Esperanza de Vida

Proyecto Grupal

Informe de actividades del equipo
Data Engineering

Octubre 2023

Metodología de trabajo

La metodología de trabajo utilizada en el proyecto "Esperanza de Vida" se basa en el marco ágil Scrum el cual permite trabajar en una serie de interacciones en equipo. Las fases que definen y en las que se divide un proceso de SCRUM son las siguientes:

- **El quién y el qué:** identifica los roles de cada uno de los miembros del equipo y define su responsabilidad en el proyecto.
- **El dónde y el cuándo:** que representan el Sprint.
- **El por qué y el cómo:** representan las herramientas que utilizan los miembros de Scrum.

Roles del equipo (quién y el qué):

Product Owner/Dueño del producto: Es "voz del cliente" el responsable de definir los objetivos del proyecto, gestionar el backlog de tareas y tomar decisiones sobre las funcionalidades a desarrollar.

Scrum Master: Es el facilitador del equipo, se encarga de asegurar que el equipo siga las prácticas y principios de Scrum, y ayuda a resolver impedimentos y dificultades del proyecto.

Development Team Members/Miembros del Equipo de desarrollo: Está compuesto por especialistas en diferentes áreas relevantes para el proyecto, como data scientist, data Analyst, data Engineer, etc.

Backlog del producto:

El Product Owner es el encargado de mantener el backlog del producto, que es una lista priorizada de todas las funcionalidades, requisitos y tareas pendientes para el proyecto.

El backlog se va actualizando y refinando a medida que se obtiene más información y se van identificando nuevas necesidades durante el desarrollo del proyecto.

Sprints:

Los sprints son iteraciones cortas de tiempo, generalmente de 2 a 4 semanas, en las cuales se desarrollan funcionalidades y se entregan resultados incrementales. Al comienzo de cada sprint, el equipo se reúne en una reunión de planificación de sprint para seleccionar las tareas a realizar del backlog y establecer los objetivos específicos para el sprint.

Durante el sprint, el equipo trabaja en las tareas asignadas y se reúne en reuniones diarias de seguimiento, llamadas "daily scrums", para compartir el progreso, identificar posibles obstáculos y ajustar el plan si es necesario.

Reuniones de revisión y retrospectiva:

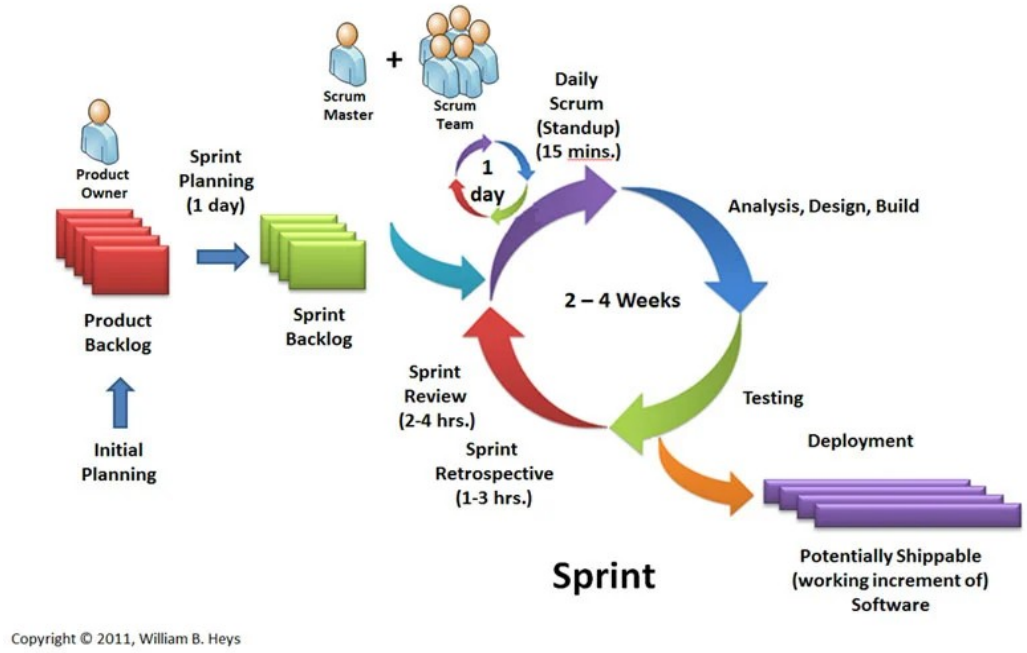
Al final de cada sprint, se lleva a cabo una reunión de revisión del sprint, en la cual el equipo presenta los resultados alcanzados durante el sprint y recibe comentarios del Product Owner y de otras partes interesadas.

Después de la reunión de revisión, se realiza una reunión de retrospectiva del sprint, en la cual el equipo reflexiona sobre el proceso de trabajo, identifica oportunidades de mejora y establece acciones para implementar en los siguientes sprints.

Entrega incremental y continua:

A medida que se completan los sprints, se van entregando incrementos funcionales del proyecto, lo que permite obtener retroalimentación temprana y hacer ajustes si es necesario.

La entrega incremental también permite que las partes interesadas vean el progreso y realicen ajustes o cambios de dirección si es necesario.



Transparencia y comunicación:

La transparencia es un principio fundamental en Scrum. Se promueve una comunicación abierta y clara entre los miembros del equipo y las partes interesadas, con el objetivo de mantener a todos informados sobre el progreso, los obstáculos y las decisiones tomadas.

El canal de comunicación para los integrantes del proyecto fueron seleccionados whatsapp y google meet canales por los cuales el equipo de desarrolladores se mantienen en continuo contacto de manera colaborativa, organizada y eficiente hacia el logro de los objetivos del proyecto.

Equipo de trabajo - Roles y responsabilidades

En el proyecto "Esperanza de Vida", se requiere la colaboración de un equipo multidisciplinario para llevar a cabo tareas clave relacionadas con la manipulación y análisis de datos. A continuación, se detallan los roles principales en el equipo y sus responsabilidades:

Data Engineer (Ingeniero de Datos: Carlos Villarreal, Leonardo Prada):

El Data Engineer es responsable de la infraestructura y el flujo de datos en el proyecto.

- Diseñar y desarrollar la arquitectura de datos responsable de diseñar y construir la infraestructura de datos necesaria para el proyecto, incluyendo la configuración de sistemas de almacenamiento y bases de datos.
- Extracción, transformación y carga de datos (ETL) se encarga de extraer los datos relevantes de diversas fuentes, transformarlos en un formato adecuado para su análisis y cargarlos en la infraestructura de datos.
- Mantenimiento y optimización de la infraestructura de datos, asegurándose de que funcione de manera eficiente y esté disponible para su uso en todo momento.
- Colaboración con otros roles trabajando en estrecha colaboración con los Data Analysts y Data Scientists para garantizar que los datos estén disponibles y sean accesibles para su análisis.

Data Analyst (Analista de Datos: Ivonn Gonzalez):

El Data Analyst se enfoca en el análisis y la interpretación de los datos para obtener información valiosa. Sus responsabilidades incluyen:

- Análisis de datos utilizando herramientas y técnicas estadísticas para analizar los datos y descubrir patrones, tendencias y relaciones relevantes.
- Preparación de informes y visualización de datos claras y comprensibles para comunicar los resultados del análisis a los demás miembros del equipo y clientes.
- Colaboración con otros roles de manera estrecha con el Data Engineer y el Data Scientist para obtener los datos necesarios y apoyar en la interpretación de los resultados.

Data Scientist (Científico de Datos: Fernanda Mosquera, Daniel Vielma)

El Data Scientist es responsable de aplicar técnicas avanzadas de análisis de datos y modelado para obtener información y generar conocimiento. Sus responsabilidades incluyen:

- Modelado y análisis predictivo utilizando técnicas de modelado avanzadas para predecir la esperanza de vida y explorar cómo diferentes factores socioeconómicos influyen en ella.
- Desarrollo de algoritmos y modelos desarrollar algoritmos y modelos para analizar los datos y descubrir relaciones no evidentes.
- Interpretación de resultados analiza y traduce los resultados del análisis en insights accionables que puedan utilizarse para recomendar al cliente las mejores alternativas (países) para la inversión mas adecuada a sus intereses
- Colaboración con otros roles trabaja en estrecha colaboración con el Data Engineer y el Data Analyst para obtener los datos necesarios, validar los modelos y comunicar los resultados del análisis.

La colaboración entre el Data Engineer, el Data Analyst y el Data Scientist es fundamental para el éxito del proyecto. El Data Engineer se encarga de proporcionar una infraestructura de datos robusta y confiable, lo que permite a los otros miembros del equipo acceder y

procesar los datos necesarios para su análisis. El Data Analyst, por su parte, utiliza técnicas estadísticas y herramientas de visualización para comprender y comunicar los hallazgos obtenidos a partir de los datos. Finalmente, el Data Scientist aplica métodos avanzados de modelado y análisis para obtener insights predictivos y proponer recomendaciones basadas en los resultados.

Análisis preliminar de calidad de datos Recopilación de datos:

Factores socioeconómicos importantes a estudiar inicialmente

1. Esperanza de vida al nacer (Total) | worldbank.org link: SP.DYN.LE00.IN
2. Esperanza de vida al nacer (female) | worldbank.org link: SP.DYN.LE00.FE.IN
3. Esperanza de vida al nacer (male) | worldbank.org link: SP.DYN.LE00.MA.IN

El indicador sobre "**Esperanza de vida al nacer"** se refiere a la cantidad de años que viviría un recién nacido *si los patrones de mortalidad vigentes al momento de su nacimiento no cambian a lo largo de la vida del infante*.

Se considerará como el Target para el modelado, así como, es factor importante para el análisis

4. Urban population (% of total population) | worldbank.org link: SP.URB.TOTL.IN.ZS

La **"Población Urbana"** se refiere a las personas que viven en áreas urbanas según lo definen las oficinas nacionales de estadística. Los datos son recopilados y suavizados por la División de Población de las Naciones Unidas.

5. Rural population (% of total population) | worldbank.org link: SP.RUR.TOTL.ZS

La **"Población Rural"** se refiere a las personas que viven en zonas rurales según la definición de las oficinas nacionales de estadística. Se calcula como la diferencia entre la población total y la población urbana.

6. Population growth (annual %) | worldbank.org link: SP.POP.GROW

La **"Tasa de Crecimiento Poblacional"** anual para el año t es la tasa exponencial de crecimiento de la población a mitad de año desde el año t-1 hasta t, expresada como porcentaje. La población se basa en la definición de facto de población, que cuenta a todos los residentes independientemente de su estatus legal o ciudadanía.

7. Inflation, consumer prices (annual %) | worldbank.org link: FP.CPI.TOTL.ZG

La **Inflación,** medida por el índice de precios al consumidor, refleja el cambio porcentual anual en el costo para el consumidor promedio de adquirir una canasta de bienes y servicios que puede fijarse o cambiarse en intervalos específicos, como por ejemplo anualmente. Generalmente se utiliza la fórmula de Laspeyres.

8. Gini index | worldbank.org link: worldbank.org link: SI.POV.GINI

El índice de **Gini** mide el grado en que la distribución del ingreso (o, en algunos casos, el gasto de consumo) entre individuos u hogares dentro de una economía se desvía de una distribución perfectamente equitativa. Una curva de Lorenz traza los porcentajes acumulados del ingreso total recibido frente al número acumulado de beneficiarios, comenzando con el individuo o el hogar más pobre. El índice de Gini mide el área entre la curva de Lorenz y una línea hipotética de igualdad absoluta, expresada como porcentaje del área máxima bajo la línea. Así, un índice de Gini de 0 representa una igualdad perfecta, mientras que un índice de 100 implica una desigualdad perfecta.

9. Inflation, GDP deflator (annual %)' | worldbank.org link: NY.GDP.DEFL.KD.ZG

La inflación medida por la tasa de crecimiento anual del deflactor implícito del PIB muestra la tasa de cambio de precios en la economía en su conjunto. El deflactor implícito del PIB es la relación entre el PIB en moneda local corriente y el PIB en moneda local constante.

10. GDP (current US$) | worldbank.org link: NY.GDP.MKTP.CD

El **PIB** a precios de comprador es la suma del valor agregado bruto de todos los productores residentes en la economía más los impuestos sobre los productos y menos los subsidios no incluidos en el valor de los productos. Se calcula sin hacer deducciones por depreciación de activos fabricados o por agotamiento y degradación de recursos naturales. Los datos están en dólares estadounidenses actuales. Las cifras en dólares del PIB se convierten a partir de monedas nacionales utilizando tipos de cambio oficiales de un solo año. Para algunos países donde el tipo de cambio oficial no refleja el tipo efectivamente aplicado a las transacciones reales de divisas, se utiliza un factor de conversión alternativo.

11. GDP per capita (current US$) | worldbank.org link: NY.GDP.PCAP.CD

El **PIB per cápita** es el producto interno bruto dividido por la población a mitad de año. El PIB es la suma del valor agregado bruto de todos los productores residentes en la economía más los impuestos sobre los productos y menos los subsidios no incluidos en el valor de los productos. Se calcula sin hacer deducciones por depreciación de activos fabricados o por agotamiento y degradación de recursos naturales. Los datos están en dólares estadounidenses actuales.

12. GDP per capita growth (annual %) | worldbank.org link: NY.GDP.PCAP.KD.ZG

Tasa de crecimiento porcentual anual del PIB per cápita basada en moneda local constante. El PIB per cápita es el producto interno bruto dividido por la población a mitad de año. El PIB a precios de comprador es la suma del valor agregado bruto de todos los productores residentes en la economía más los impuestos sobre los productos y menos los subsidios no incluidos en el valor de los productos. Se calcula sin hacer deducciones por depreciación de activos fabricados o por agotamiento y degradación de recursos naturales.

13. GNI (current US$) | worldbank.org link: NY.GNP.MKTP.CD

El INB (anteriormente PNB) es la suma del valor agregado de todos los productores residentes más cualquier impuesto sobre los productos (menos subsidios) no incluido en la valoración de la producción más los ingresos netos de ingresos primarios (compensación de los empleados e ingresos de la propiedad) del exterior. Los datos están en dólares estadounidenses actuales.

14. Current health expenditure (% of GDP) | worldbank.org link: SH.XPD.CHEX.GD.ZS

Nivel de gasto corriente en salud expresado como porcentaje del PIB. Las estimaciones de los gastos corrientes en salud incluyen los bienes y servicios de atención médica consumidos durante cada año. Este indicador no incluye gastos de capital en salud, como edificios, maquinaria, TI y reservas de vacunas para emergencias o brotes.

15. Hospital beds (per 1,000 people)'| worldbank.org link: SH.MED.BEDS.ZS

Las camas hospitalarias incluyen camas para pacientes hospitalizados disponibles en hospitales y centros de rehabilitación públicos, privados, generales y especializados. En la mayoría de los casos se incluyen camas para cuidados agudos y crónicos.

16. Domestic private health expenditure per capita (current US$)' | worldbank.org link: SH.XPD.PVTD.PC.CD

Gasto privado corriente en salud per cápita expresado en dólares estadounidenses corrientes. Las fuentes privadas nacionales incluyen fondos de hogares, corporaciones y organizaciones sin fines de lucro. Dichos gastos pueden pagarse por adelantado al seguro médico voluntario o pagarse directamente a los proveedores de atención médica.

17. Current health expenditure per capita (current US$)' | worldbank.org link: SH.XPD.CHEX.PC.CD

Gasto privado corriente en salud per cápita expresado en dólares estadounidenses corrientes. Las fuentes privadas nacionales incluyen fondos de hogares, corporaciones y organizaciones sin fines de lucro. Dichos gastos pueden pagarse por adelantado al seguro médico voluntario o pagarse directamente a los proveedores de atención médica.

18. Mortality caused by road traffic injury (per 100,000 population)' | worldbank.org link: SH.STA.TRAF.P5

La mortalidad causada por lesiones por accidentes de tránsito se estima en muertes por lesiones mortales por accidentes de tránsito por cada 100.000 habitantes.

19. People using at least basic sanitation services (% of population)' | worldbank.org link: SH.STA.BASS.ZS

El porcentaje de personas que utilizan al menos servicios básicos de saneamiento, es decir, instalaciones sanitarias mejoradas que no se comparten con otros hogares. Este indicador abarca tanto a las personas que utilizan servicios de saneamiento básicos como a aquellas

que utilizan servicios de saneamiento gestionados de forma segura. Las instalaciones de saneamiento mejoradas incluyen sistemas de descarga/vertido de agua a sistemas de alcantarillado, fosas sépticas o letrinas de pozo; letrinas de pozo mejoradas y ventiladas, sanitarios compuestos o letrinas de pozo con losas.

20. People using safely managed drinking water services (% of population)' | worldbank.org link: SH.H2O.SMDW.ZS

El porcentaje de personas que utilizan agua potable de una fuente mejorada que sea accesible en las instalaciones, disponible cuando sea necesario y libre de contaminación fecal y química prioritaria. Las fuentes de agua mejoradas incluyen agua entubada, perforaciones o pozos entubados, pozos excavados protegidos, manantiales protegidos y agua envasada o entregada.

21. Literacy rate, adult total (% of people ages 15 and above)' | worldbank.org link: SE.ADT.LITR.ZS

La tasa de alfabetización de adultos es el porcentaje de personas de 15 años o más que pueden leer y escribir y comprender una declaración breve y sencilla sobre su vida cotidiana.

22. Control of Corruption: Estimate' worldbank.org link: CC.EST

Control de la Corrupción captura percepciones de hasta qué punto el poder público se ejerce para beneficio privado, incluidas formas tanto pequeñas como grandes de corrupción, así como la "captura" del Estado por parte de élites e intereses privados. La estimación proporciona la puntuación del país en el indicador agregado, en unidades de una distribución normal estándar, es decir, entre aproximadamente -2,5 y 2,5.

23. Military expenditure (% of GDP)' | worldbank.org link: MS.MIL.XPND.GD.ZS

Los datos sobre gastos militares del SIPRI se derivan de la definición de la OTAN, que incluye todos los gastos corrientes y de capital en las fuerzas armadas, incluidas las fuerzas de mantenimiento de la paz; ministerios de defensa y otras agencias gubernamentales involucradas en proyectos de defensa; fuerzas paramilitares, si se considera que están entrenadas y equipadas para operaciones militares; y actividades espaciales militares. Dichos gastos incluyen personal militar y civil, incluidas pensiones de jubilación del personal militar y servicios sociales para el personal; operación y mantenimiento; obtención; investigación y desarrollo militar; y ayuda militar (en los gastos militares del país donante). Se excluyen la defensa civil y los gastos corrientes de actividades militares anteriores, como los beneficios para veteranos, la desmovilización, la conversión y la destrucción de armas. Sin embargo, esta definición no se puede aplicar a todos los países, ya que requeriría información mucho más detallada que la que está disponible sobre lo que se incluye en los presupuestos militares y en las partidas de gastos militares extrapresupuestarios. (Por ejemplo, los presupuestos militares pueden cubrir o no la defensa civil, las reservas y las fuerzas auxiliares, la policía y las fuerzas paramilitares, las fuerzas de doble propósito como la policía militar y civil, las subvenciones militares en especie, las pensiones del personal militar y las contribuciones a la seguridad social pagadas de una parte del gobierno a otra.)

24. General government final consumption expenditure (current US$)' | worldbank.org link: NE.CON.GOVT.CD

El gasto de consumo final del gobierno general (anteriormente consumo del gobierno general) incluye todos los gastos corrientes del gobierno para compras de bienes y servicios (incluida la remuneración de los empleados). También incluye la mayoría de los gastos en defensa y seguridad nacionales, pero excluye los gastos militares del gobierno que forman parte de la formación de capital del gobierno. Los datos están en dólares estadounidenses actuales.

25. Access to electricity (% of population)' | worldbank.org link: EG.ELC.ACCS.ZS

El acceso a la electricidad es el porcentaje de la población con acceso a la electricidad. Los datos sobre electrificación se recopilan de la industria, encuestas nacionales y fuentes internacionales.

26. Food exports (% of merchandise exports)' | worldbank.org link: TX.VAL.FOOD.ZS.UN

Los alimentos comprenden los productos de las secciones 0 (alimentos y animales vivos), 1 (bebidas y tabaco) y 4 (aceites y grasas animales y vegetales) de la CUCI y la división 22 de la CUCI (semillas, nueces y granos oleaginosos).

27. Food production index (2014-2016 = 100)' | worldbank.org link: AG.PRD.FOOD.XD

El índice de producción de alimentos cubre cultivos alimentarios que se consideran comestibles y que contienen nutrientes. Se excluyen el café y el té porque, aunque son comestibles, no tienen valor nutritivo.

28. Households and NPISHs final consumption expenditure (% of GDP)' | worldbank.org link: NE.CON.PRVT.ZS

El gasto de consumo final de los hogares (anteriormente consumo privado) es el valor de mercado de todos los bienes y servicios, incluidos los productos duraderos (como automóviles, lavadoras y computadoras domésticas), adquiridos por los hogares. Excluye las compras de viviendas pero incluye el alquiler imputado de las viviendas ocupadas por sus propietarios. También incluye pagos y tarifas a los gobiernos para obtener permisos y licencias. Aquí, el gasto de consumo de los hogares incluye los gastos de las instituciones sin fines de lucro que prestan servicios a los hogares, incluso cuando el país los informa por separado. Esta partida también incluye cualquier discrepancia estadística en el uso de recursos en relación con la oferta de recursos.

29. School enrollment, tertiary (gross), gender parity index (GPI)' | worldbank.org link: SE.ENR.TERT.FM.ZS

El índice de paridad de género para la tasa bruta de matriculación en educación terciaria es la proporción de mujeres y hombres matriculados en el nivel terciario en escuelas públicas y privadas.

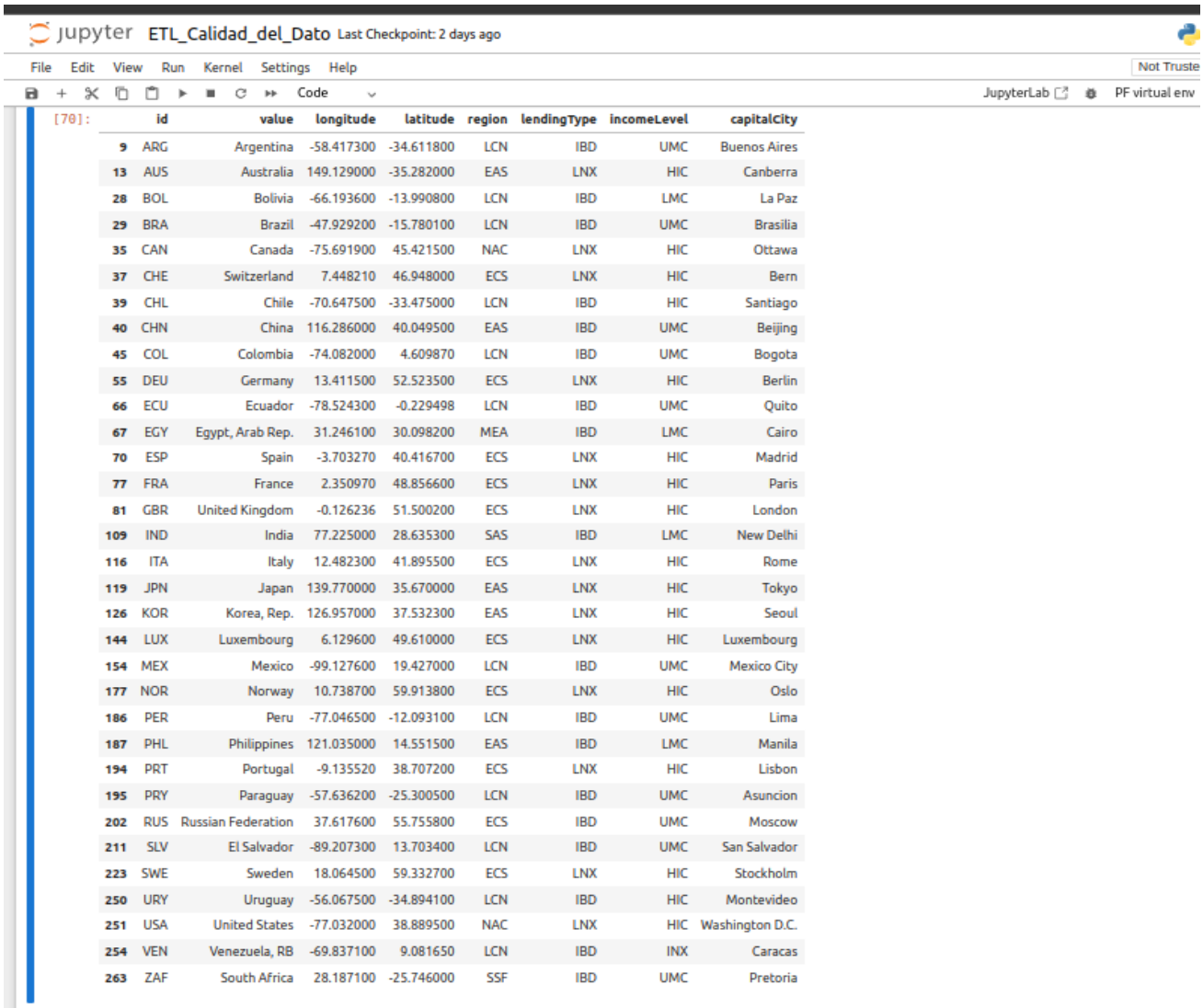
30. School enrollment, tertiary (% gross)', | worldbank.org link: [SE.ENR.TERT.FM.ZS]

El índice de paridad de género para la tasa bruta de matriculación en educación terciaria es la proporción de mujeres y hombres matriculados en el nivel terciario en escuelas públicas y privadas.

Estos factores socioeconómicos son importantes para comprender la situación de un país en términos de salud, desarrollo económico, acceso a servicios básicos y calidad de vida. Al analizar estos indicadores en conjunto, se puede obtener una imagen más completa de la situación socioeconómica de una nación y su posible impacto en la esperanza de vida de la población.

Se revisa la documentación proporcionada por las fuentes de datos para comprender el significado y la interpretación correcta de cada variable.

Países que formarán parte del Estudio.



	id	value	longitude	latitude	region	lendingType	incomeLevel	capitalCity
9	ARG	Argentina	-58.417300	-34.611800	LCN	IBD	UMC	Buenos Aires
13	AUS	Australia	149.129000	-35.282000	EAS	LNX	HIC	Canberra
28	BOL	Bolivia	-66.193600	-13.990800	LCN	IBD	LMC	La Paz
29	BRA	Brazil	-47.929200	-15.780100	LCN	IBD	UMC	Brasilia
35	CAN	Canada	-75.691900	45.421500	NAC	LNX	HIC	Ottawa
37	CHE	Switzerland	7.448210	46.948000	ECS	LNX	HIC	Bern
39	CHL	Chile	-70.647500	-33.475000	LCN	IBD	HIC	Santiago
40	CHN	China	116.286000	40.049500	EAS	IBD	UMC	Beijing
45	COL	Colombia	-74.082000	4.609870	LCN	IBD	UMC	Bogota
55	DEU	Germany	13.411500	52.523500	ECS	LNX	HIC	Berlin
66	ECU	Ecuador	-78.524300	-0.229498	LCN	IBD	UMC	Quito
67	EGY	Egypt, Arab Rep.	31.246100	30.098200	MEA	IBD	LMC	Cairo
70	ESP	Spain	-3.703270	40.416700	ECS	LNX	HIC	Madrid
77	FRA	France	2.350970	48.856600	ECS	LNX	HIC	Paris
81	GBR	United Kingdom	-0.126236	51.500200	ECS	LNX	HIC	London
109	IND	India	77.225000	28.635300	SAS	IBD	LMC	New Delhi
116	ITA	Italy	12.482300	41.895500	ECS	LNX	HIC	Rome
119	JPN	Japan	139.770000	35.670000	EAS	LNX	HIC	Tokyo
126	KOR	Korea, Rep.	126.957000	37.532300	EAS	LNX	HIC	Seoul
144	LUX	Luxembourg	6.129600	49.610000	ECS	LNX	HIC	Luxembourg
154	MEX	Mexico	-99.127600	19.427000	LCN	IBD	UMC	Mexico City
177	NOR	Norway	10.738700	59.913800	ECS	LNX	HIC	Oslo
186	PER	Peru	-77.046500	-12.093100	LCN	IBD	UMC	Lima
187	PHL	Philippines	121.035000	14.551500	EAS	IBD	LMC	Manila
194	PRT	Portugal	-9.135520	38.707200	ECS	LNX	HIC	Lisbon
195	PRY	Paraguay	-57.636200	-25.300500	LCN	IBD	UMC	Asuncion
202	RUS	Russian Federation	37.617600	55.755800	ECS	IBD	UMC	Moscow
211	SLV	El Salvador	-89.207300	13.703400	LCN	IBD	UMC	San Salvador
223	SWE	Sweden	18.064500	59.332700	ECS	LNX	HIC	Stockholm
250	URY	Uruguay	-56.067500	-34.894100	LCN	IBD	HIC	Montevideo
251	USA	United States	-77.032000	38.889500	NAC	LNX	HIC	Washington D.C.
254	VEN	Venezuela, RB	-69.837100	9.081650	LCN	IBD	INX	Caracas
263	ZAF	South Africa	28.187100	-25.746000	SSF	IBD	UMC	Pretoria

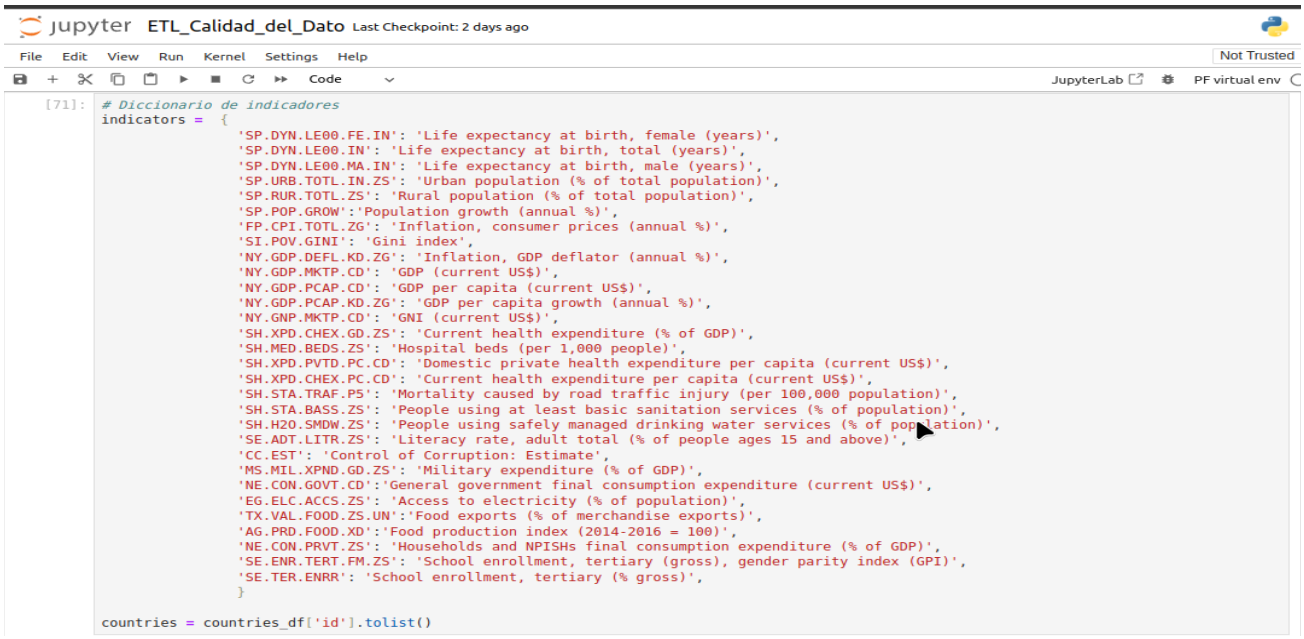
Se preseleccionaron un total de 50 países para ser evaluados con respecto a dos aspectos primordiales:

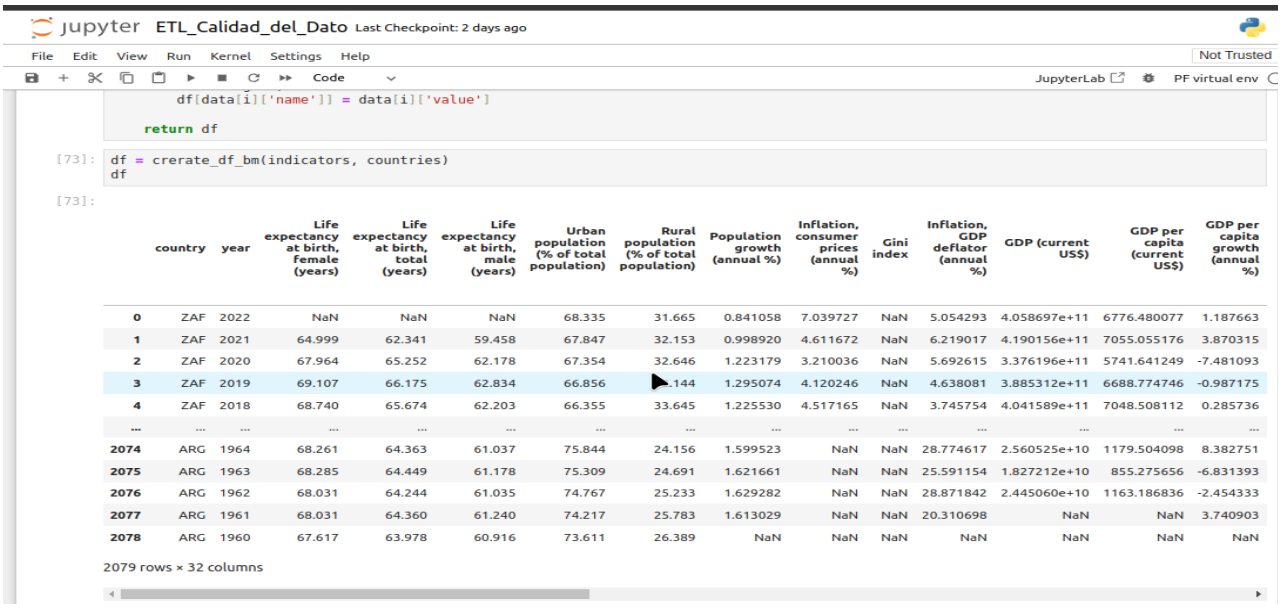
a) Los países deberán tener un a economía con **nivel de ingresos (incomen) minimo del tipo medio**
b) Los países deberán contener la mayor cantidad de datos para su evaluación respecto a los indicadores seleccionados.

De los resultados arrojados por el análisis fueron seleccionados treinta tres (33) países que se muestran en la tabla anterior.

Una vez seleccionados los países e indicadores para la realización del estudio se procede a generar un dataframe con los datos por medio de una función personalizada:

1. Listado de indicadores y dataframe resultante

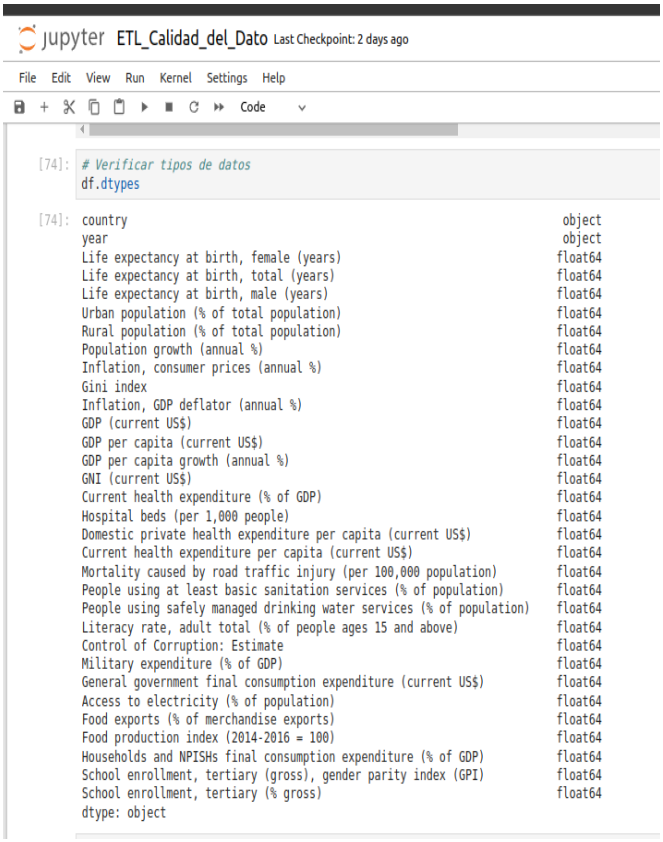




	country	year	Life expectancy at birth, female (years)	Life expectancy at birth, total (years)	Life expectancy at birth, male (years)	Urban population (% of total population)	Rural population (% of total population)	Population growth (annual %)	Inflation, consumer prices (annual %)	Gini index	Inflation, GDP deflator (annual %)	GDP (current US$)	GDP per capita (current US$)	GDP per capita growth (annual %)
0	ZAF	2022	NaN	NaN	NaN	68.335	31.665	0.841058	7.039727	NaN	5.054293	4.058697e+11	6776.480077	1.187663
1	ZAF	2021	64.999	62.341	59.458	67.847	32.153	0.998920	4.611672	NaN	6.219017	4.190156e+11	7055.055176	3.870315
2	ZAF	2020	67.964	65.252	62.178	67.354	32.646	1.223179	3.210036	NaN	5.692615	3.376196e+11	5741.641249	-7.481093
3	ZAF	2019	69.107	66.175	62.834	66.856	33.144	1.295074	4.120246	NaN	4.638081	3.885312e+11	6688.774746	-0.987175
4	ZAF	2018	68.740	65.674	62.203	66.355	33.645	1.225530	4.517165	NaN	3.745754	4.041589e+11	7048.508112	0.285736
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2074	ARG	1964	68.261	64.363	61.037	75.844	24.156	1.599523	NaN	NaN	28.774617	2.560525e+10	1179.504098	8.382751
2075	ARG	1963	68.285	64.449	61.178	75.309	24.691	1.621661	NaN	NaN	25.591154	1.827212e+10	855.275656	-6.831393
2076	ARG	1962	68.031	64.244	61.035	74.767	25.233	1.629282	NaN	NaN	28.871842	2.445060e+10	1163.186836	-2.454333
2077	ARG	1961	68.031	64.360	61.240	74.217	25.783	1.613029	NaN	NaN	20.310698	NaN	NaN	3.740903
2078	ARG	1960	67.617	63.978	60.916	73.611	26.389	NaN	NaN	NaN	NaN	NaN	NaN	NaN

2079 rows × 32 columns

2. Visión general de la estructura del dataframe, información sobre los tipos de datos, la presencia de valores nulos y la memoria utilizada.



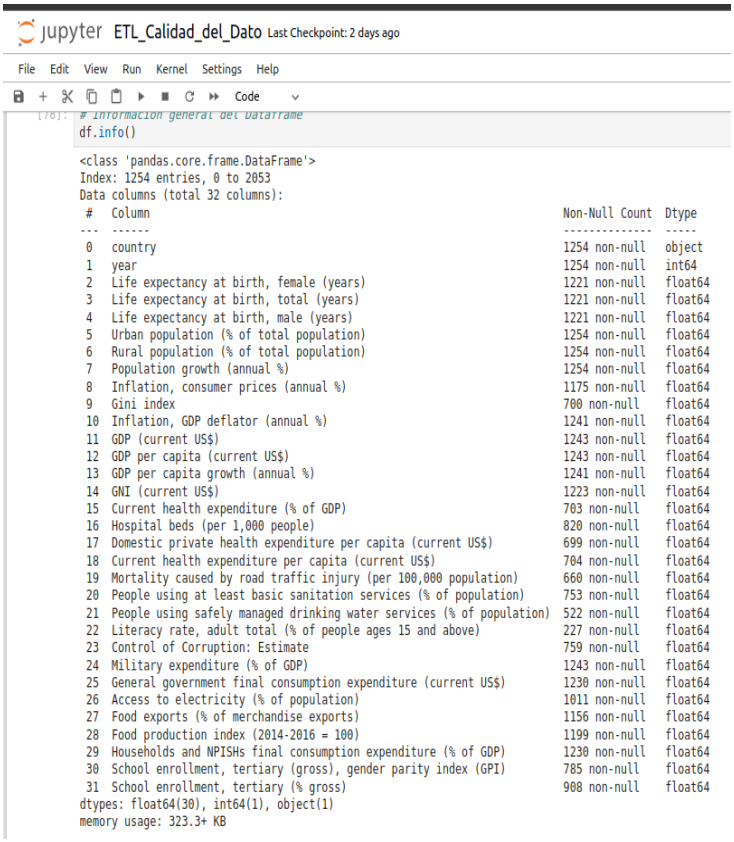
jupyter ETL_Calidad_del_Dato Last Checkpoint: 2 days ago

File Edit View Run Kernel Settings Help

Code

```
[74]: # Verificar tipos de datos
      df.dtypes
```

```
[74]: country                                                          object
      year                                                             object
      Life expectancy at birth, female (years)                        float64
      Life expectancy at birth, total (years)                         float64
      Life expectancy at birth, male (years)                          float64
      Urban population (% of total population)                        float64
      Rural population (% of total population)                        float64
      Population growth (annual %)                                    float64
      Inflation, consumer prices (annual %)                           float64
      Gini index                                                      float64
      Inflation, GDP deflator (annual %)                              float64
      GDP (current US$)                                               float64
      GDP per capita (current US$)                                    float64
      GDP per capita growth (annual %)                                float64
      GNI (current US$)                                               float64
      Current health expenditure (% of GDP)                           float64
      Hospital beds (per 1,000 people)                                float64
      Domestic private health expenditure per capita (current US$)    float64
      Current health expenditure per capita (current US$)             float64
      Mortality caused by road traffic injury (per 100,000 population) float64
      People using at least basic sanitation services (% of population) float64
      People using safely managed drinking water services (% of population) float64
      Literacy rate, adult total (% of people ages 15 and above)      float64
      Control of Corruption: Estimate                                 float64
      Military expenditure (% of GDP)                                 float64
      General government final consumption expenditure (current US$)  float64
      Access to electricity (% of population)                         float64
      Food exports (% of merchandise exports)                         float64
      Food production index (2014-2016 = 100)                         float64
      Households and NPISHs final consumption expenditure (% of GDP)  float64
      School enrollment, tertiary (gross), gender parity index (GPI)  float64
      School enrollment, tertiary (% gross)                           float64
      dtype: object
```



jupyter ETL_Calidad_del_Dato Last Checkpoint: 2 days ago

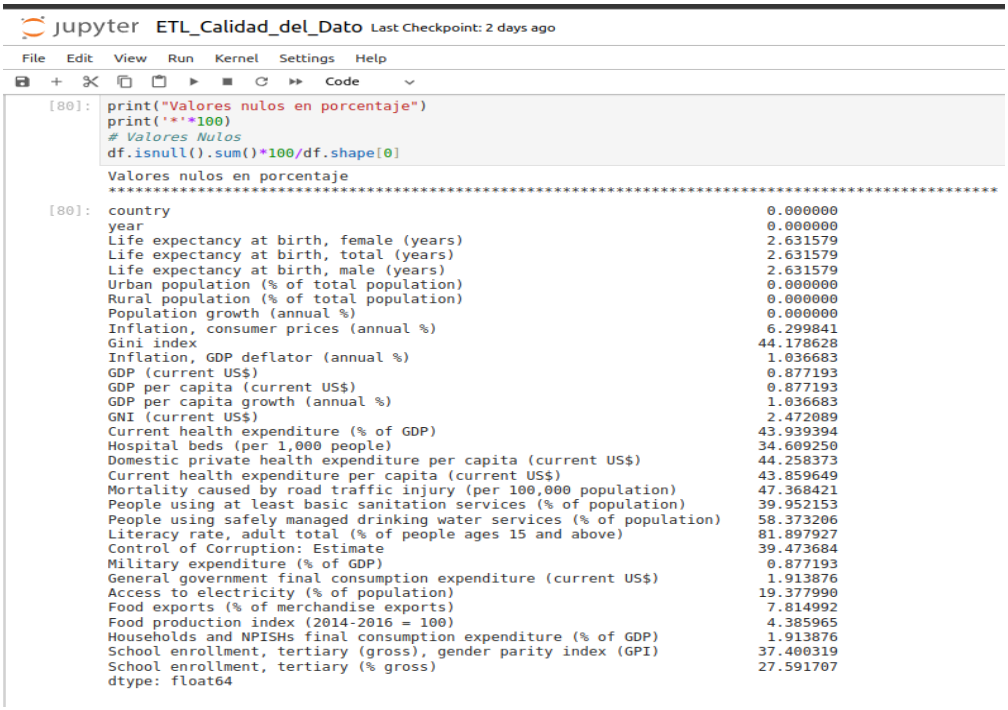
File Edit View Run Kernel Settings Help

Code

```
[76]: # Información general del DataFrame
      df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1254 entries, 0 to 2053
Data columns (total 32 columns):
 #   Column                                                          Non-Null Count  Dtype
---  ------                                                          --------------  -----
 0   country                                                         1254 non-null   object
 1   year                                                            1254 non-null   int64
 2   Life expectancy at birth, female (years)                        1221 non-null   float64
 3   Life expectancy at birth, total (years)                         1221 non-null   float64
 4   Life expectancy at birth, male (years)                          1221 non-null   float64
 5   Urban population (% of total population)                        1254 non-null   float64
 6   Rural population (% of total population)                        1254 non-null   float64
 7   Population growth (annual %)                                    1254 non-null   float64
 8   Inflation, consumer prices (annual %)                           1175 non-null   float64
 9   Gini index                                                      700 non-null    float64
 10  Inflation, GDP deflator (annual %)                              1241 non-null   float64
 11  GDP (current US$)                                               1243 non-null   float64
 12  GDP per capita (current US$)                                    1243 non-null   float64
 13  GDP per capita growth (annual %)                                1241 non-null   float64
 14  GNI (current US$)                                               1223 non-null   float64
 15  Current health expenditure (% of GDP)                           703 non-null    float64
 16  Hospital beds (per 1,000 people)                                820 non-null    float64
 17  Domestic private health expenditure per capita (current US$)    699 non-null    float64
 18  Current health expenditure per capita (current US$)             704 non-null    float64
 19  Mortality caused by road traffic injury (per 100,000 population) 660 non-null   float64
 20  People using at least basic sanitation services (% of population) 753 non-null  float64
 21  People using safely managed drinking water services (% of population) 522 non-null  float64
 22  Literacy rate, adult total (% of people ages 15 and above)      227 non-null    float64
 23  Control of Corruption: Estimate                                 759 non-null    float64
 24  Military expenditure (% of GDP)                                 1243 non-null   float64
 25  General government final consumption expenditure (current US$)  1230 non-null   float64
 26  Access to electricity (% of population)                         1011 non-null   float64
 27  Food exports (% of merchandise exports)                         1156 non-null   float64
 28  Food production index (2014-2016 = 100)                         1199 non-null   float64
 29  Households and NPISHs final consumption expenditure (% of GDP)  1230 non-null   float64
 30  School enrollment, tertiary (gross), gender parity index (GPI)  785 non-null    float64
 31  School enrollment, tertiary (% gross)                           908 non-null    float64
dtypes: float64(30), int64(1), object(1)
memory usage: 323.3+ KB
```

3. Cálculo de porcentajes de valores nulos por variable



jupyter ETL_Calidad_del_Dato Last Checkpoint: 2 days ago

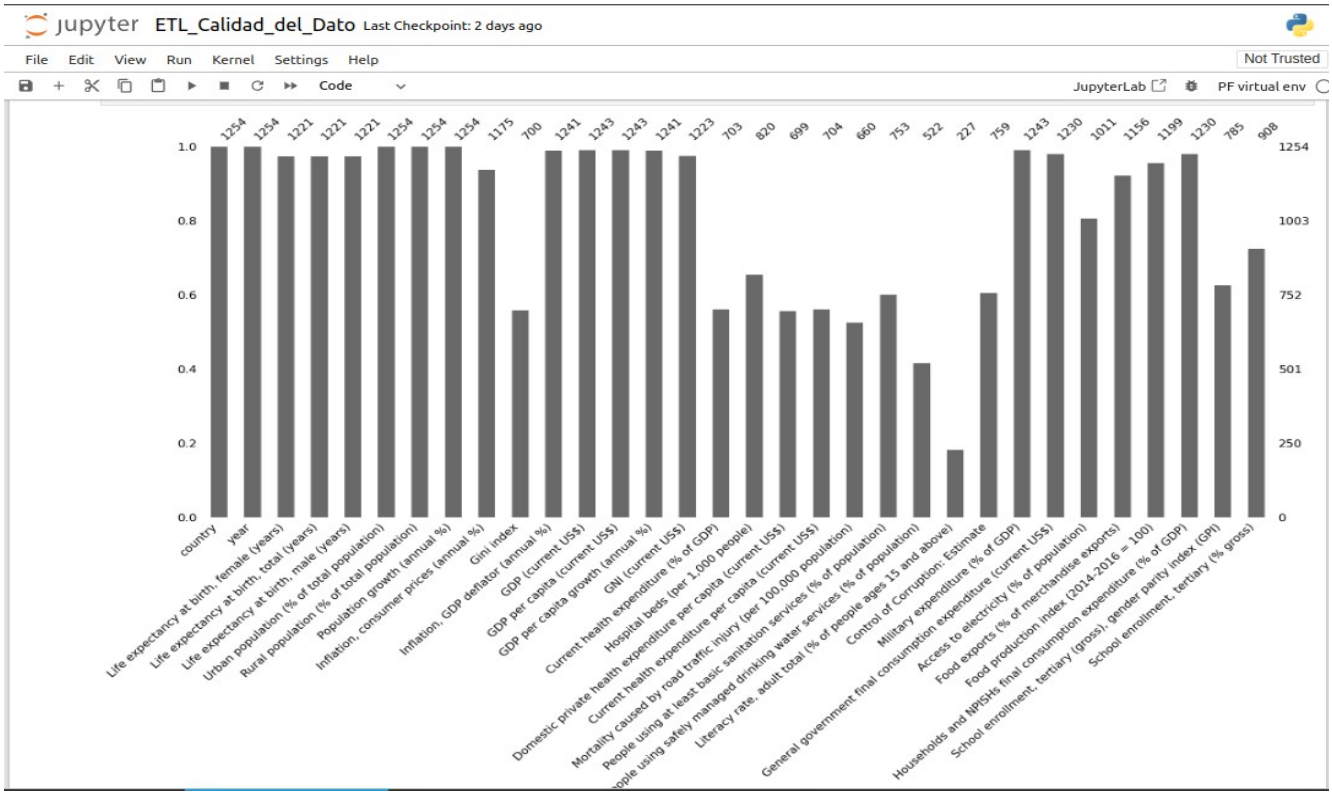
File Edit View Run Kernel Settings Help

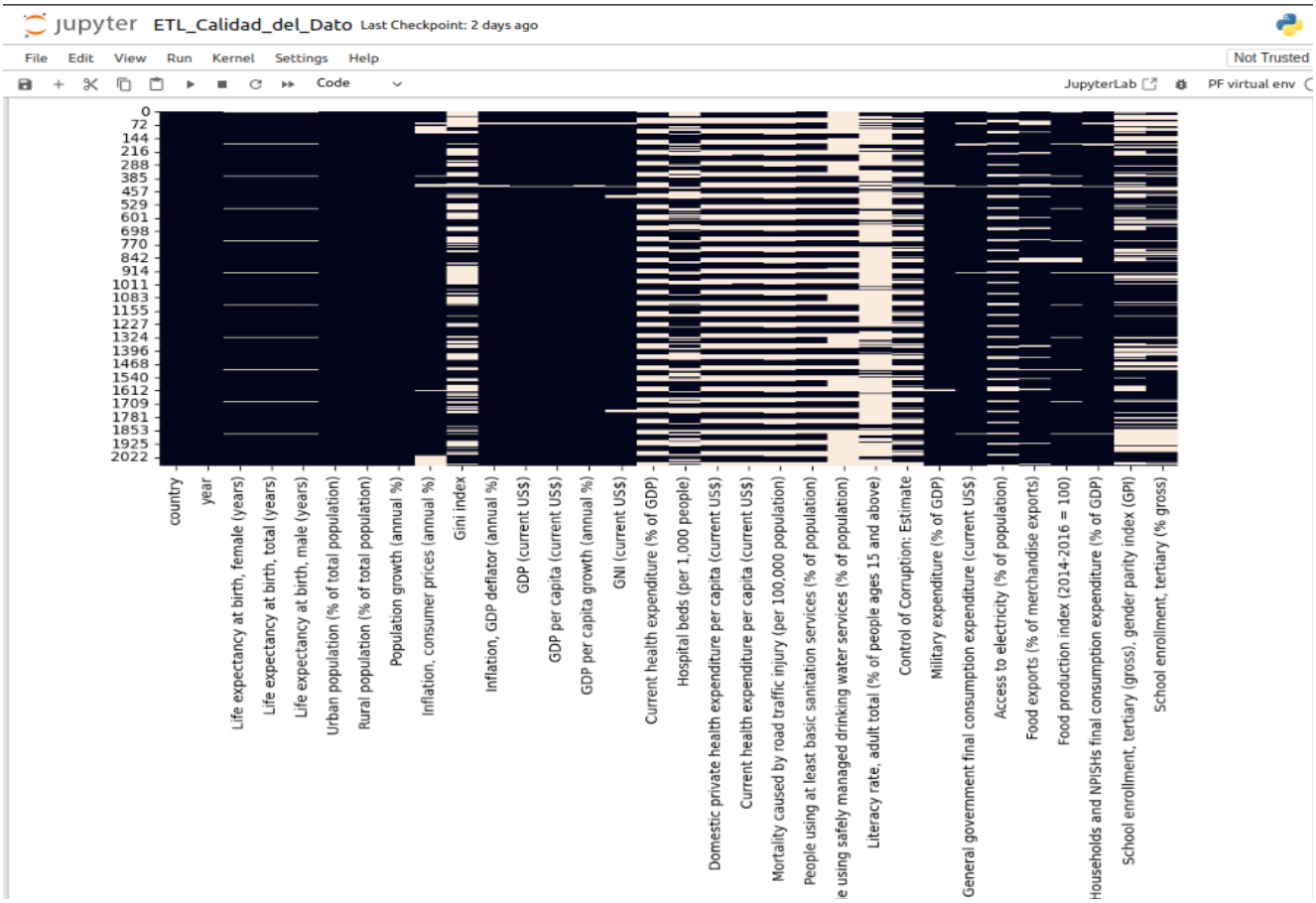
Code

```
[80]: print("Valores nulos en porcentaje")
      print('*'*100)
      # Valores Nulos
      df.isnull().sum()*100/df.shape[0]
```

```
Valores nulos en porcentaje
****************************************************************************************************
[80]: country                                                          0.000000
      year                                                             0.000000
      Life expectancy at birth, female (years)                         2.631579
      Life expectancy at birth, total (years)                          2.631579
      Life expectancy at birth, male (years)                           2.631579
      Urban population (% of total population)                         0.000000
      Rural population (% of total population)                         0.000000
      Population growth (annual %)                                     0.000000
      Inflation, consumer prices (annual %)                            6.299841
      Gini index                                                      44.178628
      Inflation, GDP deflator (annual %)                               1.036683
      GDP (current US$)                                                0.877193
      GDP per capita (current US$)                                     0.877193
      GDP per capita growth (annual %)                                 1.036683
      GNI (current US$)                                                2.472089
      Current health expenditure (% of GDP)                           43.939394
      Hospital beds (per 1,000 people)                                34.609250
      Domestic private health expenditure per capita (current US$)    44.258373
      Current health expenditure per capita (current US$)             43.859649
      Mortality caused by road traffic injury (per 100,000 population) 47.368421
      People using at least basic sanitation services (% of population) 39.952153
      People using safely managed drinking water services (% of population) 58.373206
      Literacy rate, adult total (% of people ages 15 and above)      81.897927
      Control of Corruption: Estimate                                 39.473684
      Military expenditure (% of GDP)                                  0.877193
      General government final consumption expenditure (current US$)   1.913876
      Access to electricity (% of population)                         19.377990
      Food exports (% of merchandise exports)                          7.814992
      Food production index (2014-2016 = 100)                          4.385965
      Households and NPISHs final consumption expenditure (% of GDP)   1.913876
      School enrollment, tertiary (gross), gender parity index (GPI)  37.400319
      School enrollment, tertiary (% gross)                           27.591707
      dtype: float64
```

4. Visualización de variables con valores faltantes y distribución de los mismos





5. Selección de indicadores con menos del 30% de datos nulos.

jupyter ETL_Calidad_del_Dato Last Checkpoint: 2 days ago

File Edit View Run Kernel Settings Help Not Trusted

Code JupyterLab PF virtual env

Eliminar columna con más del 30% de datos faltantes

[84]:
```
# Columnas a eliminar
drop_columns = [
                'Gini index',
                'Current health expenditure (% of GDP)',
                'Hospital beds (per 1,000 people)',
                'Domestic private health expenditure per capita (current US$)',
                'Current health expenditure per capita (current US$)',
                'Mortality caused by road traffic injury (per 100,000 population)',
                'People using at least basic sanitation services (% of population)',
                'People using safely managed drinking water services (% of population)',
                'Literacy rate, adult total (% of people ages 15 and above)',
                'Control of Corruption: Estimate',
                'School enrollment, tertiary (gross), gender parity index (GPI)'
            ]
```

6. Visualizar nueva estructura del dataframe.

jupyter ETL_Calidad_del_Dato Last Checkpoint: 2 days ago

File Edit View Run Kernel Settings Help Not Trusted

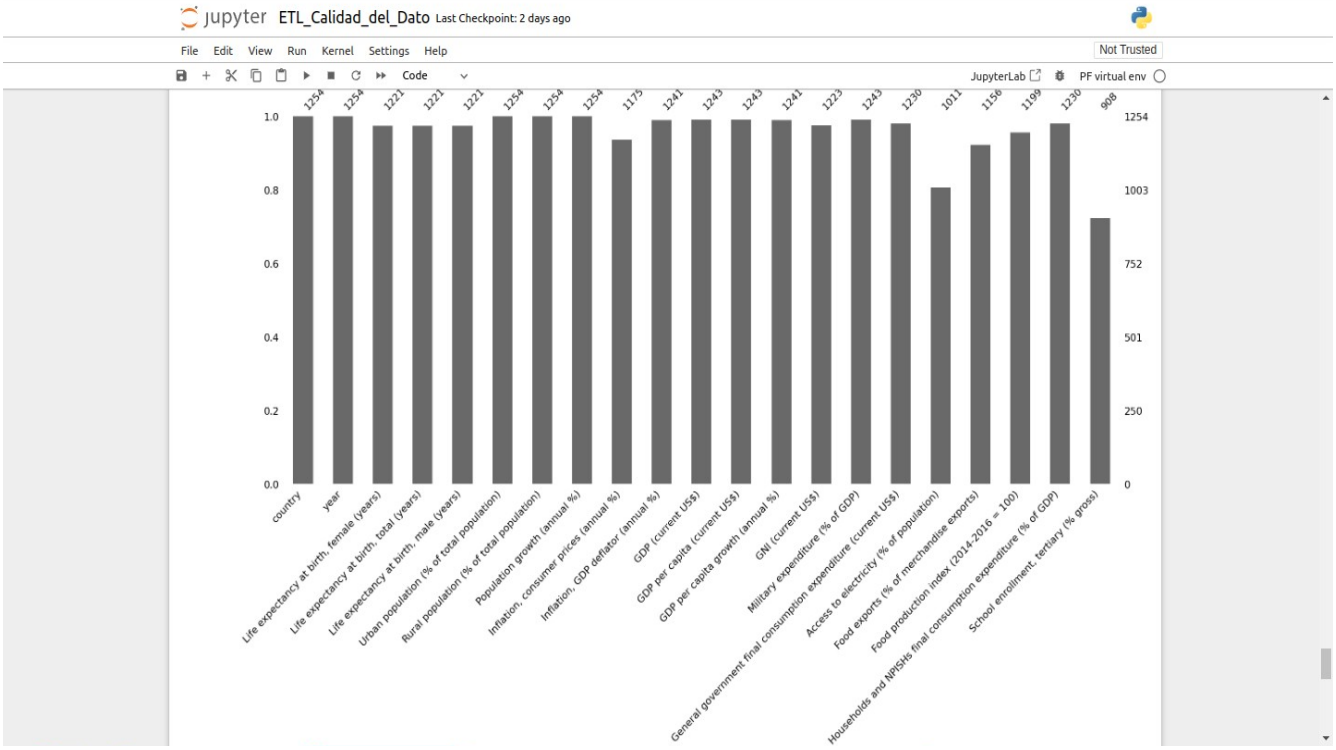
Code JupyterLab PF virtual env

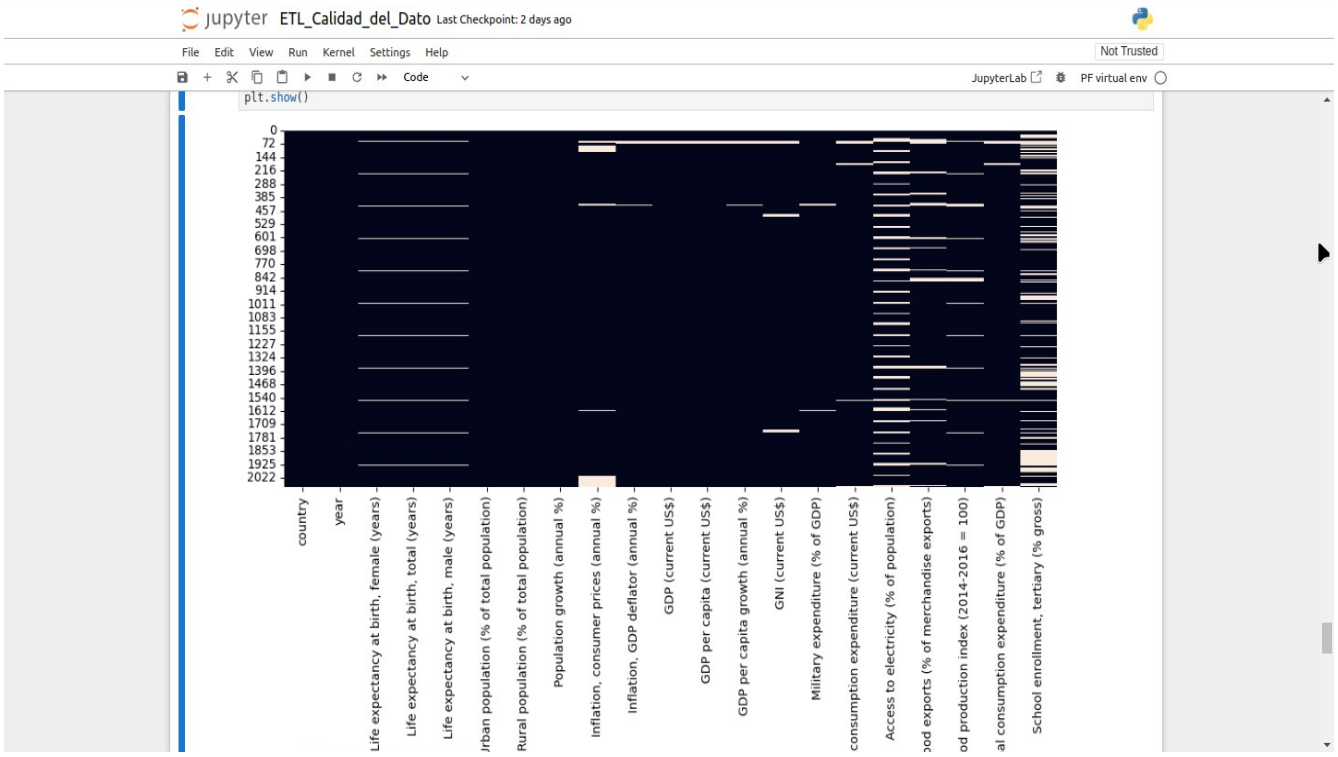
[85]:

	country	year	Life expectancy at birth, female (years)	Life expectancy at birth, total (years)	Life expectancy at birth, male (years)	Urban population (% of total population)	Rural population (% of total population)	Population growth (annual %)	Inflation, consumer prices (annual %)	Inflation, GDP deflator (annual %)	GDP (current US$)	GDP per capita (current US$)	GDP per capita growth (annual %)	GN
0	ZAF	2022	NaN	NaN	NaN	68.335	31.665	0.841058	7.039727	5.054293	4.058697e+11	6776.480077	1.187663	3.97
1	ZAF	2021	64.999	62.341	59.458	67.847	32.153	0.998920	4.611672	6.219017	4.190156e+11	7055.055176	3.870315	4.10
2	ZAF	2020	67.964	65.252	62.178	67.354	32.646	1.223179	3.210036	5.692615	3.376196e+11	5741.641249	-7.481093	3.31
3	ZAF	2019	69.107	66.175	62.834	66.856	33.144	1.295074	4.120246	4.638081	3.885312e+11	6688.774746	-0.987175	3.78
4	ZAF	2018	68.740	65.674	62.203	66.355	33.645	1.225530	4.517165	3.745754	4.041589e+11	7048.508112	0.285736	3.92
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2049	ARG	1989	74.735	71.425	68.090	86.613	13.387	1.487655	NaN	3046.091152	7.662973e+10	2382.338066	-8.527911	7.02
2050	ARG	1988	74.381	71.052	67.711	86.233	13.767	1.508087	NaN	381.246344	1.268902e+11	4004.009589	-2.570392	1.21
2051	ARG	1987	74.004	70.564	67.148	85.843	14.157	1.510974	NaN	127.539918	1.088109e+11	3485.690232	1.164690	1.03
2052	ARG	1986	73.638	70.119	66.655	85.445	14.555	1.511328	NaN	77.292237	1.058724e+11	3443.191431	4.561114	1.01
2053	ARG	1985	73.253	69.651	66.141	85.038	14.962	1.513403	NaN	607.447498	8.815089e+10	2910.508329	-6.613093	8.34

1254 rows × 21 columns

7. Se verifican las variables con datos faltantes una vez realizado la eliminación de las columnas con problemas y la nueva distribución



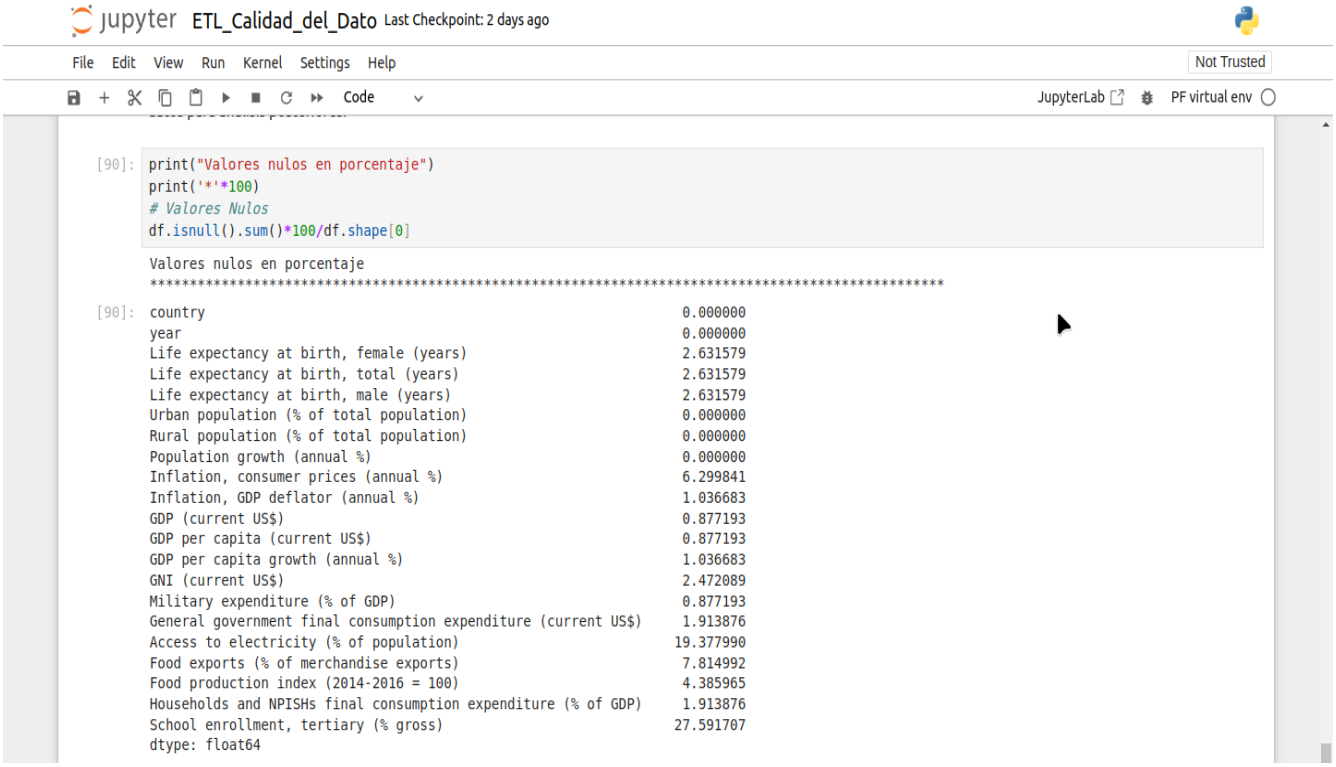


8. Evaluación de integridad de los datos

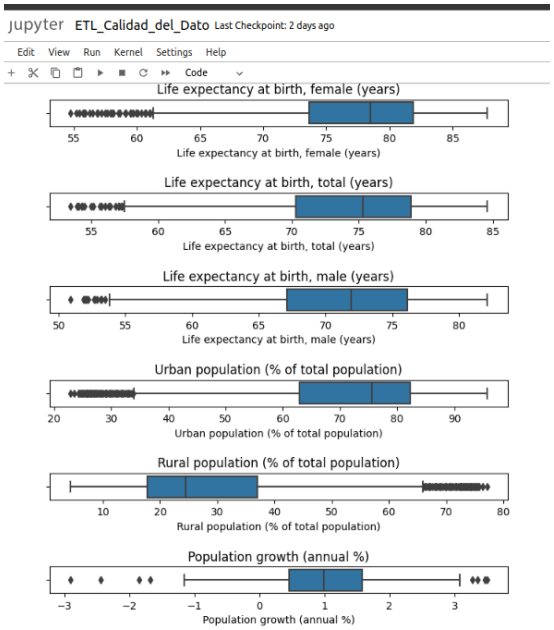
Se verifica si los datos recopilados tienen todos los campos requeridos, es decir si existe algún campo faltante o si existes registros duplicados. Se examina la consistencia de los datos en términos de formatos, unidades de medida

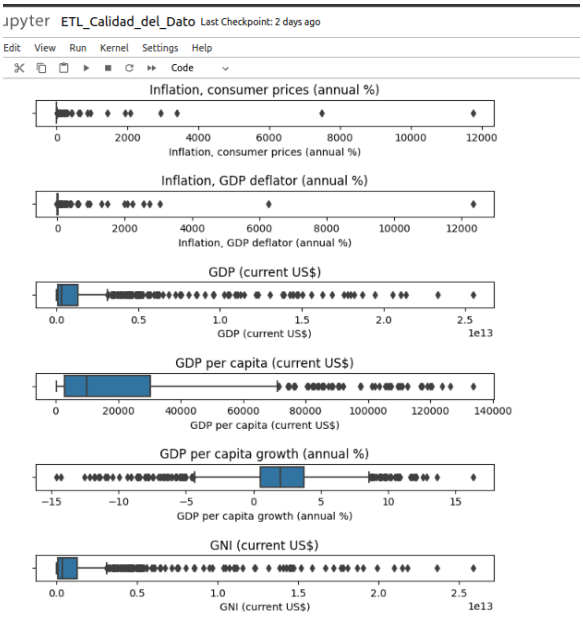
8.1. Verificación de registros repetidos y eliminación de los mismos de existir

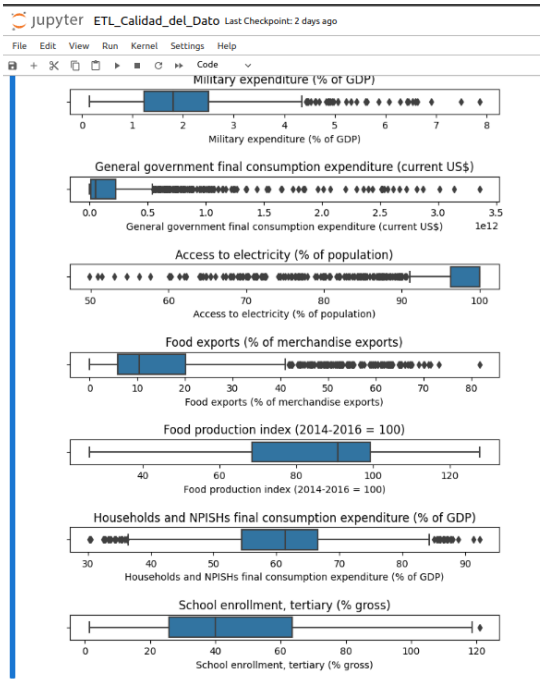
8.2. Determinación de porcentaje de valores nulos por variable.



8.3. Evaluación de valores extremos (outliers).







Observaciones:

En los gráficos anteriores se evidencian valores que se encuentran fuera de los parámetros normales los cuales se pueden tratar como outliers, sin embargo tomando en cuenta que la data r**epresenta economías de distintos tamaños y fortaleza** la existencia de esos valores es factible, por lo que se decide mantener todo el conjunto de datos para análisis posteriores.

Calidad de los Datos

Aquí detallamos las seis dimensiones principales de la calidad de los datos:

1. **Precisión:** los datos deben ser reales; la medida de precisión se puede confirmar con una fuente verificable.
2. **Completitud:** es la capacidad de los datos para entregar efectivamente todos los valores requeridos que están disponibles. Tus datos deben estar completos. La información incompleta muchas veces es inutilizable o lleva a errores operacionales.
3. **Consistencia:** se refiere a la uniformidad de los datos a medida que se mueven a través de redes y aplicaciones. Los mismos valores de datos almacenados en ubicaciones diferentes no deben entrar en conflicto entre sí. Si la información disponible no coincide con la almacenada en otras bases de datos, todos los datos se ponen en duda.
4. **Validez:** los datos deben recopilarse de acuerdo con reglas y parámetros comerciales definidos, y deben ajustarse al formato correcto y estar dentro del rango correcto.
5. **Unicidad:** esta dimensión garantiza que no haya duplicaciones ni superposiciones de valores en todos los conjuntos de datos. La limpieza y la deduplicación de datos pueden ayudar a remediar una puntuación de unicidad baja.
6. **Oportunidad:** La información cambia constantemente y los datos obsoletos pueden no ser representativos de la situación actual. Los datos oportunos son datos que están disponibles cuando se requieren. Implican la actualización constante, en tiempo real, para garantizar que estén fácilmente disponibles y accesibles.

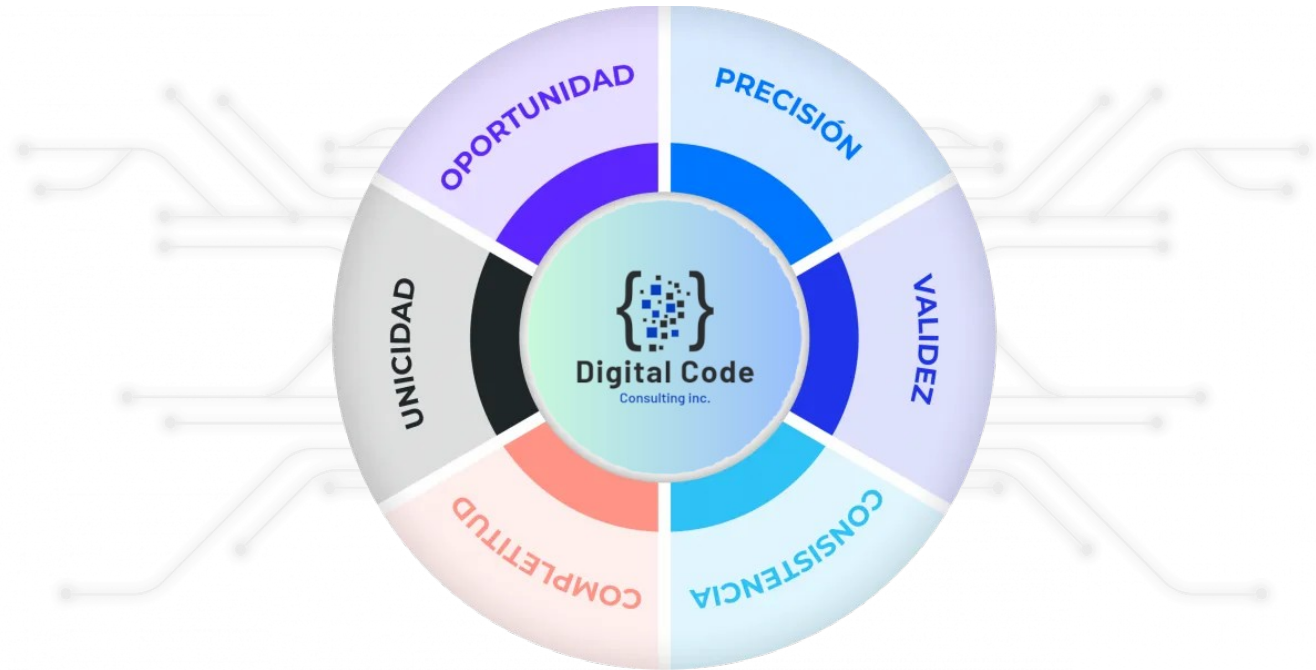
La calidad de los datos se evalúa en función de una serie de dimensiones que pueden variar según la fuente de información. Estas dimensiones se utilizan para categorizar las métricas de calidad de datos:

Completitud: representa la cantidad de datos que que se pueden utilizar o que están completos. Si hay un alto porcentaje de valores omitidos, se puede generar un análisis sesgado o erróneo si los datos no son representativos de una muestra de datos típica.
Unicidad: representa la cantidad de datos duplicados en un conjunto de datos. Por ejemplo, al revisar los datos de cliente, lo lógico es que cada cliente tenga un ID de cliente único.
Validez: esta dimensión mide cuántos datos coinciden con el formato necesario para las reglas de negocio. El formateo suele incluir metadatos, como tipos de datos válidos, rangos, patrones, etc.
Oportunidad: esta dimensión hace referencia a la disponibilidad de los datos dentro de un marco de tiempo previsible. Por ejemplo, los clientes esperan recibir un número de pedido inmediatamente después de haber realizado una compra, y esos datos se deben generar en tiempo real.
Exactitud: esta dimensión se refiere a la precisión de los valores de datos en función de la «fuente de verdad» acordada. Dado que puede haber varios orígenes que informen de la

misma métrica, es importante designar un origen de datos primario, y se pueden utilizar otros orígenes de datos para confirmar la exactitud del primario. Por ejemplo, las herramientas pueden comprobar si la tendencia de cada origen de datos sigue la misma dirección para reforzar la confianza en la exactitud de los datos.

Coherencia: esta dimensión evalúa los registros de datos de dos conjuntos de datos diferentes. Como ya hemos mencionado, varios orígenes pueden informar sobre una misma métrica. El uso de diferentes orígenes para comprobar el comportamiento y las las tendencias de coherencia de los datos aporta fiabilidad a la información procesable de los análisis de las organizaciones. Esta lógica también se puede aplicar en torno a las relaciones entre datos. Por ejemplo, el número de empleados de un departamento no debe superar al número total de empleados de una compañía.

Adecuación para un propósito: por último, la adecuación a un propósito permite comprobar que el activo de datos cumple con una necesidad de negocio. Esta dimensión puede ser difícil de evaluar, en especial en el caso de nuevos conjuntos de datos emergentes.



Data Engineering

La etapa de **Data Engineering** se centrará en la construcción de la infraestructura de almacenamiento, procesamiento y transformación de los datos de manera eficiente y oportuna para ser servidos a los procesos de Data Analitics (inteligencia de negocios) y Machine Learning (Modelado) con alta calidad y de manera confiable.

Modelo Entidad - Relación:

Se iniciará esta etapa con el modelo de datos relacionales el cual tiene como objetivo el crear un modelo entidad relación ER adecuado y detallado, se deben especificar las tablas que

representan entidades relevantes, las relaciones entre éstas y los tipos de datos que se utilizaran, siempre en el contexto del proyecto que nos ocupa “Esperanza de Vida al Nacer”.

Diseño del modelo Entidad – Relación ER:

A continuación y siguiendo las pautas para un diseño adecuado del modelo ER para el proyecto identificamos las entidades, se definen los atributos y se establece la relación entre estas.

a) **Entidades:**

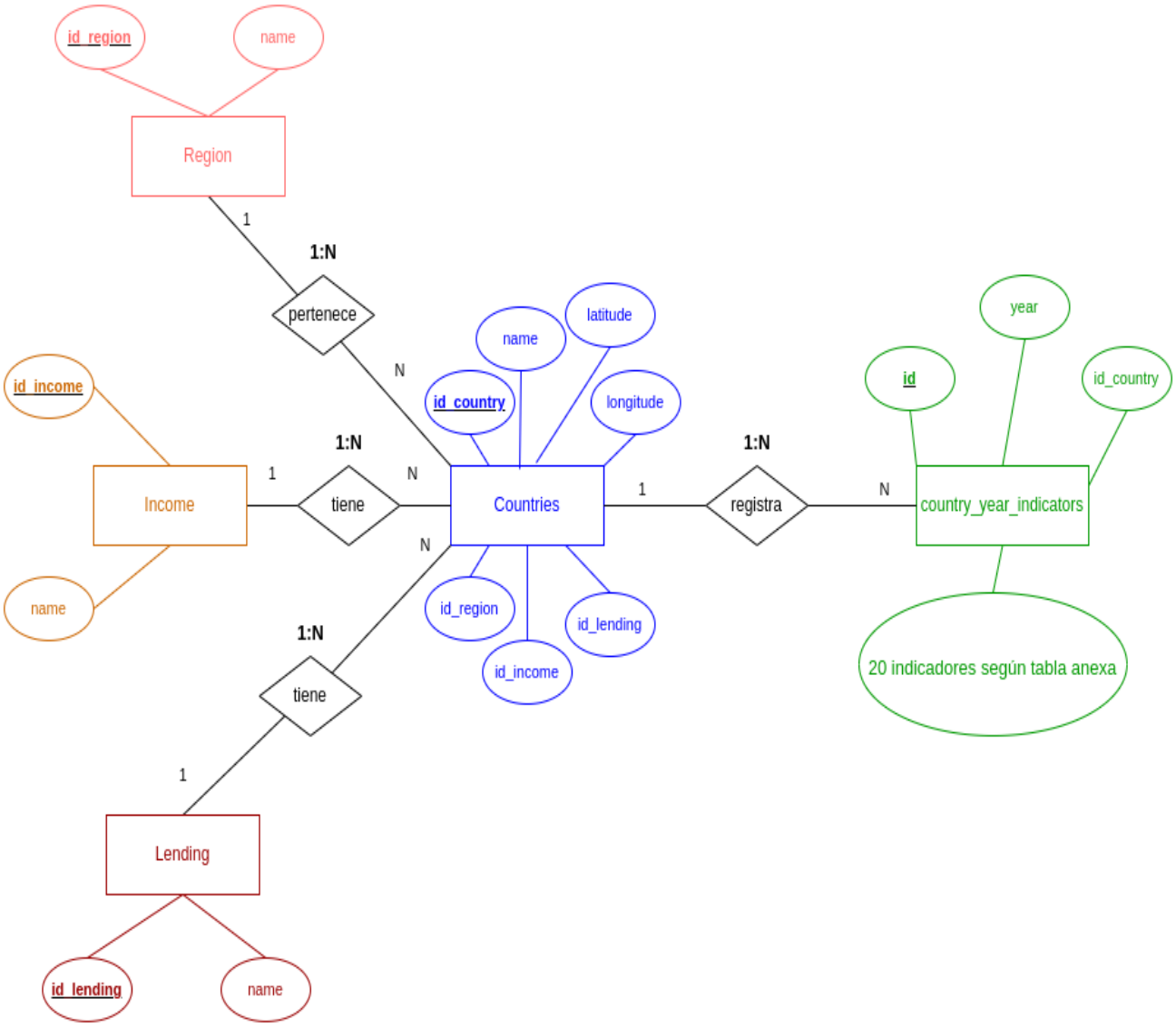
Región – Tabla: Region	
id_region	varchar(5) not Null primary_key
name	Varchar(250) not null
Ingresos – Tabla: Income	
id_income	varchar(5) not Null primary_key
name	Varchar(250) not null
Préstamo – Tabla: Lending	
id_lending	varchar(5) not Null primary_key
name	Varchar(250) not null
Países – Tabla: Countries	
id_country	varchar(5) not Null primary_key
name	Varchar(250) not null
longitude	Decimal(10,7) not null
latitude	Decimal(10,7) not null
id_region	Varchar(5) (foreign key on table Region)
id_income	Varchar(5) (foreign key on table Income)
id_lending	Varchar(5) (foreign key on table Lending)
Indicadores – Tabla: country_year_indicators	
Id	int auto_increment primary key
year	int not null
id_country	varchar(5) not Null (foreign key on table Countries)
Life expectancy at birth, female (years)	
Life expectancy at birth, total (years)	

Life expectancy at birth, male (years)	
Urban population (% of total population)	
Rural population (% of total population)	
Population growth (annual %)	
Inflation, consumer prices (annual %)	
Inflation, GDP deflator (annual %)	
GDP (current US$)	
GDP per capita (current US$)	
GDP per capita growth (annual %)	
GNI (current US$)	
Military expenditure (% of GDP	
General government final consumption expenditure (current US$)	
Access to electricity (% of population)	
Food exports (% of merchandise exports)	
Food production index (2014-2016 = 100)	
Households and NPISHs final consumption expenditure (% of GDP)	
School enrollment, tertiary (% gross)	
Index of Economic Freedom	

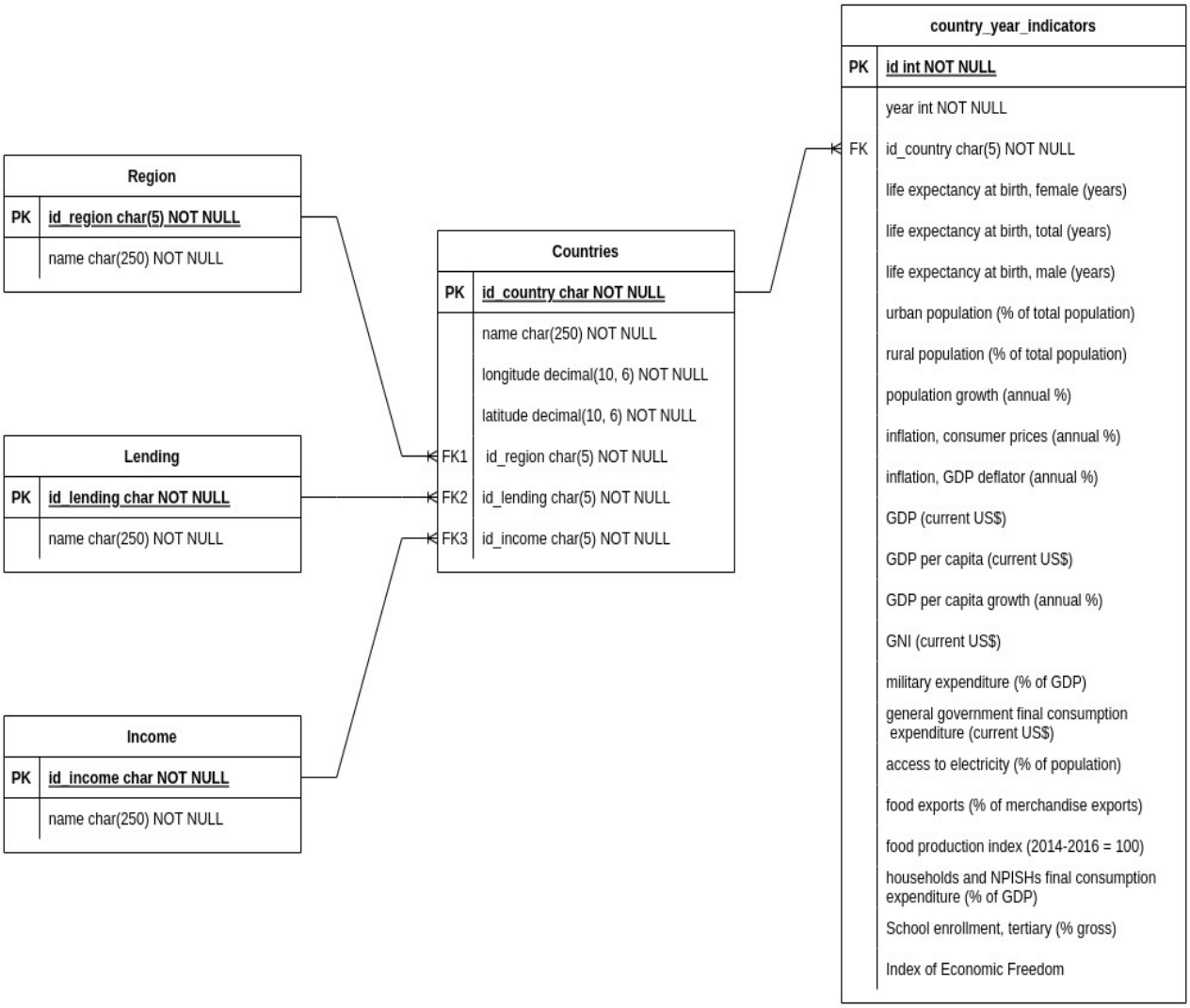
b) Relaciones

Tabla	Tipo Relación	Tabla
Countries	**N : 1**	Region
Countries	**N : 1**	Income
Countries	**N : 1**	Lending
Countries	**1 : N**	Country_year_indicators

c) Diagramas


id_region
name
Region
1
1:N
pertenece
N
name
latitude
id_country
longitude
id_income
1:N
1:N
Income
1
tiene
N
Countries
1
registra
N
country_year_indicators
id
year
id_country
name
N
id_region
id_lending
id_income
20 indicadores según tabla anexa
1:N
tiene
1
Lending
id_lending
name

Modelo Entidad – Relación (MER)



Modelo Relacional (MR)

La relaciones entre las diferentes entidades establecida e indicadas en los gráficos anteriores queda de la siguiente manera:

Tabla: Region:

- Relacionado con tabla **Countries**
  - **Cardinalidad : Uno a Muchos → Una “Region” se puede asociar a uno o varios países identificados en la tabla Countries**.
  - **Participación : Region tiene una participación Total (Todas las regiones deben tener al menos un país asociados) | Country tiene una participación Total (Todos los paises deben estar asociados a una Region)**

Tabla: Lending:

- Relacionada con tabla **Countries**
  - **Uno a Muchos → Una “Lending” (Tipo de Préstamo) se puede asociar a uno o varios países identificados en la tabla Countries**.
  - **Participación : Lending tiene una participación Total (Todas las Lending deben tener al menos un país asociados) | Country tiene una participación Total (Todos los países deben estar asociados a una Lending)**

Tabla: Income:

- Relacionada con tabla **Countries**
  - **Uno a Muchos → Una “Income” (Tipo de Economía) se puede asociar a uno o varios países identificados en la tabla Countries**.
  - **Participación : Income tiene una participación Total (Todas las Income deben tener al menos un país asociados) | Country tiene una participación Total (Todos los países deben estar asociados a una Income)**

Tabla: Countries:

- Relacionada con tabla **Region**
  - **Muchos a Uno → Varios países se puede asociar a una región identificados en la tabla Region**.
- Relacionada con tabla **Lending**
  - **Muchos a Uno → Varios países se puede asociar a un tipo de préstamo identificados en la tabla Leanding**.
- Relacionada con tabla **Income**
  - **Muchos a Uno → Varios países se puede asociar a un tipo de economía identificados en la tabla Income**.
- Relacionada con tabla **Country_year_indicators**
  - **Uno a Muchos → Un “País” se puede asociar a uno o varios registros identificados en la tabla Country_year_indicators.**

Tabal: Country_year_indicators

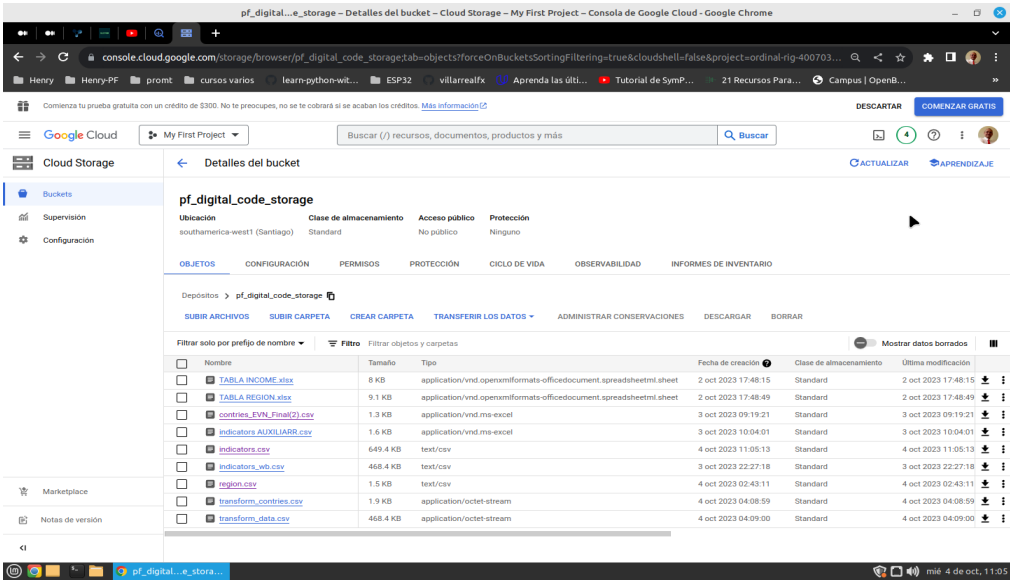
- **Relacionada con tabla Countries**
  - **Muchos a Uno → Varios registros se puede asociar a un país identificados en la tabla Countries**.
  - **Participación : Indicators tiene una participación Total (Todas las Indicadores por fecha deben tener al menos un país asociados) | Country tiene una participación Parcial (puede existir un país sin tener indicadores asociados)**

Estas asignaciones de cardinalidades y participación indican las reglas y restricciones de las relaciones entre las entidades, lo que ayuda a definir la estructura y la integridad del modelo de datos.

Disponibilidad de los datos en la nube

Pipelines para alimentar el DW

En el contexto del proyecto, los pipelines desempeñan un papel crucial en la alimentación y actualización del Data Lake (WL) y Data Warehouse (DW). Los pipelines son flujos de trabajo automatizados que permiten el procesamiento, transformación y carga de datos desde diversas fuentes hasta el DW, asegurando que los datos estén disponibles y actualizados para su posterior análisis.



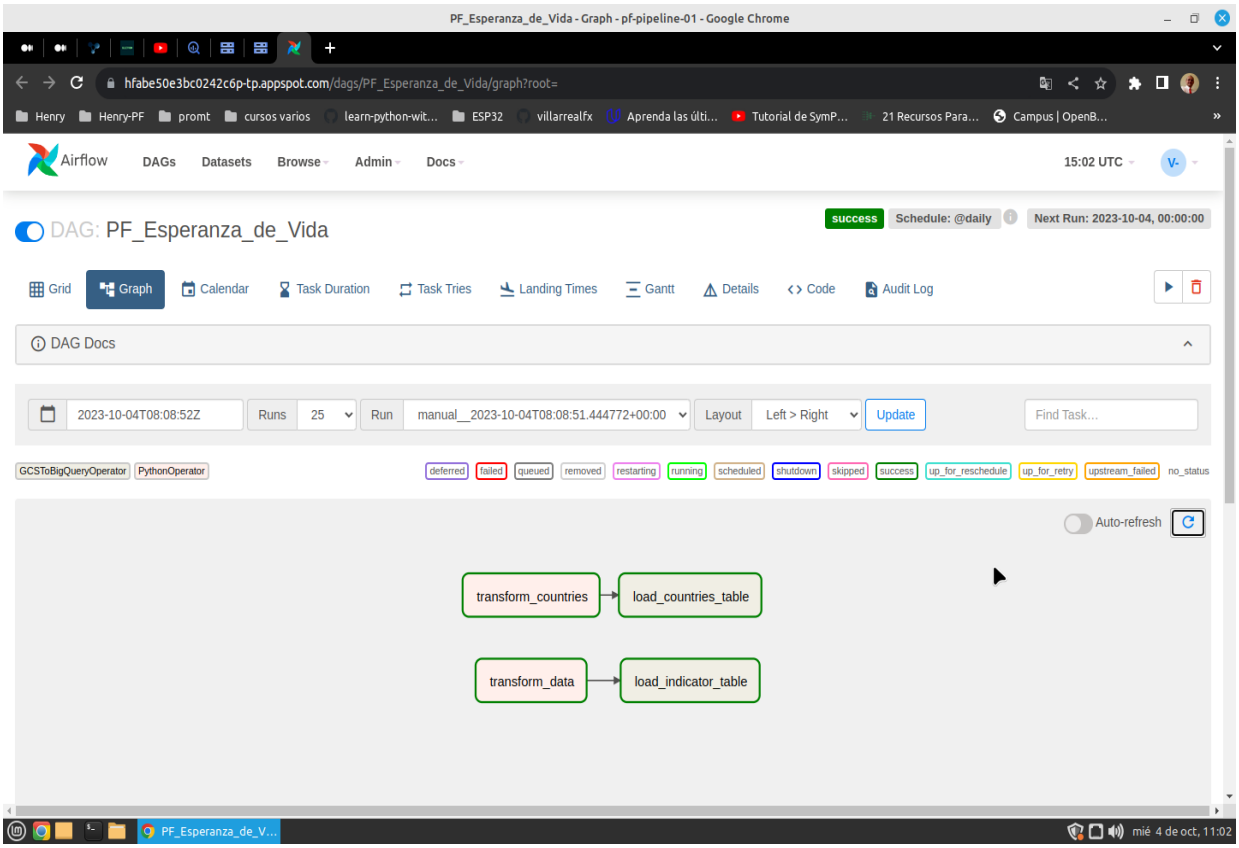
Extracción de datos:

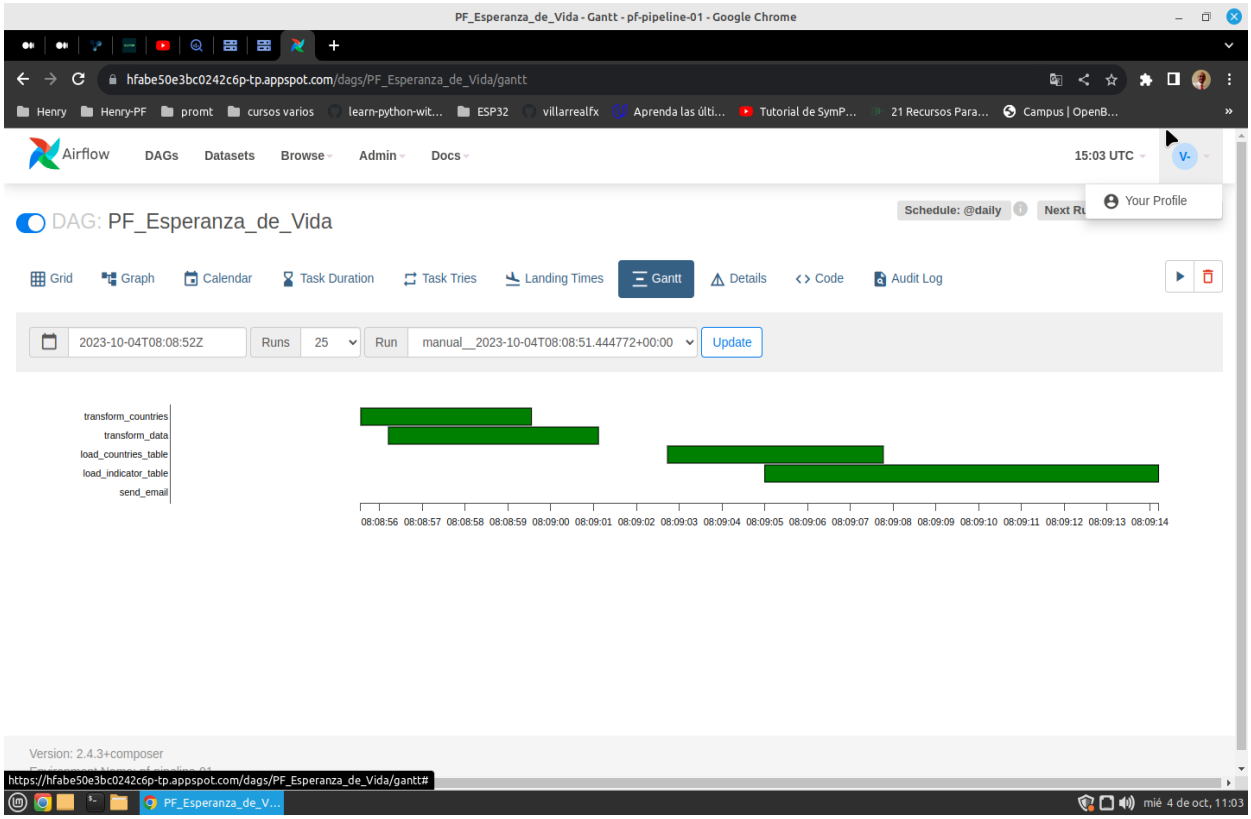
El primer paso en el pipeline es la extracción de datos desde diversas fuentes de información relevantes para el proyecto. Estas fuentes incluyen bases de datos externas, principalmente las bases de datos alojadas y administradas por el Banco Mundial a través de su API, las cuales suministran la información en formato JSON. Se utilizan herramientas o librerías específicas, como wbgapi y pandas, para acceder a los datos y obtenerlos en un formato estructurado pero en crud, de esta manera poder almacenarlos en el Data Lake en nuestro caso es un Bucket de Cloud Storage.

Seguidamente y de manera automática por medio de del servicio de Cloud Composer con airflow se realiza transformación de los datos requeridos, se adecúan según las necesidades de los servicios que los utilizaran (Inteligencia de Negocios y Modelado de Machine Learning) se desnormalizan y se almacenan en una instania de Cloud BigQuery para que sean consumidos posteriormente. Adicionalmente en el proceso, se crea una copia redundante de los datos actualizados y se resguarda como copia de seguridad en el Data Lake, ésta copia

permite restaurar el data werehouse de manera inmediata en caso de que suceda alguna eventualidad.

Para nuestro caso el sistema de pipeline automático tiene una frecuencia de búsqueda diaria de los datos en sus fuentes de origen, la transformación y carga en el Data Werehouse.



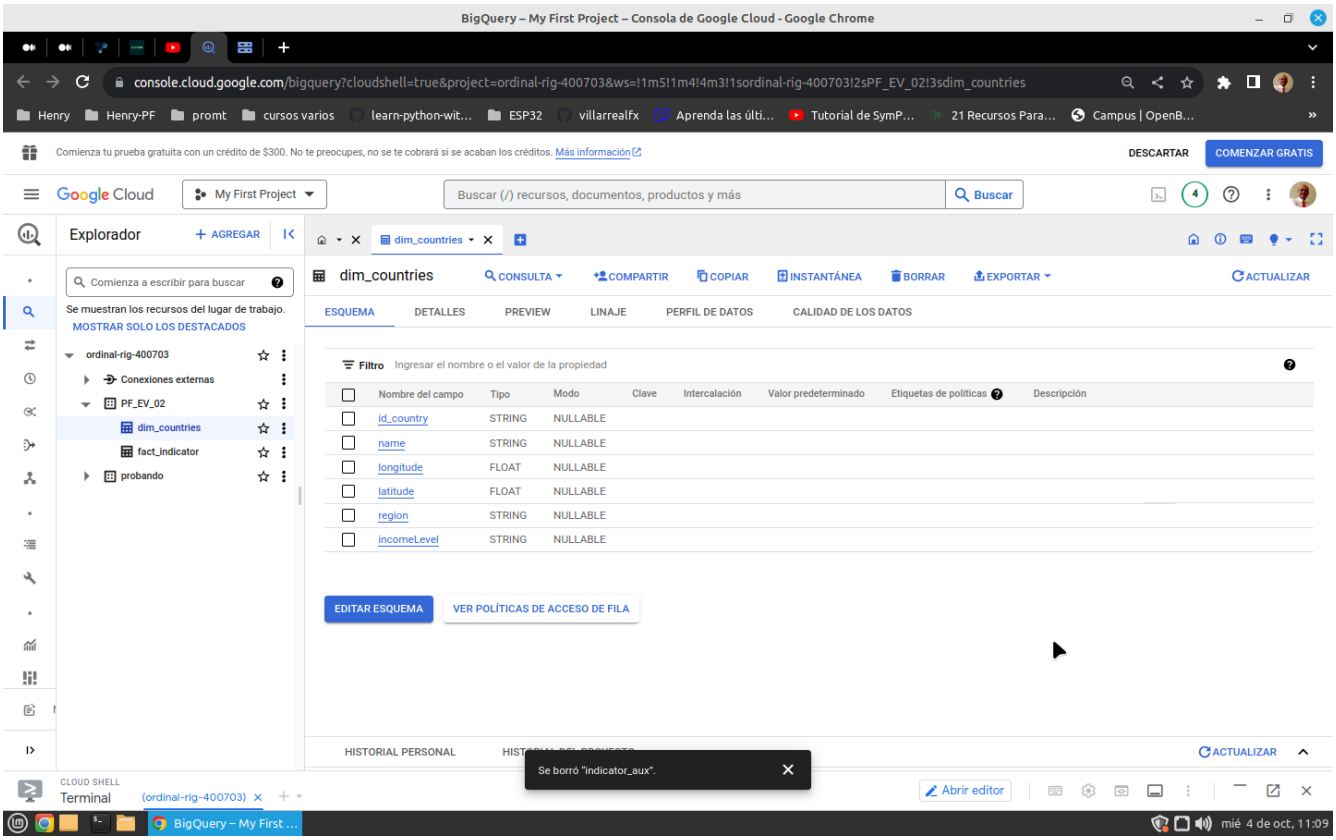


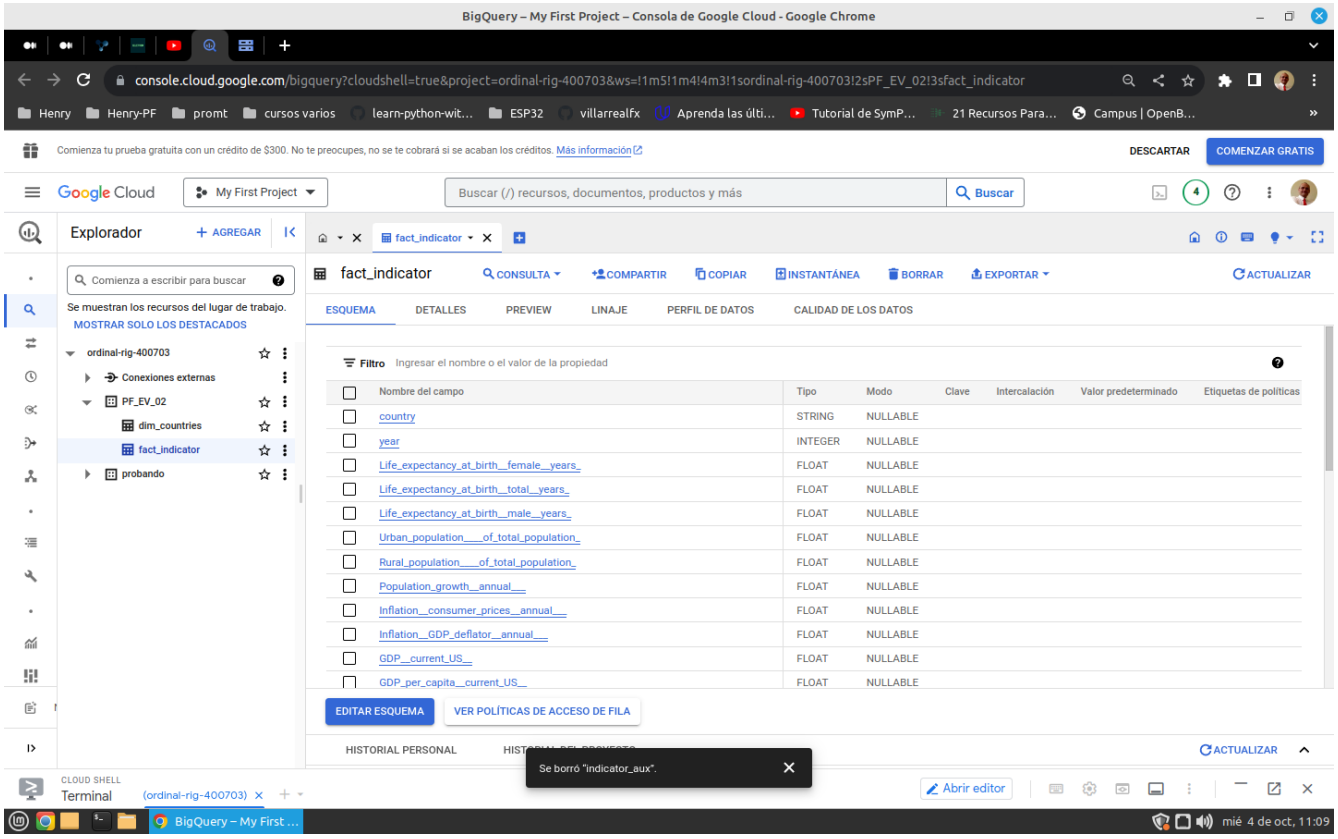
Transformación de datos:

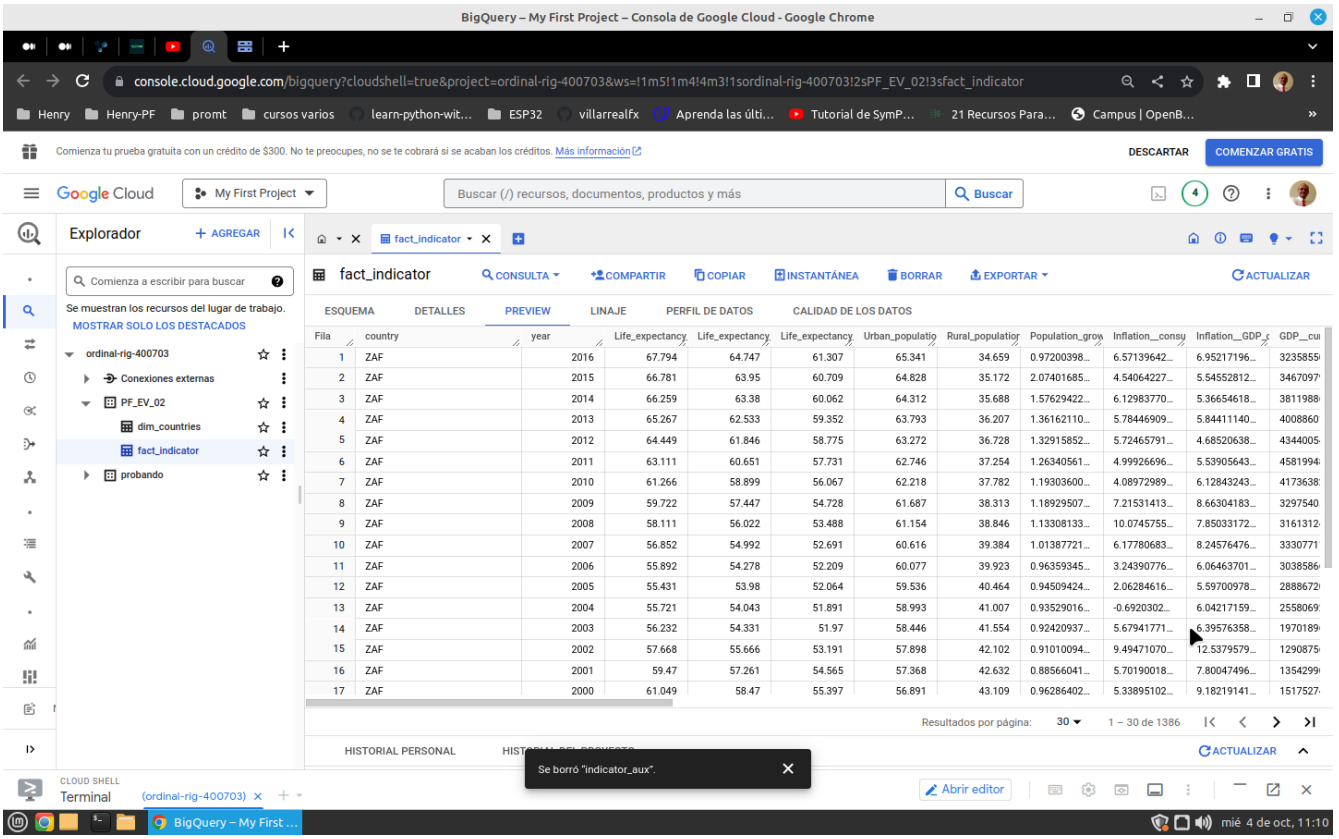
Una vez que los datos se han extraído, se realiza la etapa de transformación. Durante esta fase, se aplican diversas operaciones para limpiar y preparar los datos antes de cargarlos en el DW. Esto puede incluir la eliminación de registros duplicados, la corrección de valores inconsistentes o nulos, la normalización de datos, la agregación de información, entre otros. Las librerías como pandas y NumPy son útiles para llevar a cabo estas transformaciones.

Disponibilización de datos:

Después de la transformación, los datos se cargan en el DW que es implementado utilizando tecnologías como Cloud BigQuery. Esta plataforma de almacenamiento y procesamiento distribuido permiten la carga eficiente de grandes volúmenes de datos y brindan un entorno analítico de alto rendimiento.




BigQuery – My First Project – Consola de Google Cloud - Google Chrome
console.cloud.google.com/bigquery?cloudshell=true&project=ordinal-rig-400703&ws=!1m5!1m4!4m3!1sordinal-rig-400703!2sPF_EV_02!3sfact_indicator
Comienza tu prueba gratuita con un crédito de $300. No te preocupes, no se te cobrará si se acaban los créditos. Más información
DESCARTAR
COMENZAR GRATIS
Google Cloud
My First Project
Buscar (/) recursos, documentos, productos y más
Buscar
Explorador
+ AGREGAR
Comienza a escribir para buscar
Se muestran los recursos del lugar de trabajo.
MOSTRAR SOLO LOS DESTACADOS
ordinal-rig-400703
Conexiones externas
PF_EV_02
dim_countries
fact_indicator
probando
fact_indicator
CONSULTA
COMPARTIR
COPIAR
INSTANTÁNEA
BORRAR
EXPORTAR
ACTUALIZAR
ESQUEMA
DETALLES
PREVIEW
LINAJE
PERFIL DE DATOS
CALIDAD DE LOS DATOS
Filtro
Ingresar el nombre o el valor de la propiedad
Nombre del campo
Tipo
Modo
Clave
Intercalación
Valor predeterminado
Etiquetas de políticas
country STRING NULLABLE
year INTEGER NULLABLE
Life_expectancy_at_birth__female__years_ FLOAT NULLABLE
Life_expectancy_at_birth__total__years_ FLOAT NULLABLE
Life_expectancy_at_birth__male__years_ FLOAT NULLABLE
Urban_population____of_total_population_ FLOAT NULLABLE
Rural_population____of_total_population_ FLOAT NULLABLE
Population_growth__annual___ FLOAT NULLABLE
Inflation__consumer_prices__annual___ FLOAT NULLABLE
Inflation__GDP_deflator__annual___ FLOAT NULLABLE
GDP__current_US__ FLOAT NULLABLE
GDP_per_capita__current_US__ FLOAT NULLABLE
EDITAR ESQUEMA
VER POLÍTICAS DE ACCESO DE FILA
HISTORIAL PERSONAL
Se borró "indicator_aux".
CLOUD SHELL
Terminal
(ordinal-rig-400703)
Abrir editor
BigQuery – My First ...
mié 4 de oct, 11:09


BigQuery – My First Project – Consola de Google Cloud - Google Chrome
console.cloud.google.com/bigquery?cloudshell=true&project=ordinal-rig-400703&ws=!1m5!1m4!4m3!1sordinal-rig-400703!2sPF_EV_02!3sfact_indicator
Comienza tu prueba gratuita con un crédito de $300. No te preocupes, no se te cobrará si se acaban los créditos. Más información
DESCARTAR
COMENZAR GRATIS
Google Cloud
My First Project
Buscar (/) recursos, documentos, productos y más
Buscar
Explorador
+ AGREGAR
Comienza a escribir para buscar
Se muestran los recursos del lugar de trabajo.
MOSTRAR SOLO LOS DESTACADOS
ordinal-rig-400703
Conexiones externas
PF_EV_02
dim_countries
fact_indicator
probando
fact_indicator
CONSULTA
COMPARTIR
COPIAR
INSTANTÁNEA
BORRAR
EXPORTAR
ACTUALIZAR
ESQUEMA
DETALLES
PREVIEW
LINAJE
PERFIL DE DATOS
CALIDAD DE LOS DATOS
Fila
country
year
Life_expectancy
Life_expectancy
Life_expectancy
Urban_populatio
Rural_populatior
Population_grow
Inflation__consu
Inflation__GDP_c
GDP__cu
1 ZAF 2016 67.794 64.747 61.307 65.341 34.659 0.97200398... 6.57139642... 9.95217196... 3235855
2 ZAF 2015 66.781 63.95 60.709 64.828 35.172 2.07401685... 4.54064227... 5.54552812... 3467097
3 ZAF 2014 66.259 63.38 60.062 64.312 35.688 1.57629422... 6.12983770... 5.36654618... 3811988
4 ZAF 2013 65.267 62.533 59.352 63.793 36.207 1.36162110... 5.78446909... 5.84411140... 4008860
5 ZAF 2012 64.449 61.846 58.775 63.272 36.728 1.32915852... 5.72465791... 4.68520638... 4344005
6 ZAF 2011 63.111 60.651 57.731 62.746 37.254 1.26340561... 4.99926696... 5.53905643... 4581994
7 ZAF 2010 61.266 58.899 56.067 62.218 37.782 1.19303600... 4.08972989... 6.12843243... 4173638
8 ZAF 2009 59.722 57.447 54.728 61.687 38.313 1.18929507... 7.21531413... 8.66304183... 3297540
9 ZAF 2008 58.111 56.022 53.488 61.154 38.846 1.13308133... 10.0745755... 7.85033172... 3161312
10 ZAF 2007 56.852 54.992 52.691 60.616 39.384 1.01387721... 6.17780683... 8.24576476... 3330771
11 ZAF 2006 55.892 54.278 52.209 60.077 39.923 0.96359345... 3.24390776... 6.06463701... 3038586
12 ZAF 2005 55.431 53.98 52.064 59.536 40.464 0.94509424... 2.06284616... 5.59700978... 2888672
13 ZAF 2004 55.721 54.043 51.891 58.993 41.007 0.93529016... -0.6920302... 6.04217159... 2558069
14 ZAF 2003 56.232 54.331 51.97 58.446 41.554 0.92420937... 5.67941771... 6.39576358... 1970189
15 ZAF 2002 57.668 55.666 53.191 57.898 42.102 0.91010094... 9.49471070... 12.5379579... 1290875
16 ZAF 2001 59.47 57.261 54.565 57.368 42.632 0.88566041... 5.70190018... 7.80047496... 1354299
17 ZAF 2000 61.049 58.47 55.397 56.891 43.109 0.96286402... 5.33895102... 9.18219141... 1517527
Resultados por página: 30
1 – 30 de 1386
HISTORIAL PERSONAL
ACTUALIZAR
Se borró "indicator_aux".
CLOUD SHELL
Terminal
(ordinal-rig-400703)
Abrir editor
BigQuery – My First ...
mié 4 de oct, 11:10

Workflow detallando tecnologías