflación, corrupción, inversión en la salud, inversión en la educación, Acceso a agua potable, desnutrición infantil. Posteriormente realizaremos el proceso del limpieza y preparación para luego realizar el análisis exploratorio.

# 2. Recolección de Datos

Existen diversas formas de recolectar datos, y la elección del método adecuado depende del contexto y los objetivos del análisis. En este proyecto, se emplearán tres enfoques principales: la extracción de datos desde archivos planos, la técnica de *web scraping* para obtener información directamente de páginas web, y el uso de APIs que permiten acceder a datos estructurados proporcionados por plataformas externas. Cabe mencionar que estaremos trabajando solamente en Python.

## 2.1 obtención de datos a través de Archivos planos

Fuente: <https://www.worldbank.org/ext/en/home>

Para este apartado decidí descargar la BD de World Bank Group. esta base de datos proporciona aproximadamente 1,600 indicadores para 217 economías, abarcando temas como pobreza, salud, demografía, PIB, comercio y medio ambiente. Los datos están disponibles para su descarga en formatos como CSV y Excel, y se pueden acceder a través de su herramienta DataBank.

Para guardar todos estos datos en Python se me fue necesario tener instalado la librería pandas ya que esta nos da muchas herramientas para extraer datos desde archivos planos y convertirlos en un Data Frame. Mediante la siguiente línea de código podemos extraer todos los datos desde nuestro archivo xlsx.

**pd.read\_excel (“Nombre del archivo”, shet\_name= “Nombre de la hoja en excel”)**

Mediante este código pude realizar la extracción de datos de archivos planos.

En esta ocasión se extrajeron los siguientes datos para todos los países del mundo: “Deuda Externa”:"DT.DOD.DSTC.CD"   
“inversión Extranjera”:"BX.KLT.DINV.CD.WD"  
Y posteriormente se los guardo en un data frame llamado df\_economic\_indicator

También se extrajeron datos como:  
“inversión en salud”:"SH.XPD.CHEX.GD.ZS"  
 “inversión en educación”:"SE.XPD.TOTL.GD.ZS"  
 “Acceso a agua potable”:"SH.H2O.SMDW.ZS”  
“desnutrición infantil”:"SH.STA.MALN.ZS"

Y posteriormente se los guardo en un data frame llamado df\_bienestar\_desarrollo

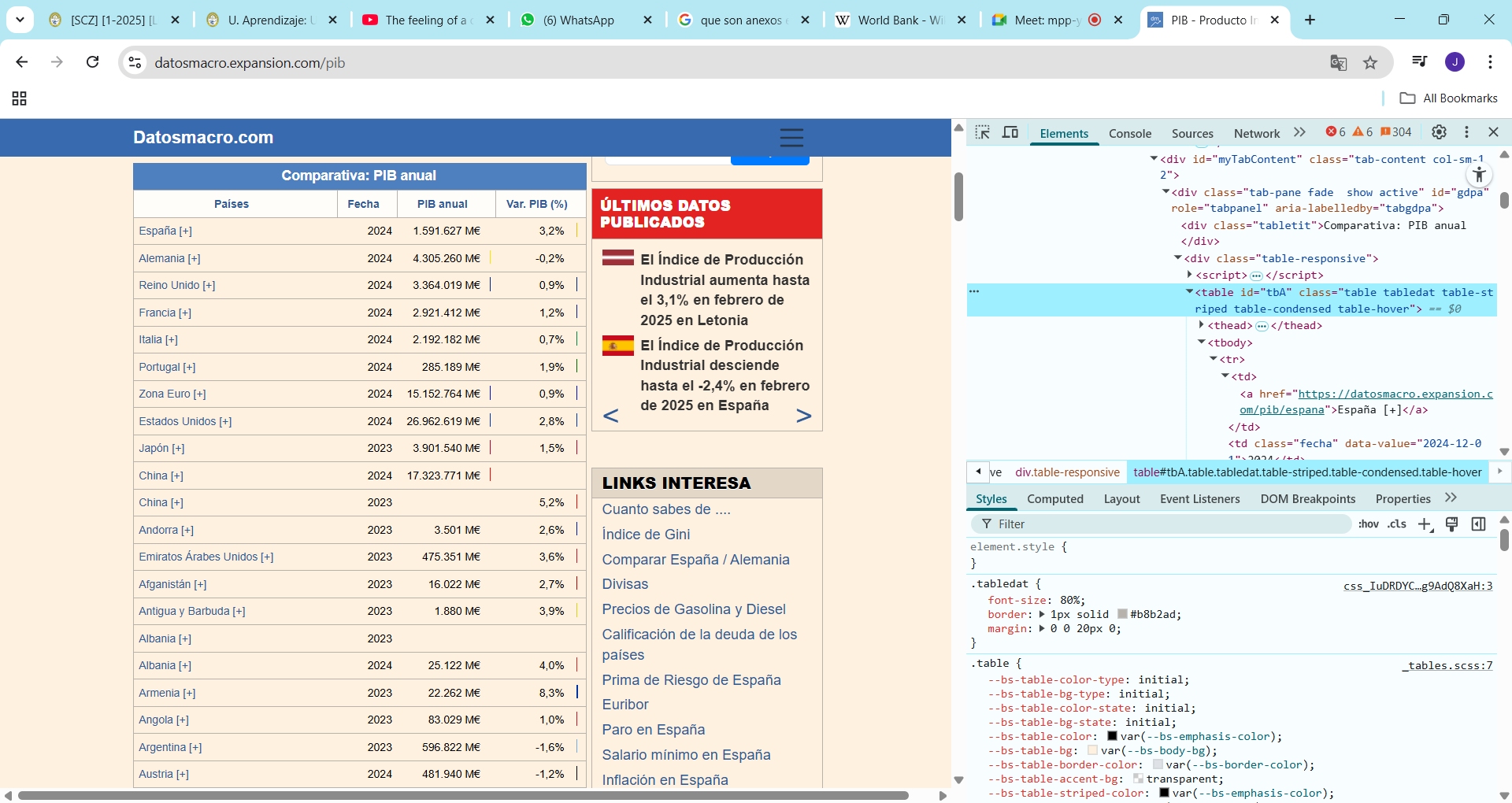
Los data frames (df\_economic\_indicator, df\_bienestar\_desarrollo )con los datos recolectados del archivo plano, pueden verse a más detalle en **anexos**.

## 2.2 obtención de datos haciendo Web Scraping

Extraemos la información de la página <https://tradingeconomics.com/country-list/corruption-index> que nos dará el **ranking de países con mayor a menos control de corrupción en el mundo.**

El cuadro que nos interesa está en el apartado div class =’card en el html, así que aplicamos web Scraping a solo esa tabla y extraemos todos sus valores. Adicionalmente, guardaremos toda la tabla en un data frame llamado df\_corruption\_indicators\_2024, el cual posteriormente limpiaremos si es necesario y lo prepararemos para el análisis.

También obtuvimos información del PBI de los países del mundo en el año 2024 de la página <https://datosmacro.expansion.com/pib> y aplicamos web Scraping al cuadro que se nos muestra a continuación:



Finalmente guardamos los datos de la tabla en un data frame llamado df\_pbi\_2024 el cual posteriormente lo limpiaremos y prepararemos para realizar el análisis exploratorio. Los data frames df\_pbi\_2024, df\_corruption\_indicators\_2024 se podrán ver mas a detalle en el apartado de **anexos**

## 2.3 obtención de datos mediante API

Extraemos nuevamente información de World Bank Group esta vez con una pequeña diferencia, haremos uso de APIs. La pagina cuenta con una API a la cual puedes acceder a todos los datos dependiendo el dato que quieras saber, por ejemplo:

<https://api.worldbank.org/v2/country/all/indicator/>

es la URL que usamos para conectarnos con World Bank, adicionalmente debemos colocar el código del indicador del cual queremos que nos traiga los datos, ejemplo:

Digamos quiero traer el 'PIB' su código seria 'NY.GDP.MKTP.CD', entonces adjuntaría el código a la URL y quedaría tal que así: <https://api.worldbank.org/v2/country/all/indicator/NY.GDP.MKTP.CD> y así la Api devuelve los datos en cuanto al PBI de todos los países del mundo.

De esta manera se realizó el uso de APIs para la recolección de datos, en mi caso traje los siguientes indicadores de todos los países del mundo usando la Api:

    'PIB': 'NY.GDP.MKTP.CD',

    'Tasa de Desempleo': 'SL.UEM.TOTL.ZS',

    'Esperanza de Vida': 'SP.DYN.LE00.IN',

    'Acceso a Internet': 'IT.NET.USER.ZS',

    'Densidad Poblacional': 'EN.POP.DNST',

    'Alfabetización': 'SE.ADT.LITR.ZS',

    'Inflación': 'FP.CPI.TOTL.ZG',

    'Índice de corrupción':'CC.PER.RNK'

Todos estos datos finalmente fuero guardados en un data frame llamado, df\_general\_information en los **anexos** podrán verlo más a detalle.

# 3. Limpieza y Transformación

Los datos recolectados anteriormente haciendo uso de archivos planos y APIs contenían datos que abarcaban desde 1960 a 2023. Al realizar un análisis de faltantes gracias a la función df.isnull().sum() que no da la cantidad de valores nulos por A screenshot of a computer

AI-generated content may be incorrect.cada columna o también podemos verlo a través de la tabla:

A screenshot of a computer

AI-generated content may be incorrect.

En este caso es evidente la cantidad de valores faltantes des de los años 1960 a 1999 por lo que decidimos reducir nuestra fuente de datos desde el año 2000 a 2023, de esta forma eliminamos muchos nulos y la calidad de datos es mucho mejor. Adicional mente las columnas que representaban las fechas 1960, 1961, 1962, 1963……..2023 se las convierte a filas usando el comando pd.melt de la librería pandas. El resultado quedo tal que así:

A screenshot of a computer program

AI-generated content may be incorrect.

Finalmente, realizamos pívot a los valores de indicator name y los convertimos a columnas, esto para facilitar la extracción de datos y merges al momento de querer realizar un análisis exploratorio. El resultado es el siguiente:  
A screenshot of a computer

AI-generated content may be incorrect.

De esta forma quedo nuestro Data Frame.

Finalmente se realiza el conteo de nulos gracias a la función df.isnull.sum () de la librería de pandas y vemos que:

External debt stocks, short-term (DOD, current US$) 5817  
Foreign direct investment, net inflows (BoP, current US$) 3106

Hay bastantes países con muchos valores nulos. No obstante, estos no serán problema al momento de sacar un promedio, cuartiles o valor de mediana ya que el programa los ignora. Mas adelante al momento de realizar correlaciones si se hará sumo control en cuanto a nulos eliminando cada fila que los contenga con la función drop.na () esto con la finalidad de no afectar al valor de correlación.

De igual manera que se realizo limpieza y preparación de datos recolectados mediante Apis y archivos planos, se realizo la limpieza de los datos extraídos con Web Scraping

A screenshot of a computer program

AI-generated content may be incorrect.

La df\_pbi\_2024 al momento de realizar Web Scraping la columna referencia, que guardaba el año, tenía sus valores tipo objeto, **lo cual puede causar problemas** al momento de realizar un análisis exploratorio ya que **todos los datos de año de nuestras data frames están en formato date-time** por lo que se realizó el casting de tipo objeto a date time y solo nos quedamos con el año. También cambiamos el nombre de la columna “Referencia” por “date” ya que todas la data frames que se recolectaron a lo largo de este proyecto guardan los años en una columna llamada “date” por lo que valga la redundancia cambiaremos el nombre de la columna “Referencia” por “date”.

A screenshot of a computer program

AI-generated content may be incorrect.

Esa es nuestra df\_corruption\_indicators\_2024 ya limpiada y preparada.

Por último tenemos el siguiente data frame con los datos del pbi del 2024 extraídos mediante web scraping

A screenshot of a computer program

AI-generated content may be incorrect.

Podemos darnos cuenta que los valores de las columnas PBI\_EUR, PBI\_USD son de tipo objeto ya que contienen caracteres especiales como M$ o M€ por lo que lo procedemos a limpiar las columnas quitándole los caracteres especiales M$ y M€ y realizar un casting de tipo de dato objecto a float64 a las columnas PBI\_EUR y PBI\_USD para que se sus valores sean flotantes y se pueda realizar un análisis estadísticos con ellos. Finalmente, este es nuestro data frame limpio y listo para realizar un análisis exploratorio

A screenshot of a computer program

AI-generated content may be incorrect.

Estos son todos los data frames con nuestros datos recolectados ya listos para realizar un análisis exploratorio.

Nota: El código, con el cual se recolectaron, limpiaron y prepararon los datos, se puede encontrar en el archivo. ipynb en el proyecto o en el apartado de anexos.

# 4. EDA

El análisis exploratorio es de suma importancia ya que es la parte donde nuestros datos empiezan a hablar. Previamente se había recolectado, limpiado y preparado los datos en distintos data frames para su análisis, por lo cual es bueno saber cómo se comportan esto datos, su media, mediana, rango, desvío estándar, sesgo y curtosis. Esto nos dará un panorama mucho más amplio de nuestros datos.

A screenshot of a computer

AI-generated content may be incorrect.

**Datos estadísticos de nuestros datos en el data frame df\_general\_information**

A screenshot of a computer

AI-generated content may be incorrect.

**Datos estadísticos de nuestros datos en el data frame df\_bienestar\_desarrollo**

* **Analizamos el promedio de PBI a nivel mundial a lo largo de los años:**

A screen shot of a computer

AI-generated content may be incorrect.

*Esta serie de tiempo nos muestra que el PBI de todos los países en conjunto va aumentando y en consecuencia también el índice de desarrollo humano. Esto nos muestra que el mundo esta en constante crecimiento y evolución económicamente.*

* **A screen shot of a computer

  AI-generated content may be incorrect.El índice de desnutrición en niños en todos los países del mundo a lo largo de los años**

*Estos datos demuestran que el índice de niños desnutridos, por debajo de su peso normal, disminuye en el mundo al pasar los años.*

* **Analizaremos si un país tiene una baja economía porque hay alta corrupción, para ello analizaremos la correlación entre el PBI y la corrupción en el mundo**

A red and blue squares

AI-generated content may be incorrect.

Como podemos observar en el grafico de calor, la relación entre el PBI de los países del mundo y la corrupción en los países del mundo es muy baja, por no decir mediocre. Esto cobra sentido cuando nos damos cuenta que el PBI no representa al 100% la economía de un país. Para ello estudiaremos otra variable llamada desarrollo económico, y la correlacionaremos con la corrupción en los países del mundo.

A screenshot of a computer

AI-generated content may be incorrect.

Sorprendentemente al comparar la corrupción de un país con un dato que representa más la economía del mismo, la correlación se hace más fuerte con gran notoriedad, a lo que podemos concluir que, a más corrupto el país, menos desarrollo económico.

* **Alfabetización a nivel mundial:**

A graph of a graph

AI-generated content may be incorrect.

*A nivel mundial los datos nos dicen que en promedio la alfabetización en los países es del 85% en su población.*

* **PBI de los países del mundo 2024**

*A white screen with black text

AI-generated content may be incorrect.*

*En conclusión, los países del mundo en su mayoría no pasan de 1 billón en su PBI, además hay muchos valores atípicos, es decir países cuyos PBI se disparan. Esto nos indica la gran brecha en cuanto a economía respecta existe entre países.*

*PD: El país con 3 billones en PBI pertenece a Estados Unidos.*

* **Desnutrición infantil en el mundo%**

*A diagram of a graph

AI-generated content may be incorrect.*

*Podemos concluir que el índice de desnutrición infantil en los paises del mundo es en promedio del 10%.*

* **Cantidad de personas por km^2:**

*A white background with black dots and circles

AI-generated content may be incorrect.*

*La mayoría de los paises tienen menos de 2500 personas por km^2. No obstante existen paises que incluso llegan a los 17500 habitante por km^2*

*indicator Country Name Population density (people per sq. km of land area)*

*3968 Monaco 18680.902111*

*Mónaco incluso cuenta con 18680.902111 personas por km^2*

* **Sabiendo la cantidad de personas por km^2, cual es la situación de china siendo el país con mayor sobrepoblación.**

**A graph with a line

AI-generated content may be incorrect.**

*Como observamos en la serie de tiempo, China ha ido aumentando el número de personas por km^2 a pesar de tener una gran población. Por último el anho 2023 alcanzo como máximo 151 personas por km^2*

* **Control de corrupción a Nivel mundial**

A diagram of a graph

AI-generated content may be incorrect.

*En promedio el control de corrupción en los paises es del 50%, una cifra considerablemente baja.*

*En este caso los paises más seguros son:*

*indicator Country Name Control of Corruption: Percentile Rank*

*1411 Denmark 100.000000*

*1963 Finland 99.523811*

*5092 Singapore 99.047623*

*4229 New Zealand 98.571426*

*5619 Sweden 98.095238*

*Dinamarca con un 100% de control contra corrupción posicionándose como el país menos corrupto del mundo*

* **Tasa de desempleo a nivel mundial**

A diagram of a graph

AI-generated content may be incorrect.

*La tasa de desempleo en los paises del mundo es de un 6% en promedio, no obstantes observamos valores atípicos (paises que tienen alto índice de desempleo). 2019*

* **Gastos del gobierno en educación al nivel mundial**

***A diagram of a graph

AI-generated content may be incorrect.***

*Los paises del mundo gastan en promedio casi 4% de su PBI en educación*

* **¿Sabiendo lo anterior, cuanto china gasta de su PBI en educación?**

A diagram of a graph

AI-generated content may be incorrect.

*China en promedio gasta 4% de su PBI en educación y lo máximo que ha llegado a gastar es de 4.25 % de su PBI. Esto cobra sentido cuando sabemos que el PBI de China es de más de 17 billones*

* Centrándonos más en Bolivia cual es el PBI de este

*A graph with a line going up

AI-generated content may be incorrect.*

*El PBI de Bolivia va en crecimiento considerable, no obstante, en el anho 2020 cae inusualmente. Si analizamos bien la fecha y el contexto social vivido nos damos cuenta de que en esas épocas el covid-19 azotó el mundo, por lo cual, el PBI en Bolivia se vio afectado y con varias razones, todo esto explica la caída en el PBI*

* Tasa de desempleo en Bolivia

A graph with a line going up

AI-generated content may be incorrect.

**📉 Descenso entre 2008 y 2012**

* Durante este período, Bolivia vivió una **época de estabilidad económica** y un **buen precio de las materias primas**, especialmente el gas natural y los minerales (zinc, plata, estaño).
* El gobierno usó **políticas de inversión pública** intensas en infraestructura y programas sociales.
* Esto ayudó a reducir el desempleo, ya que **aumentó la demanda de mano de obra** en construcción, servicios y minería.

**📈 Subida en 2013**

* Esta subida podría reflejar una **ralentización en el crecimiento**, o el inicio de una reducción en los precios internacionales de las materias primas.
* También puede estar ligada a una **entrada de nuevos trabajadores al mercado laboral** (por ejemplo, jóvenes o migrantes internos) sin que haya un crecimiento equivalente en los empleos formales.

**📉 Caída entre 2013 y 2015**

* En este período, el desempleo baja nuevamente, posiblemente por los **programas de empleo del gobierno** y una economía aún activa, aunque con menos impulso que años anteriores.
* En 2014 hubo aún precios aceptables del gas, y se mantenía una relativa estabilidad.

**📈 Aumento constante hasta 2019**

* Aquí Bolivia empieza a sentir más fuerte la **caída en los precios del gas y minerales**, que son su principal fuente de ingresos.
* Además, la economía comienza a **desacelerarse** y **la informalidad crece**.
* También hubo una **transición política tensa en 2019**, con conflictos y cambios de gobierno que generaron **incertidumbre económica** y frenaron inversiones.

**📈 Pico en 2020**

* Este es **el efecto de la pandemia de COVID-19**.
* Bolivia, como el resto del mundo, **cerró actividades económicas**, y muchos sectores (turismo, comercio, informalidad) se paralizaron.
* La tasa de desempleo se disparó, como puede verse claramente en el gráfico.
* **Control de corrupción en Bolivia**

**A graph with blue lines and dots

AI-generated content may be incorrect.**

**📈 2000: Punto alto inicial (~40)**

* En esta etapa Bolivia tenía un índice relativamente alto, pero no necesariamente por ausencia de corrupción, sino porque los mecanismos de control aún estaban bajo fuertes influencias internacionales (ajuste estructural y reformas del Banco Mundial).
* Este dato puede estar más vinculado a la percepción de transparencia institucional desde afuera.

**📉 Caída 2001-2004 (~22 a 24)**

* En este periodo hubo una fuerte crisis política y social en Bolivia.
* Gobiernos inestables, protestas, la "guerra del gas", renuncias presidenciales (como la de Gonzalo Sánchez de Lozada en 2003).
* Todo eso aumentó la percepción de corrupción e ineficiencia estatal.

**📈 Aumento notable en 2006-2008 (~39-40)**

* Coincide con el inicio del gobierno de Evo Morales (2006), donde se promovieron cambios institucionales con fuerte respaldo social.
* Inicialmente hubo percepción de mejora en la gestión pública, nacionalización de sectores estratégicos y promesas de mayor control estatal, lo cual mejoró el índice.

**📉 Descenso progresivo de 2011 a 2015 (~40 a ~30)**

* Aunque hubo crecimiento económico, se empezaron a ver señales de concentración de poder, falta de transparencia en contrataciones públicas y debilidad institucional.
* Casos de corrupción comenzaron a surgir más frecuentemente en medios.

**📉 Caída continua 2016-2022 (~27 a ~21)**

* Durante esta etapa, el índice sigue bajando constantemente. Esto refleja:
  + Acusaciones de corrupción dentro del Estado.
  + Falta de independencia del poder judicial.
  + Crisis institucional en 2019, donde hubo un cambio de gobierno tras protestas por posible fraude electoral.
  + Transición política con acusaciones cruzadas de corrupción tanto al gobierno saliente como al interino.
  + En pandemia (2020), se denunciaron actos de corrupción en compras públicas, como los respiradores con sobreprecio.
* **A screenshot of a computer

  AI-generated content may be incorrect.Por último una mega correlación entre todos los datos recolectados**

Rescatemos correlaciones fuertes:

* Deuda externa con PBI 🡪 Alta correlación positiva
* Inversiones en el extranjero con inversión en la salud 🡪Correlación positiva aceptable
* Inversiones en el extranjero con PBI 🡪 Alta correlación positiva
* Personas tomando agua con Esperanza de vida 🡪 Alta correlación positiva
* inversión en la salud con Deuda externa 🡪 Alta correlación negativa
* PBI con inversiones en la salud 🡪 correlación positiva aceptable
* Esperanza de vida en personas con personas con acceso a internet% población 🡪 Alta correlación positiva

El **uso de internet** suele estar relacionado con:

* Mayor acceso a información de salud
* Educación más amplia
* Comunicación más eficiente
* Servicios médicos digitales (telemedicina, citas virtuales)
* Conectividad en zonas rurales o aisladas

Todo esto contribuye, de manera indirecta, a **mejorar la calidad de vida y aumentar la esperanza de vida**.

# 5. Conclusiones

Si bien a lo largo del proyecto no se planteó una hipótesis específica, fue posible identificar correlaciones relevantes entre los datos recolectados. Por ejemplo, se observó que los países con mayor Producto Bruto Interno (PBI) tienden a tener niveles más altos de deuda externa, lo cual es coherente con su mayor capacidad de endeudamiento. Asimismo, se evidenció cómo la percepción de corrupción puede influir negativamente en el desarrollo económico de una nación.

A lo largo del análisis exploratorio, también se trabajó con series de tiempo, lo que permitió visualizar la evolución de distintas variables a lo largo de los años y extraer patrones temporales significativos. Por otra parte, la utilización de gráficos de caja permitió entender la distribución de los datos y detectar valores atípicos.

Para alcanzar estos resultados, fue fundamental aplicar técnicas rigurosas de recolección, limpieza y transformación de datos. Dado que las fuentes eran diversas, los datos presentaban numerosos problemas: presencia de caracteres no deseados, tipos de datos inconsistentes, y estructuras heterogéneas. Fue necesario normalizar columnas, establecer claves (keys) para vincular registros, y organizar los datos en estructuras unificadas mediante dataframes compatibles, facilitando así operaciones como merge o join.

Finalmente, se realizó una evaluación cuidadosa de las variables disponibles, con el objetivo de seleccionar aquellas con mayor potencial explicativo o informativo. Dada la gran cantidad de datos, este proceso fue esencial para identificar relaciones significativas y formular observaciones relevantes sobre los distintos países analizados.

# 6. Recomendaciones

**Recomendaciones**

1. **Asegurarse de que no haya demasiados valores nulos:** Es fundamental verificar la cantidad de valores nulos en el conjunto de datos, ya que su presencia puede afectar el análisis y los resultados, especialmente en el cálculo de correlaciones. Si se encuentran demasiados nulos, es recomendable reducir el rango temporal o geográfico de los datos. Por ejemplo, en mi caso, la base de datos contenía información desde 1960, pero debido a la alta cantidad de valores nulos en esos años, decidí limitar el análisis a los periodos de 2000 a 2023, evitando así que los valores nulos distorsionaran los resultados.
2. **Asegurarse de que los datos tengan una estructura uniforme:** Aunque no es estrictamente necesario, tener una estructura de datos coherente y homogénea es altamente beneficioso, especialmente cuando se requiere realizar operaciones de combinación como joins o merge usando una clave común. Una estructura unificada facilita la integración de datos provenientes de diversas fuentes y mejora la interpretación de los resultados.
3. **Validar la confiabilidad de las fuentes de datos:** Es crucial asegurarse de que las fuentes de datos sean confiables y estén actualizadas. La calidad de los datos depende en gran medida de la credibilidad de las fuentes. Se recomienda utilizar fuentes oficiales, como organismos gubernamentales, instituciones académicas o bases de datos reconocidas, y siempre verificar la metodología utilizada en su recolección. Además, es importante realizar un análisis preliminar para identificar posibles inconsistencias o sesgos que puedan afectar los resultados del análisis.
4. **Verificar la coherencia y precisión de los datos:** Antes de realizar cualquier análisis, es importante validar que los datos sean consistentes y precisos. Esto incluye la identificación y corrección de errores tipográficos, la verificación de rangos de valores y la aseguración de que las unidades de medida sean consistentes. La precisión y coherencia de los datos son fundamentales para garantizar la fiabilidad de los análisis y las conclusiones derivadas de ellos.
5. **Documentar todos los pasos del proceso de limpieza y preparación de datos:** Es recomendable mantener un registro detallado de todas las decisiones tomadas durante el proceso de recolección, limpieza y transformación de los datos. Esto incluye la justificación de la eliminación o modificación de registros, la normalización de las variables y la creación de nuevas variables derivadas. Una documentación adecuada asegura la transparencia y reproducibilidad del análisis.

# 7.Anexos

El código fuentes lo podrán ver junto a la documentación en el repositorio GitHub