

Perspectivas Integradas: Análisis Prospectivo de la Mortalidad por Infecciones Respiratorias en Cinco Potencias Globales (2020-2025)

Marilyn Rojas, Adriana Rojas, Paula Galindo, Ana Maria Cruz

Maestria Analitica de Datos

Universidad Central

Curso de Bases de Datos

Bogotá, Colombia

mrojasa7@ucentral.edu.co, arojasa4@ucentral.edu.co, pgalindop@ucentral.edu.co

acruz3@ucentral.edu.co

November 23, 2023

Contents

1	Introducción	3
2	Características del proyecto de investigación que hace uso de Bases de Datos	3
2.1	Título del proyecto de investigación	3
2.2	Objetivo general	3
2.2.1	Objetivos específicos	4
2.3	Alcance	4
2.4	Pregunta de investigación	4
2.5	Hipótesis	4
3	Reflexiones sobre el origen de datos e información	5
3.1	¿Cuál es el origen de los datos e información?	5
3.2	¿Cuáles son las consideraciones legales o éticas del uso de la información?	5
3.3	¿Cuáles son los retos de la información y los datos que utilizará en la base de datos en términos de la calidad y la consolidación?	6
3.4	¿Qué espera de la utilización de un sistema de Bases de Datos para su proyecto?	6
4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)	6
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto	6
4.2	Diagrama modelo de datos	7
4.3	Imágenes de la Base de Datos	8
4.4	Código SQL - lenguaje de definición de datos (DDL)	9
4.5	Código SQL - Manipulación de datos (DML)	12
4.6	Código SQL + Resultados: Vistas	14
4.7	Código SQL + Resultados: Triggers	15
4.8	Código SQL + Resultados: Funciones	16
4.9	Código SQL + Resultados: procedimientos almacenados	17

5	Bases de Datos No-SQL	18
5.1	Diagrama Bases de Datos No-SQL	18
5.2	SMBD utilizado para la Base de Datos No-SQL	19
6	Aplicación de ETL (Extract, Transform, Load)	21
6.1	Ejemplo de aplicación de ETL	21
7	Proximos pasos	29
8	Lecciones aprendidas	31
9	Bibliografía	33

1 Introducción

En el panorama global actual, la comprensión profunda de los patrones de salud y las variables que influyen en la mortalidad es esencial para abordar desafíos críticos. El proyecto "Perspectivas Integradas" surge como un estudio de la mortalidad por infecciones respiratorias en cinco potencias globales: Estados Unidos, China, Japón, Alemania y Brasil. Con un enfoque abarcador desde 1990 hasta 2025, esta investigación se sumerge en datos reales recopilados de diversas fuentes confiables.

Guiado por la premisa de que los datos son cruciales para la comprensión, el proyecto emplea técnicas avanzadas de gestión y análisis de datos, aprovechando la potencia de los Sistemas Manejadores de Bases de Datos (SMBD) tanto relacionales como No-SQL. La combinación de la estructura robusta de las bases de datos tradicionales y la flexibilidad de las bases de datos No-SQL, especialmente MongoDB, proporciona el marco ideal para explorar patrones, tendencias y relaciones en la mortalidad por infecciones respiratorias.

Desde la imputación de datos faltantes hasta la construcción de modelos de pronóstico, el proyecto se embarca en un viaje analítico que abarca variables económicas, ambientales y socio-demográficas. Al hacerlo, busca responder preguntas cruciales sobre las tendencias históricas y los posibles escenarios futuros de la mortalidad, explorando la compleja interacción entre factores económicos, ambientales y de salud pública.

La elección de tecnologías líderes, como Oracle Cloud y MongoDB, para la gestión de bases de datos refleja el compromiso del proyecto con la eficiencia, la escalabilidad y la seguridad en el tratamiento de datos a gran escala. A través de este proyecto, se aspira no solo a comprender el pasado y proyectar el futuro de la mortalidad por infecciones respiratorias, sino también a sentar las bases para enfoques informados y estratégicos en la promoción de la salud a nivel global.

2 Características del proyecto de investigación que hace uso de Bases de Datos

- **Uso de datos reales:** Este proyecto se basa en datos reales recopilados de diversas fuentes confiables, proporcionando una sólida base para la investigación y análisis de la mortalidad por infecciones respiratorias en cinco potencias globales.
- **Gestión eficiente de datos:** Implica la gestión eficiente de un conjunto significativo de datos, aprovechando sistemas avanzados de analítica para procesar la información de manera efectiva y garantizar su integridad. La focalización en cinco países específicos permite una gestión más precisa y detallada de los datos.
- **Enfoque Analítico:** Se destaca el enfoque analítico del proyecto, utilizando métodos estadísticos y técnicas de modelado de bases de datos para identificar patrones, tendencias y relaciones en los datos específicos de Estados Unidos, China, Japón, Alemania y Brasil. Además, se hace hincapié en la importancia de la identificación y manejo de datos faltantes, normalización y verificación de fuentes para asegurar la calidad de los resultados en este contexto específico.

2.1 Título del proyecto de investigación

Perspectivas Integradas: Análisis Prospectivo de la Mortalidad por Infecciones Respiratorias en Cinco Potencias Globales (2020-2025).

2.2 Objetivo general

Analizar el pronóstico de mortalidad por infecciones respiratorias en cinco potencias globales (Estados Unidos, China, Japón, Alemania y Brasil) para el periodo de 2020 a 2025, integrando datos históricos

de 1990 a 2019 y considerando variables económicas y ambientales, con el fin de comprender las tendencias, identificar factores determinantes y proyectar posibles escenarios futuros.

2.2.1 Objetivos específicos

- Realizar la imputación de datos faltantes en la base de datos mediante técnicas estadísticas apropiadas, asegurando la integridad y calidad de los datos.
- Realizar un análisis exploratorio de datos para comprender la distribución y variabilidad de los indicadores relevantes en los conjuntos de datos específicos de cada país.
- Construir un modelo de pronóstico para las tasas de mortalidad por infecciones respiratorias en los cinco países seleccionados, basado en los datos imputados y el análisis exploratorio previos.

2.3 Alcance

Este proyecto tiene como objetivo realizar una evaluación de la mortalidad por infecciones respiratorias en cinco potencias globales a lo largo del extenso período que abarca desde 1990 hasta 2025. La investigación se centrará en la recopilación, análisis y correlación de datos históricos y contemporáneos de cinco países específicos: Estados Unidos, Japón, China, Alemania y Brasil. Estos países fueron seleccionados con el propósito de determinar el pronóstico de muertes, considerando tanto el número de habitantes (es decir, aquellos con mayor población mundial) como el Producto Interno Bruto (PIB) de cada uno (priorizando aquellos con mayores índices de PIB).

Para llevar a cabo este análisis, se tendrán en cuenta diversas variables clave, tales como el Producto Interno Bruto (PIB), emisiones de CO₂, población, Índice de Desarrollo Humano (IDH) y esperanza de vida. Estas variables se consideran fundamentales para comprender y contextualizar la relación entre la mortalidad por infecciones respiratorias y los indicadores socioeconómicos y ambientales de los países seleccionados.

El análisis de tendencias históricas permitirá identificar patrones significativos, mientras que la proyección hacia el futuro se enfocará en el periodo de 2020 a 2025. Se explorará la influencia de las variables económicas y ambientales en la mortalidad por infecciones respiratorias, buscando comprender las interrelaciones entre estos factores. Sin embargo, el alcance no incluirá un análisis detallado de factores específicos de salud pública o intervenciones médicas.

La conclusión del proyecto ofrecerá una visión integral de las tendencias pasadas y perspectivas futuras de la mortalidad por infecciones respiratorias, considerando su relación con aspectos económicos y ambientales a escala global.

2.4 Pregunta de investigación

¿Cuáles son las tendencias históricas de mortalidad por infecciones respiratorias en los países seleccionados (Estados Unidos, Japón, China, Alemania y Brasil) entre 1990 y 2019, y cómo se proyecta el pronóstico de mortalidad para el periodo de 2020 a 2025, considerando la influencia de variables económicas (PIB) y ambientales (emisiones de CO₂), así como su interacción con indicadores sociodemográficos (población, IDH, esperanza de vida)?

2.5 Hipótesis

Se plantea que existe una asociación significativa entre las variables económicas, ambientales y sociodemográficas consideradas y la mortalidad por infecciones respiratorias en los cinco países seleccionados (Estados Unidos, Japón, China, Alemania y Brasil). Se anticipa que estas relaciones han experimentado cambios a lo largo del tiempo, influyendo en el pronóstico de mortalidad para el periodo de 2020 a 2025. En términos específicos, se espera que las tendencias históricas de mortalidad por infecciones

respiratorias revelan variaciones en consonancia con evoluciones socioeconómicas, ambientales y demográficas.

Además, se hipotetiza una correlación inversa entre el Producto Interno Bruto (PIB) de los países y la mortalidad, sugiriendo que un mayor desarrollo económico podría asociarse con tasas inferiores de mortalidad. Asimismo, se anticipa una correlación positiva entre las emisiones de CO₂ y la mortalidad por infecciones respiratorias, indicando una posible contribución de la contaminación atmosférica al aumento de la mortalidad. Estas hipótesis constituyen la base teórica sobre la cual se fundamenta el análisis y la interpretación de los datos recopilados durante el transcurso de la investigación.

3 Reflexiones sobre el origen de datos e información

La obtención de datos socioeconómicos, ambientales y de mortalidad fundamentales para construir la base de datos se realiza mediante fuentes específicas y confiables, tales como el Banco Mundial, la Organización Mundial de la Salud, así como plataformas reconocidas como Kaggle y populationpyramid. Se presta especial atención al origen de estos datos, reconociendo la importancia crucial de la calidad para garantizar la integridad y veracidad de la información. En la evaluación de la calidad de los datos, se identifican campos específicos en variables como población, índice de desarrollo humano e índices ambientales, que requieren normalización o eliminación mediante técnicas estadísticas previamente aplicadas. Se realiza una evaluación de la consistencia de los datos y la metodología de recopilación, considerando factores como encuestas, registros de países e informes globales. Además, se verifica la frecuencia de actualización de los datos y se asegura que mantengan un periodo de referencia coherente con la información actual. Finalmente, es crucial destacar que la información sensible está bajo control y es de carácter público a nivel mundial.

3.1 ¿Cuál es el origen de los datos e información?

El origen de los datos e información en la base de datos se fundamenta en una búsqueda focalizada principalmente en la Organización Mundial de la Salud (OMS). A partir del año 1990, la OMS ha llevado a cabo la "Carga Global de Enfermedad", un informe que mide la pérdida de salud en cada país y facilita la comparación directa de la carga de mortalidad por diferentes enfermedades entre naciones. Estos informes, publicados regularmente en diversos institutos de Medición y Evaluación de la Salud (IHME) desde 1993, son ampliamente reconocidos como una fuente confiable de datos en este ámbito.

Adicionalmente, la mortalidad en cada país desde 1990-2019 se obtuvo del sitio web Kaggle (<https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world>), creado a partir de Our World in Data y disponible bajo la licencia Creative Commons BY. Además, se incluyen enlaces adicionales que respaldan la integración de otras variables relevantes para el estudio, como la población de cada país de 1990-2019 (<https://www.populationpyramid.net/es/poblacion-por-pais/1990/>) y diversos indicadores obtenidos de <https://datos.macroe.expansion.com>.

3.2 ¿Cuáles son las consideraciones legales o éticas del uso de la información?

Las bases de datos empleadas en esta investigación son de acceso público, disponibles de manera libre y gratuita, y, por ende, no están sujetas a restricciones de privacidad y confidencialidad. Se asegura el cumplimiento de las leyes de protección de datos y regulaciones de privacidad tanto en Colombia como en los países de origen de los datos, los cuales se derivan de fuentes internacionales.

Es importante destacar que los datos utilizados son de origen internacional y provienen de fuentes oficiales que siguen un proceso adecuado para su uso, clasificándose como "datos abiertos". Este término implica que están disponibles para el público sin implicaciones legales.

La investigación se lleva a cabo de manera ética y responsable, comprometida con la integridad de los datos y la presentación honesta de los resultados. Además, se respetan plenamente los derechos

de propiedad intelectual de las fuentes de datos. Este ajuste busca mejorar la fluidez y claridad del texto. Asegúrate de que refleje con precisión las prácticas éticas y legales específicas aplicadas en tu investigación.

3.3 ¿Cuáles son los retos de la información y los datos que utilizará en la base de datos en términos de la calidad y la consolidación?

Dentro del proyecto de investigación, se enfrentaron desafíos en la calidad y consolidación de los datos de la base. En primer lugar, se encontraron datos faltantes, los cuales requerían el uso de técnicas estadísticas de imputación para garantizar la integridad del análisis. Además, al integrar información de diversas fuentes, surgió la complejidad de normalizar y verificar los datos para mantener uniformidad. La evaluación de la confiabilidad de las fuentes, que incluyó la revisión de metodologías y la representatividad de las muestras, también se consideró un aspecto crítico. En todos estos procesos, se priorizó el cumplimiento de las regulaciones de protección de datos y se garantizó la ética y la integridad en la investigación.

3.4 ¿Qué espera de la utilización de un sistema de Bases de Datos para su proyecto?

La implementación de un sistema de bases de datos en nuestro proyecto de investigación es esencial para garantizar una gestión eficaz de los datos y alcanzar los objetivos propuestos. Nuestra expectativa principal es que el sistema proporcione una solución integral que aborde los desafíos relacionados con la calidad y consolidación de la información. En primer lugar, buscamos que el sistema asegure la integridad de los datos a lo largo de todo el proyecto, implementando controles de calidad para mantener la precisión y coherencia de la información almacenada.

Además, esperamos que el sistema facilite un acceso rápido y preciso a los datos, lo cual es crucial para llevar a cabo análisis avanzados y obtener resultados de manera oportuna. La seguridad y privacidad de los datos son aspectos críticos, y confiamos en que el sistema cumpla con los estándares de seguridad necesarios y las regulaciones de privacidad para proteger nuestra información de manera robusta.

Aspiramos a que el sistema facilite la realización de análisis avanzados, incluyendo modelos de series temporales y técnicas estadísticas, permitiéndonos explorar patrones y tendencias en los datos de manera efectiva. Además, la escalabilidad es un requisito esencial para adaptarse a nuestras necesidades en evolución a medida que el proyecto avanza, así como para incorporar nuevos datos de manera eficiente en el futuro.

4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)

4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto

Para el desarrollo del proyecto se usa como SMBD Oracle Cloud que es una plataforma de nube ofrecida por Oracle Corporation que proporciona una amplia variedad de servicios en la nube, incluido un sistema de gestión de bases de datos llamado Oracle Database Cloud Service. Algunas de sus características más importantes son:

- *Escalabilidad:* Oracle Cloud permite escalar verticalmente y horizontalmente la capacidad de tu base de datos según las necesidades de tu aplicación. Puedes ajustar los recursos de CPU, memoria y almacenamiento de manera flexible.

- *Rendimiento:* Oracle Database en la nube ofrece un rendimiento de base de datos de alto nivel con tecnologías avanzadas de optimización y ajuste de rendimiento.
- *Disponibilidad:* Oracle Cloud ofrece opciones de alta disponibilidad, incluida la replicación de datos y la tolerancia a fallos para garantizar que tus aplicaciones estén disponibles de manera continua.
- *Seguridad:* Oracle Database Cloud proporciona una amplia gama de medidas de seguridad, incluida la autenticación, el cifrado de datos, el control de acceso y la auditoría para proteger tus datos.
- *Gestión de bases de datos:* Oracle Cloud ofrece herramientas de gestión de bases de datos que facilitan la administración, la monitorización y el ajuste de tus bases de datos. También puedes automatizar tareas de mantenimiento y copias de seguridad.
- *Copias de seguridad y recuperación:* Oracle Cloud ofrece opciones de copia de seguridad automatizadas y recuperación de datos para garantizar la integridad y la disponibilidad de tus datos.
- *Integración:* Oracle Cloud Database se integra fácilmente con otras soluciones de Oracle, como Oracle Cloud Applications, para permitir una integración fluida de datos y procesos comerciales.
- *Soporte para múltiples bases de datos:* Además de Oracle Database, Oracle Cloud admite otras bases de datos populares, como MySQL y PostgreSQL, lo que te permite elegir la mejor opción para tu aplicación.
- *Movilidad de datos:* Oracle Cloud Database admite la migración de datos desde y hacia entornos locales, lo que facilita la transición a la nube o la implementación híbrida.
- *Monitoreo y diagnóstico:* Oracle Cloud ofrece herramientas avanzadas de monitoreo y diagnóstico que te permiten supervisar el rendimiento de tu base de datos y solucionar problemas de manera eficiente.

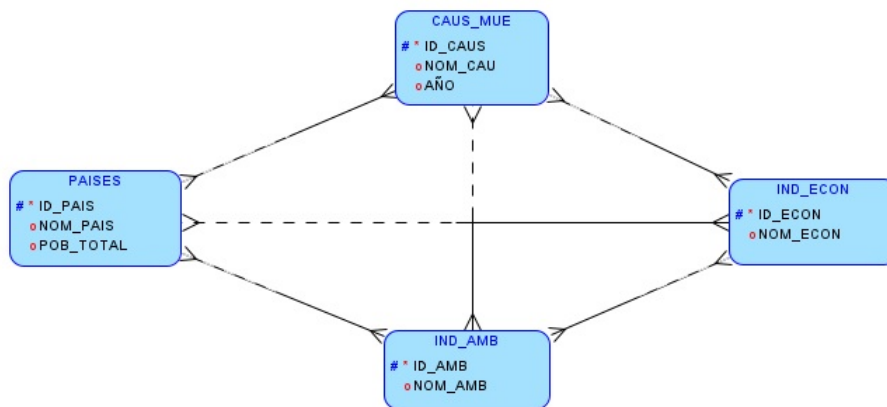
Por las características antes mencionadas se decide manejar la base de datos en este sistema.

4.2 Diagrama modelo de datos

En el desarrollo de sistemas de información y bases de datos, la representación estructural es esencial para comprender la organización de los datos y garantizar una gestión eficiente de la información. En este contexto, hemos creado y diseñado dos modelos fundamentales: el Modelo Lógico y el Modelo Entidad - Relación (ER).

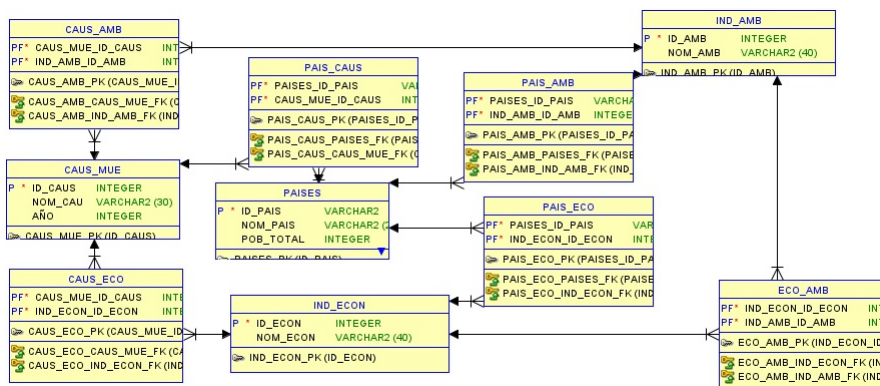
• Modelo Lógico

Este modelo posibilita la definición de la estructura de la base de datos de manera abstracta, poniendo énfasis en la lógica y organización de los datos, sin necesidad de preocuparse por los detalles de implementación.



• Modelo Entidad - Relación

Este modelo proporciona una representación visual de las entidades, atributos y relaciones en la base de datos, ayudándonos a comprender la interconexión de los datos de una manera clara y efectiva.



En este proyecto las bases de datos tienen una relación de “muchos a muchos” entre las tablas, lo que se refiere a un tipo de relación en la que múltiples registros en este caso; muchos países están relacionados con múltiples causas de muertes, múltiples indicadores económicos e indicadores ambientales.

4.3 Imágenes de la Base de Datos

Tabla Países

```
-- verificación contenido creado
select * from países
```

ID_PAIS	NOM_PAIS	POB_TOTAL
1	Afganistán	10694796
2	Albania	3295066
3	Algeria	25518074
4	Samoa	168186
5	Andorra	54507

Tabla Causas de muerte

```
-- verificación contenido creado
select * from caus_mue
```

ID_CAUS	NOM_CAU	AÑO
0	Meningitis	1990
2	Alzheimer	1991
3	Parkinson	1992
4	Malaria	1993
1	VIH/SIDA	1994

Tabla Indicadores Económicos

```
-- verificación contenido creado
select * from ind_econ
```

ID_ECON	NOM_ECON
11	PIB
12	IDH
15	Esperanza de vida
13	Esperanza de vida - Mujeres
14	Esperanza de vida - Hombres

Tabla Indicadores Ambientales

```
-- verificación contenido creado
select * from ind_amb
```

ID_AMB	NOM_AMB
1	Emisiones Totales CO2
2	Emisiones CO2 por 1000USD del PIB
3	Emisiones CO2 toneladas per capita

4.4 Código SQL - lenguaje de definición de datos (DDL)

Creación de tablas y llaves

1. Creación de tabla Causas de muerte

```
CREATE TABLE caus_mue (
    id_caus INTEGER NOT NULL,
    nom_cau VARCHAR2(30),
    año     INTEGER
);
```

```
ALTER TABLE caus_mue ADD CONSTRAINT caus_mue_pk PRIMARY KEY ( id_caus );
```

2. Creación de tabla Indicadores Ambientales

```
CREATE TABLE ind_amb (
    id_amb INTEGER NOT NULL,
    nom_amb VARCHAR2(40)
);
```

```
ALTER TABLE ind_amb ADD CONSTRAINT ind_amb_pk PRIMARY KEY ( id_amb );
```

3. Creación de tabla Indicadores económicos

```
CREATE TABLE ind_econ (  
    id_econ INTEGER NOT NULL,  
    nom_econ VARCHAR2(40)  
);
```

```
ALTER TABLE ind_econ ADD CONSTRAINT ind_econ_pk PRIMARY KEY ( id_econ );
```

4. Creación de tabla Países

```
CREATE TABLE paises (  
    id_pais INTEGER NOT NULL,  
    nom_pais VARCHAR2(20),  
    pob_total INTEGER  
);
```

```
ALTER TABLE paises ADD CONSTRAINT paises_pk PRIMARY KEY ( id_pais );
```

Creación de tablas entidad - relaciones

5. Creación de tabla Causas Muerte - Indicador Ambiental

```
CREATE TABLE caus_amb (  
    caus_mue_id_caus INTEGER NOT NULL,  
    ind_amb_id_amb INTEGER NOT NULL  
);
```

```
ALTER TABLE caus_amb ADD CONSTRAINT caus_amb_pk PRIMARY KEY ( caus_mue_id_caus,  
ind_amb_id_amb );
```

6. Creación de tabla Causas Muerte - Indicadores económicos

```
CREATE TABLE caus_eco (  
    caus_mue_id_caus INTEGER NOT NULL,  
    ind_econ_id_econ INTEGER NOT NULL  
);
```

```
ALTER TABLE caus_eco ADD CONSTRAINT caus_eco_pk PRIMARY KEY ( caus_mue_id_caus,  
ind_econ_id_econ );
```

7. Creación de tabla Indicadores económicos - indicador Ambiental

```
CREATE TABLE eco_amb (  
    ind_econ_id_econ INTEGER NOT NULL,  
    ind_amb_id_amb INTEGER NOT NULL  
);
```

```
ALTER TABLE eco_amb ADD CONSTRAINT eco_amb_pk PRIMARY KEY ( ind_econ_id_econ,  
ind_amb_id_amb );
```

8. Creación de tabla Países - Indicador Ambiental

```
CREATE TABLE pais_amb (
    paises_id_pais INTEGER NOT NULL,
    ind_amb_id_amb INTEGER NOT NULL
);
```

```
ALTER TABLE pais_amb ADD CONSTRAINT pais_amb_pk PRIMARY KEY ( paises_id_pais,
ind_amb_id_amb );
```

9. Creación de tabla Países - Causa Muerte

```
CREATE TABLE pais_caus (
    paises_id_pais INTEGER NOT NULL,
    caus_mue_id_cau INTEGER NOT NULL
);
```

```
ALTER TABLE pais_caus ADD CONSTRAINT pais_caus_pk PRIMARY KEY ( paises_id_pais,
caus_mue_id_cau );
```

10. Creación de tabla Países - Indicadores económicos

```
CREATE TABLE pais_eco (
    paises_id_pais INTEGER NOT NULL,
    ind_econ_id_econ INTEGER NOT NULL
);
```

```
ALTER TABLE pais_eco ADD CONSTRAINT pais_eco_pk PRIMARY KEY ( paises_id_pais,
ind_econ_id_econ );
```

Llaves foraneas

```
ALTER TABLE caus_amb
    ADD CONSTRAINT caus_amb_caus_mue_fk FOREIGN KEY ( caus_mue_id_caus )
    REFERENCES caus_mue ( id_caus );
```

```
ALTER TABLE caus_amb
    ADD CONSTRAINT caus_amb_ind_amb_fk FOREIGN KEY ( ind_amb_id_amb )
    REFERENCES ind_amb ( id_amb );
```

```
ALTER TABLE caus_eco
    ADD CONSTRAINT caus_eco_caus_mue_fk FOREIGN KEY ( caus_mue_id_caus )
    REFERENCES caus_mue ( id_caus );
```

```
ALTER TABLE caus_eco
    ADD CONSTRAINT caus_eco_ind_econ_fk FOREIGN KEY ( ind_econ_id_econ )
    REFERENCES ind_econ ( id_econ );
```

```
ALTER TABLE eco_amb
    ADD CONSTRAINT eco_amb_ind_amb_fk FOREIGN KEY ( ind_amb_id_amb )
    REFERENCES ind_amb ( id_amb );
```

```
ALTER TABLE eco_amb
    ADD CONSTRAINT eco_amb_ind_econ_fk FOREIGN KEY ( ind_econ_id_econ )
    REFERENCES ind_econ ( id_econ );
```

```

ALTER TABLE pais_amb
  ADD CONSTRAINT pais_amb_ind_amb_fk FOREIGN KEY ( ind_amb_id_amb )
    REFERENCES ind_amb ( id_amb );

ALTER TABLE pais_amb
  ADD CONSTRAINT pais_amb_paises_fk FOREIGN KEY ( paises_id_pais )
    REFERENCES paises ( id_pais );

ALTER TABLE pais_caus
  ADD CONSTRAINT pais_caus_caus_mue_fk FOREIGN KEY ( caus_mue_id_caus )
    REFERENCES caus_mue ( id_caus );

ALTER TABLE pais_caus
  ADD CONSTRAINT pais_caus_paises_fk FOREIGN KEY ( paises_id_pais )
    REFERENCES paises ( id_pais );

ALTER TABLE pais_eco
  ADD CONSTRAINT pais_eco_ind_econ_fk FOREIGN KEY ( ind_econ_id_econ )
    REFERENCES ind_econ ( id_econ );

ALTER TABLE pais_eco
  ADD CONSTRAINT pais_eco_paises_fk FOREIGN KEY ( paises_id_pais )
    REFERENCES paises ( id_pais );

```

4.5 Código SQL - Manipulación de datos (DML)

1. Carga de datos en la tabla países

```

insert into paises values('1','Afganistán','10694796');
insert into paises values('2','Albania','3295066');
insert into paises values('3','Algeria','25518074');
insert into paises values('4','Samoa','168186');
insert into paises values('5','Andorra','54507');

```

2. Carga de datos en la tabla Causas Muerte

```

insert into caus_mue values('0','Meningitis','1990');
insert into caus_mue values('2','Alzheimer','1991');
insert into caus_mue values('3','Parkinson','1992');
insert into caus_mue values('4','Malaria','1993');
insert into caus_mue values('1','VIH/SIDA','1994');

```

3. Carga de datos en la tabla Indicadores Económicos

```

insert into ind_econ values('11','PIB');
insert into ind_econ values('12','IDH');
insert into ind_econ values('15','Esperanza de vida');
insert into ind_econ values('13','Esperanza de vida - Mujeres');
insert into ind_econ values('14','Esperanza de vida - Hombres');

```

4. Carga de datos en la tabla Indicadores Ambientales

```
insert into ind_amb values('1','Emisiones Totales CO2');
insert into ind_amb values('2','Emisiones CO2 por 1000USD del PIB');
insert into ind_amb values('3','Emisiones CO2 toneladas per capita');
```

5. Carga de datos en las tablas relaciones

```
-- carga de datos en la tabla caus_amb
insert into caus_amb values('0','1');
insert into caus_amb values('2','2');
insert into caus_amb values('3','3');
insert into caus_amb values('4','1');
insert into caus_amb values('1','2');

-- carga de datos en la tabla caus_eco
insert into caus_eco values('0','11');
insert into caus_eco values('2','12');
insert into caus_eco values('3','13');
insert into caus_eco values('4','14');
insert into caus_eco values('1','15');

-- carga de datos en la tabla eco_amb
insert into eco_amb values('11','1');
insert into eco_amb values('12','2');
insert into eco_amb values('13','3');
insert into eco_amb values('14','1');
insert into eco_amb values('15','2');

-- carga de datos en la tabla pais_amb
insert into pais_amb values('1','1');
insert into pais_amb values('2','2');
insert into pais_amb values('3','3');
insert into pais_amb values('4','1');
insert into pais_amb values('5','2');

-- carga de datos en la tabla pais_caus
insert into pais_caus values('1','0');
insert into pais_caus values('2','2');
insert into pais_caus values('3','3');
insert into pais_caus values('4','4');
insert into pais_caus values('5','1');

-- carga de datos en la tabla pais_eco
insert into pais_eco values('1','11');
insert into pais_eco values('2','12');
insert into pais_eco values('3','13');
insert into pais_eco values('4','14');
insert into pais_eco values('5','15');
```

6. Eliminación de vistas, trigger, procedimiento almacenado, funciones y tablas

```
DROP VIEW VISTA_PAIS_MENINGITIS_1990;
DROP VIEW vista_causas_muerte_pais;
```

```

DROP VIEW vista_ind_econ_pais;
DROP TRIGGER actualizacion_pais;
DROP FUNCTION PROMPOBLA;
DROP PROCEDURE ActualizarPaisAzar;

```

```

DROP TABLE caus_amb
DROP TABLE caus_eco
DROP TABLE caus_mue
DROP TABLE eco_amb
DROP TABLE ind_amb
DROP TABLE ind_econ
DROP TABLE pais_amb
DROP TABLE pais_caus
DROP TABLE pais_eco
DROP TABLE paises

```

4.6 Código SQL + Resultados: Vistas

Las vistas en una base de datos son consultas predefinidas que simplifican el acceso y la gestión de datos. Estas consultas personalizadas permiten a los usuarios ver datos específicos de manera organizada y segura, facilitando la recuperación de información relevante sin la necesidad de escribir consultas complejas.

Creación de vistas

1. Creación de vista para ver nombre del país que tiene causa de muerte Meningitis en el año 1990

```

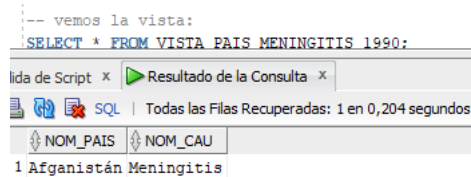
CREATE VIEW vista_pais_meningitis_1990 AS
SELECT
    p.nom_pais AS NOM_PAIS,
    cm.nom_cau AS NOM_CAU
FROM paises p
INNER JOIN
    pais_caus pc ON p.id_pais = pc.paises_id_pais
INNER JOIN
    caus_mue cm ON pc.caus_mue_id_cau = cm.id_caus
WHERE cm.nom_cau = 'Meningitis' AND cm.año = 1990;

```

```

-- Ver vista:
SELECT * FROM VISTA_PAIS_MENINGITIS_1990;

```



2. Creación de vista para ver los paises y nombres de las causas de muerte

```

CREATE VIEW vista_causas_muerte_pais AS

```

```

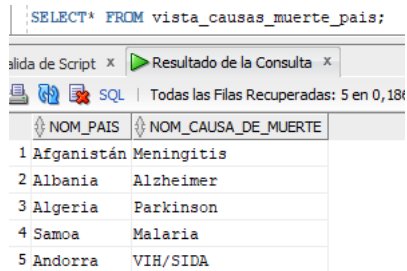
SELECT p.nom_pais AS NOM_PAIS, cm.nom_cau AS NOM_CAUSA_DE_MUERTE
FROM paises p
INNER JOIN pais_caus pc ON p.id_pais = pc.paises_id_pais
INNER JOIN caus_mue cm ON pc.caus_mue_id_cau = cm.id_caus;

```

```

SELECT* FROM vista_causas_muerte_pais;

```



The screenshot shows a SQL query window with the query: `SELECT* FROM vista_causas_muerte_pais;`. The results are displayed in a table with two columns: `NOM_PAIS` and `NOM_CAUSA_DE_MUERTE`. The table contains 5 rows of data.

NOM_PAIS	NOM_CAUSA_DE_MUERTE
1 Afganistán	Meningitis
2 Albania	Alzheimer
3 Algeria	Parkinson
4 Samoa	Malaria
5 Andorra	VIH/SIDA

3. Creación de vista para ver los paises y los indicadores económicos

```

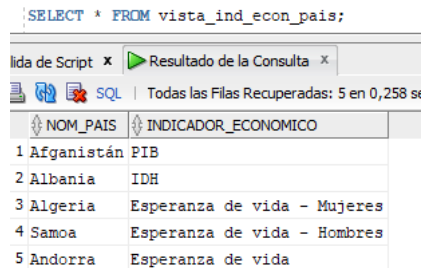
CREATE VIEW vista_ind_econ_pais AS
SELECT p.nom_pais AS NOM_PAIS, ie.nom_econ AS INDICADOR_ECONOMICO
FROM paises p
INNER JOIN pais_eco pe ON p.id_pais = pe.paises_id_pais
INNER JOIN ind_econ ie ON pe.ind_econ_id_econ = ie.id_econ;

```

```

SELECT * FROM vista_ind_econ_pais;

```



The screenshot shows a SQL query window with the query: `SELECT * FROM vista_ind_econ_pais;`. The results are displayed in a table with two columns: `NOM_PAIS` and `INDICADOR_ECONOMICO`. The table contains 5 rows of data.

NOM_PAIS	INDICADOR_ECONOMICO
1 Afganistán	PIB
2 Albania	IDH
3 Algeria	Esperanza de vida - Mujeres
4 Samoa	Esperanza de vida - Hombres
5 Andorra	Esperanza de vida

4.7 Código SQL + Resultados: Triggers

Los triggers son fragmentos de código que se activan automáticamente en respuesta a ciertos eventos en una base de datos, como inserciones, actualizaciones o eliminaciones. Estos eventos desencadenan acciones predefinidas, lo que permite automatizar tareas y aplicar lógica personalizada en la gestión de datos.

Creación de Triggers

1. Creación de Trigger que registre el nombre del país de cada insercción

```

CREATE OR REPLACE TRIGGER actualizacion_pais
BEFORE UPDATE ON paises
FOR EACH ROW
BEGIN
    -- Actualiza la fecha y hora de la última actualización
    :NEW.NOM_PAIS := SYSDATE;

```

```

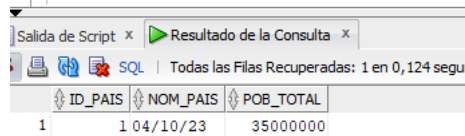
END;

-- Actualización
UPDATE paises
SET pob_total = 35000000
WHERE id_pais = 1; -- Donde 1 es el ID del país que se actualiza

-- actualización y activación del trigger
SELECT * FROM paises WHERE id_pais = 1;

```

```
-- actualización y activación del trigger
SELECT * FROM paises WHERE id_pais = 1;
```



ID_PAIS	NOM_PAIS	POB_TOTAL
1	1 04/10/23	35000000

4.8 Código SQL + Resultados: Funciones

En una base de datos, las funciones son bloques de código reutilizable que realizan tareas específicas. Estas tareas pueden incluir cálculos, manipulación de datos o transformaciones. Las funciones permiten simplificar consultas complejas al encapsular lógica personalizada que puede ser utilizada en múltiples partes de una base de datos.

Creación de funciones

1. Cración de función llamada *PROMEDIO POBLACIÓN*

```

CREATE OR REPLACE FUNCTION PROMPOBLA
RETURN NUMBER
IS
    v_Promedio NUMBER;
BEGIN
    SELECT AVG(POB_TOTAL) INTO v_Promedio
    FROM paises;
    RETURN v_Promedio;
END PROMPOBLA;

-- Ejecutar la función
DECLARE
    resultado NUMBER;
BEGIN
    resultado := PROMPOBLA();
    DBMS_OUTPUT.PUT_LINE('El promedio de la población es: ' || resultado);
END;

--- ver la función
SELECT PROMPOBLA() FROM dual;

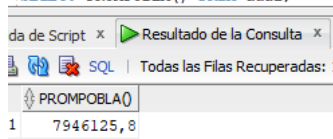
```



```

--- ver la función
SELECT PROMPOBLA() FROM dual;

```



The screenshot shows a database interface with a query window and a results window. The query is `SELECT PROMPOBLA() FROM dual;`. The results window shows a single row with the value `7946125,8`.

4.9 Código SQL + Resultados: procedimientos almacenados

Los procedimientos almacenados son conjuntos de instrucciones SQL que se guardan en la base de datos para ser reutilizados posteriormente. Actúan como funciones o rutinas que realizan operaciones específicas, como inserciones, actualizaciones o consultas complejas.

Creación de Procedimientos almacenados

1. Creación de Procedimientos almacenados para actualizar población del País Andorra

```

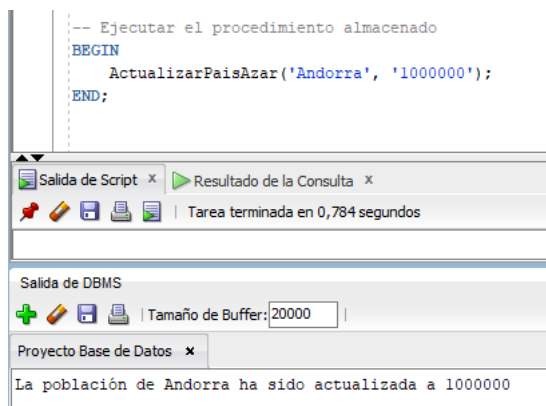
CREATE OR REPLACE PROCEDURE ActualizarPaisAzar(
    nombre_pais_in VARCHAR2,
    nueva_poblacion_in NUMBER
)
IS
BEGIN
    UPDATE paises
    SET pob_total = nueva_poblacion_in
    WHERE nom_pais = nombre_pais_in;
    DBMS_OUTPUT.PUT_LINE('La población de '
    || nombre_pais_in || ' ha sido actualizada a ' || nueva_poblacion_in);
END ActualizarPaisAzar;

```

```

-- Ejecutar el procedimiento almacenado
BEGIN
    ActualizarPaisAzar('Andorra', '1000000');
END;

```

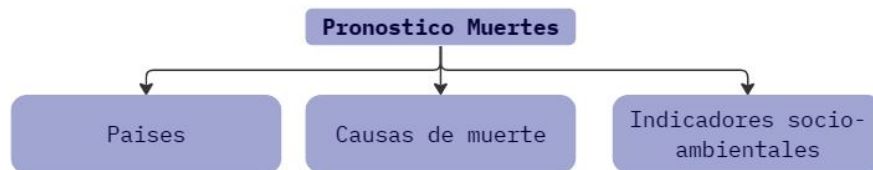


The screenshot shows a database interface with a query window and a results window. The query is `-- Ejecutar el procedimiento almacenado`. The results window shows the output of the stored procedure: `La población de Andorra ha sido actualizada a 1000000`.

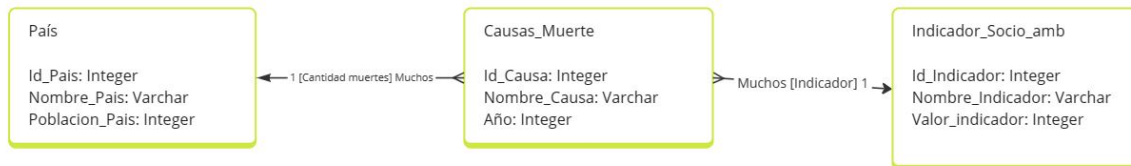
5 Bases de Datos No-SQL

5.1 Diagrama Bases de Datos No-SQL

- Modelo conceptual



- Modelo lógico



- Modelo Físico



País: "Afganistán" Codigo_País: "AFG" Año: numberInt_1990 Población: 10,694,796 Meningitis: numberInt_2159 Alzheimer y otras demencias: numberInt_1116 Parkinson: numberInt_371 . PIB: \$ - IDH: 0,273 Esperanza de vida - Mujeres: 48,4	País: Albania Codigo_País: "ALB" Año: 2000 Población: 3,182,021 Meningitis: numberInt_48 Alzheimer y otras demencias: numberInt_438 Parkinson: numberInt_117 . PIB: \$ 3.480 IDH: 0,677 Esperanza de vida - Mujeres: 78,42	País: Algeria Codigo_País: "DZA" Año: numberInt_1990 Población: 25,518,074 Meningitis: numberInt_838 Alzheimer y otras demencias: numberInt_1337 Parkinson: numberInt_439. . PIB: \$ 62.049 IDH: - Esperanza de vida - Mujeres: 68,52
--	--	---

5.2 SMBD utilizado para la Base de Datos No-SQL

La base de datos para el proyecto "Perspectivas Integradas" se carga en MongoDB, seleccionándolo como el sistema de gestión de bases de datos (SGBD) debido a su naturaleza altamente escalable y flexible en el ámbito de las bases de datos NoSQL. Esta elección se fundamenta en la necesidad de manipular extensos conjuntos de datos que abarcan información detallada de todos los países desde 1990 hasta 2019.

MongoDB destaca por su capacidad para manejar datos complejos y diversos, como estadísticas de población, causas de muerte, indicadores económicos y ambientales. Esta capacidad posibilita una representación eficiente y ágil de la información heterogénea esencial para el desarrollo del proyecto.

La elección de MongoDB se justifica aún más por su capacidad para gestionar grandes volúmenes de lecturas y escrituras, lo cual es crucial en proyectos que involucran análisis retrospectivos y prospectivos. La capacidad de escalamiento horizontal de MongoDB permite la distribución de la carga de trabajo en clústeres, ofreciendo beneficios significativos al manejar grandes cantidades de datos y consultas complejas.

El proceso de carga de la base de datos se llevó a cabo mediante el siguiente código:

```
pip install pymongo
from pymongo.mongo_client import MongoClient
from pymongo.server_api import ServerApi

uri = "mongodb+srv://paula:P123456p@cluster0.e9bmxy.mongodb.net/?retryWrites=true&w=
majority"

# Create a new client and connect to the server
client = MongoClient(uri, server_api=ServerApi('1'))

# Send a ping to confirm a successful connection
try:
    client.admin.command('ping')
    print("Pinged your deployment. You successfully connected to MongoDB!")
except Exception as e:
    print(e)
# Seleccionar una base de datos específica
db = client["Proyecto"]

# Ahora puedes usar 'db' para referirte a tu base de datos
coleccion = db["Pronostico Muertes"]
# Intentando determinar el delimitador correcto del archivo CSV
import csv
import pandas as pd

# Ruta al archivo CSV
ruta_archivo_csv = '/content/BASE FINAL CAUSAS MUERTES 1990-2019.csv'

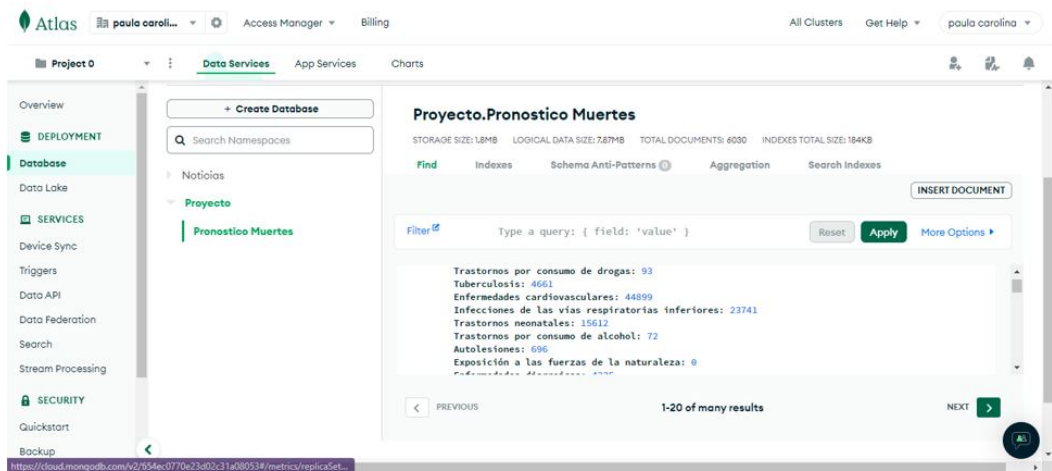
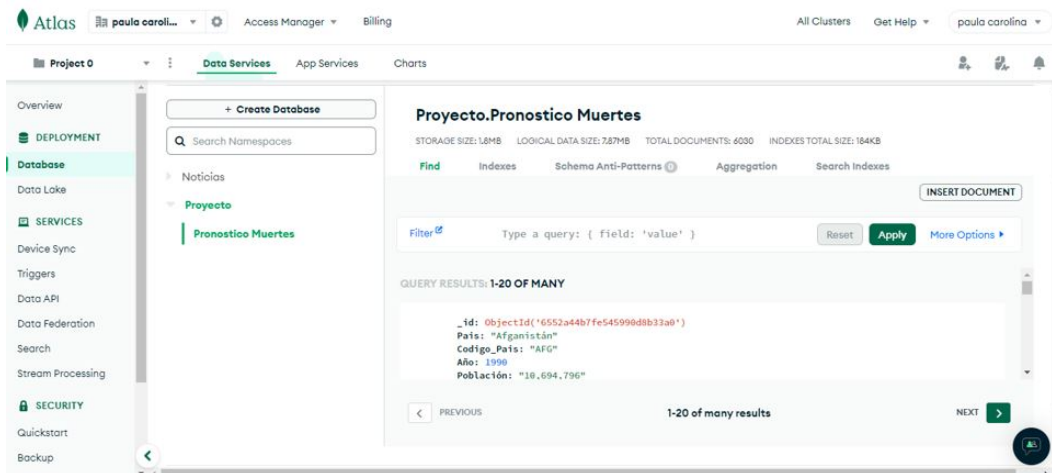
# Leer las primeras líneas del archivo para determinar el delimitador
with open(ruta_archivo_csv, 'r', encoding='utf-8') as archivo:
    muestra = archivo.read(1024)

delimitador = csv.Sniffer().sniff(muestra).delimiter

# Cargar el archivo CSV con el delimitador detectado
df = pd.read_csv(ruta_archivo_csv, delimiter=delimitador)
```

```
# Mostrar las primeras filas del DataFrame
df.head()
# Convertir el DataFrame de pandas a una lista de diccionarios
datos_para_insertar = df.to_dict(orient='records')

# Ahora puedes usar 'insert_many' con esta lista de diccionarios
coleccion.insert_many(datos_para_insertar)
```



6 Aplicación de ETL (Extract, Transform, Load)

En esta sección, detallamos el flujo de trabajo del proceso ETL utilizado en nuestro proyecto. Este enfoque, que comprende las fases de Extract, Transform y Load, es fundamental para asegurar la integración efectiva de datos en nuestra base de datos. A continuación, se describe cada etapa:

6.1 Ejemplo de aplicación de ETL

Etapas 1. Extracción

Durante esta fase, se identificaron y recuperaron los datos necesarios para nuestro proyecto desde diversas fuentes. La extracción de los datos no fue realizada a través de algoritmos y por lo tanto no fue requerido ningún lenguaje de programación, no obstante, a continuación, se describe el proceso llevado a cabo para obtener los datos.

1. Ingresar a la página web de Kaggle

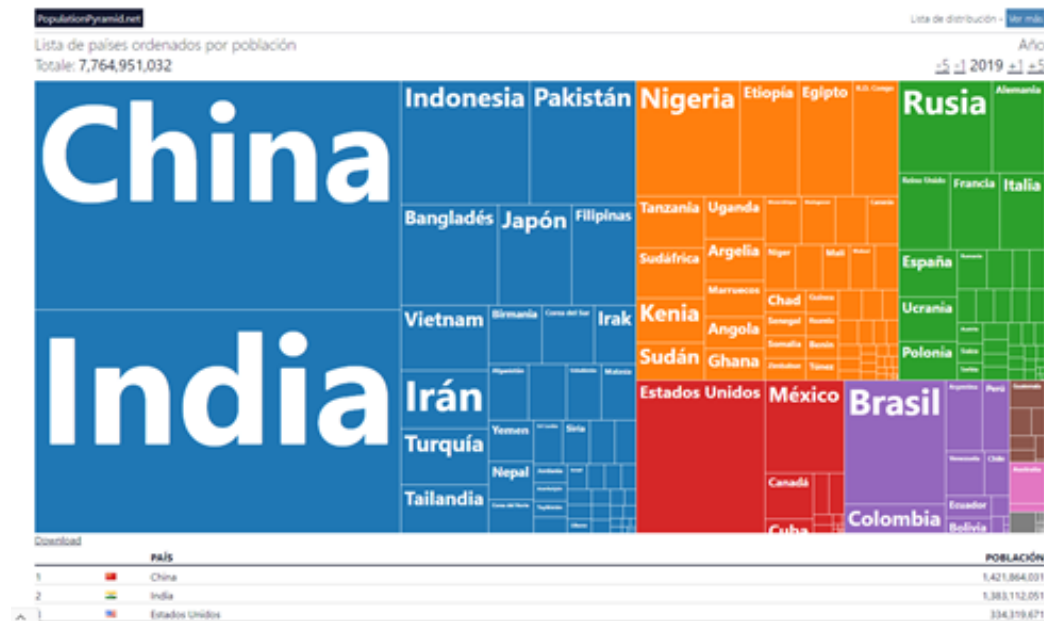
(<https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world>)



2. Descargamos la base Causas de Muerte en el Mundo

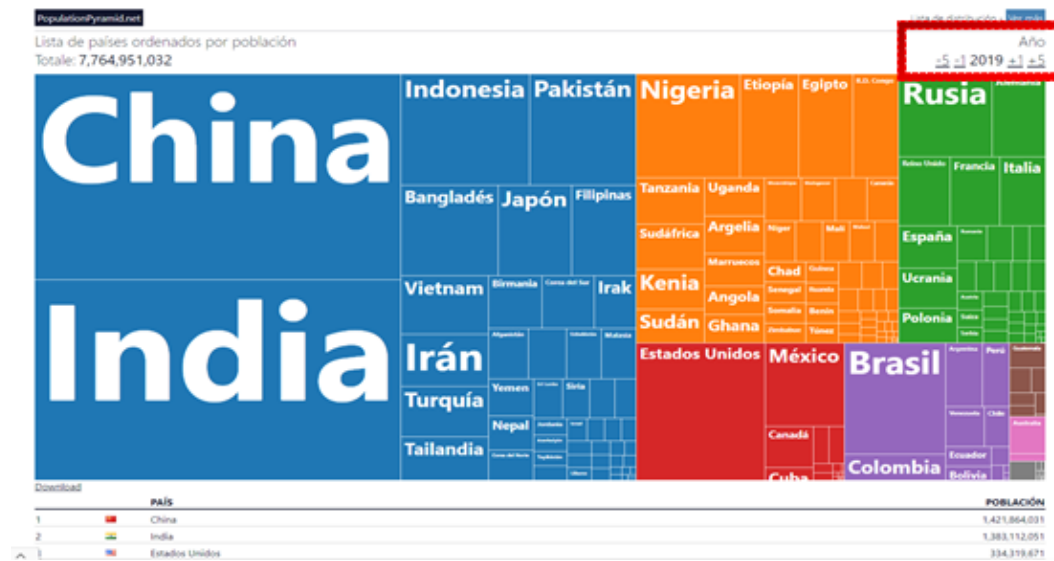


3. Ingresar a la página web de PopulationPyramid.net
(<https://www.populationpyramid.net/es/población-por-pais/2020/>)

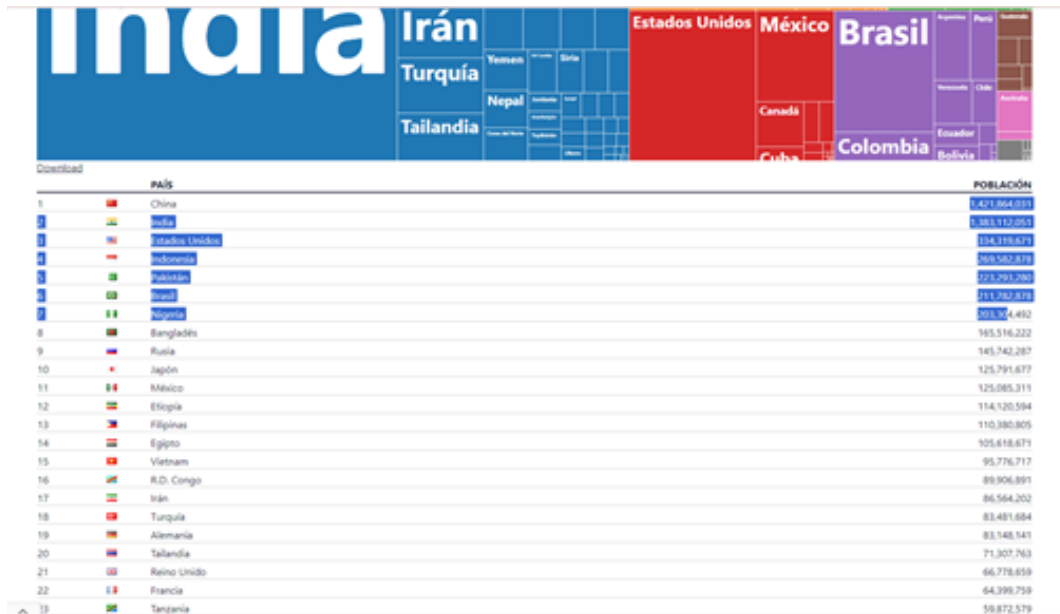


4. Seleccionamos los años que eran necesario para la información y en la tabla extraíamos de forma manual el total de población por país:

- Cambio de año



- Extracción de los datos de forma manual



5. Ingresar a la página web de Datosmacro.com
(<https://datosmacro.expansion.com/>)



6. Seleccionamos los indicadores económicos y ambientales, de esta página extraemos la información de los indicadores económicos y ambientales

- Selección de los indicadores:

The screenshot shows the Datosmacro.com website. The navigation menu includes: Datos, Países, Análisis, Diccionario, and Buscador. The main content area features a large map of the world with color-coded regions, a section for 'Calidad de vida' (Quality of Life) with a photo of a person, and a 'COMPARAR PAÍSES' (Compare Countries) section. The 'COMPARAR PAÍSES' section has dropdown menus for 'Nauru' and 'Ghana' and a 'Comparar' button. Below this, there are 'ÚLTIMOS DATOS PUBLICADOS' (Latest Published Data) with news items about Canada's IPC and Portugal's industrial prices.

- Vista de los indicadores

The screenshot shows the Datosmacro.com website with detailed views of various indicators. The left panel displays 'Emisiones de CO2' (CO2 Emissions) and 'Emisiones de metano (CH4)' (Methane Emissions) with tables of data for different countries. The right panel displays 'Cuentas Nacionales / Gobierno' (National Accounts / Government) and 'PIB anual a precios de mercado' (Annual GDP at market prices) with tables of data for different countries. The bottom section shows 'Deuda pública en Millones de EUR' (Public Debt in Millions of EUR) and 'Déficit público' (Public Deficit) with tables of data for different countries. The website also features a 'COMPARAR PAÍSES' (Compare Countries) section and a 'ÚLTIMOS DATOS PUBLICADOS' (Latest Published Data) section.

- Las cifras se sacaron de forma manual.

Datosmacro.com
Datos
Países
Análisis
Diccionario
Buscador
€
Search
f
t
in

Expansión / Datosmacro.com
ORBYT. Ediciones

BOLSAS
BONOS
DEUDA
DIVISAS
DÉFICIT
EPA
HIPOTECAS
IPC
MATERIAS PRIMAS
PIB
PARO
PRIMA DE RIESGO
RATINGS
SALARIO SMI
TIPOS

Manage All IIS Service Logs
Get an IIS log file analyzer for free with our 30-day fully functional free trial.
ManageEngine

Download

PIB - Producto Interior Bruto
¿Quieres que te enviemos a tu correo?
f
t
in

Afghanistan - PIB anual a precios de merca...
Ver además> Deuda pública Déficit público G. Público (M.€) Ratings

PIB anual
PIB Per Capita
PIB Trimestral
PIB Trim Per Capita

Comparativa: PIB anual

Países	Fecha	PIB anual	Var. PIB (%)
España [+]	2022	1.400.528 M\$	5,8%
Alemania [+]	2022	4.075.395 M\$	1,8%
Reino Unido [+]	2022	3.070.600 M\$	4,3%
Francia [+]	2022	2.782.905 M\$	2,5%
Italia [+]	2022	2.012.013 M\$	3,7%
Portugal [+]	2022	252.381 M\$	6,8%
Zona Euro [+]	2022	14.040.894 M\$	3,4%
Estados Unidos [+]	2022	25.744.100 M\$	1,9%
Japón [+]	2022	4.233.538 M\$	1,0%

#COVID-19: 09/11/2023
Confirmados: 771.820.173 (+0)
Muertos: 6.978.162 (+0)
Completamente vacunadas: 5.176.746.302 (65,62%)

COMPARAR PAÍSES
Burkina Faso
Kiribati
Comparar

BUSINESS STUDY TOURS BY ESIC
ME INTERESA

Etapa 2. Transformación

Los datos extraídos fueron sometidos a procesos de transformación para adaptarse a la estructura y requisitos de nuestra base de datos. Esto incluyó la limpieza, normalización y enriquecimiento de los datos para garantizar su coherencia y utilidad.

En el presente proyecto se realizó una primera transformación fue la unificación de los datos para poder con una sola base consolidada y robusta.

La base de datos inicial reveló la presencia de numerosos valores nulos en las variables económicas y ambientales. Para abordar esta limitación, se llevó a cabo una transformación mediante la imputación de datos en Python, utilizando la técnica de vecinos cercanos:

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import plotly.express as px
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import pairwise_distances, silhouette_score, silhouette_samples,
davies_bouldin_score
from sklearn.cluster import KMeans, AgglomerativeClustering
from scipy.cluster.hierarchy import dendrogram
import scipy.cluster.hierarchy as sch
from scipy import cluster
from scipy.stats import f_oneway, chi2_contingency, kruskal
from sklearn.impute import KNNImputer
from yellowbrick.cluster import KElbowVisualizer, silhouette_visualizer

path="/content/BASE FINAL CAUSAS MUERTES 1990-2019.xlsx"
base=pd.read_excel(path, index_col=0)
base

# conocemos las columnas de la base
base.columns
```

```
#Dentificamos las columnas que contienen información cuantitativa esencial para
nuestro análisis.
basecuanti=base[['pib', 'idh', 'esperanza_vida_m','esperanza_vida_h',
'esperanza_vida', 'emisiones_CO2','emisiones_CO2_PIB', 'emisiones_CO2_per_capita']]
#Configuración del Imputador KNN
imputer = KNNImputer(n_neighbors=3, weights="uniform")
#Aplicamos el imputador a nuestras columnas cuantitativas seleccionadas.
imputados_uni = imputer.fit_transform(basecuanti)
baseNA = pd.DataFrame(imputados_uni, columns=['pib', 'idh',
'esperanza_vida_m','esperanza_vida_h', 'esperanza_vida', 'emisiones_CO2',
'emisiones_CO2_PIB', 'emisiones_CO2_per_capita'],
index=base.index)
baseNA
#Confirmamos que no existan nulos en las variables imputadas
baseNA.isnull().sum()

#Creamos y descargamos una base sin nulos:
base_final.to_excel('BASE FINAL CAUSAS MUERTES 1990-2019 SIN NULOS.xlsx', index=True)
```

Finalmente, disponemos de una base de datos completa, sin valores nulos, que nos permite llevar a cabo el análisis y pronóstico correspondiente, cumpliendo así con el objetivo establecido en este proyecto.

- Oracle Server

Los datos extraídos se sometieron a un proceso de transformación para ajustarse al modelo de datos relacional de Oracle Server. Esto incluyó la normalización de tablas, la validación de integridad referencial y la adaptación de tipos de datos.

- MongoDB

Dado que MongoDB es una base de datos NoSQL, la fase de extracción y transformación se simplificó en comparación con Oracle Server. Los datos se extrajeron en su formato original y se

realizaron transformaciones mínimas, ya que MongoDB es más flexible en términos de esquema.

Etapa 3. Carga

La fase de carga de datos es esencial para consolidar la información transformada en un entorno estructurado y accesible. Este proceso se llevó a cabo mediante la implementación de bases de datos en varios sistemas de gestión. A continuación, se detallan los aspectos clave de este proceso para cada sistema:

- Oracle Server

La carga de datos se llevó a cabo de manera integrada con la creación de tablas mediante scripts SQL. Estos scripts no solo definieron la estructura de las tablas, sino que también realizaron la inserción de datos transformados. Durante este proceso, se implementaron mecanismos para garantizar la coherencia y la integridad de los datos recién ingresados.

- MongoDB

La base de datos para el proyecto "Perspectivas Integradas" se cargó en MongoDB, seleccionándolo como el Sistema de Gestión de Bases de Datos (SGBD) debido a su naturaleza altamente escalable y flexible en el ámbito de las bases de datos NoSQL. Esta elección se basó en la necesidad de manipular extensos conjuntos de datos que abarcan información detallada de todos los países desde 1990 hasta 2019.

7 Proximos pasos

La fase siguiente del proyecto se enfocará en llevar a cabo los procesos esenciales para alcanzar el objetivo establecido: analizar la mortalidad por infecciones respiratorias en cinco potencias globales (Estados Unidos, China, Japón, Alemania y Brasil) durante el periodo conocido (1990-2019) y el periodo a predecir (2020-2025). A partir de la información integrada en la base de datos, se han delineado los siguientes pasos a seguir:

- **Análisis Exploratorio de Datos:**

Después de completar la imputación de datos faltantes utilizando las técnicas estadísticas previamente mencionadas durante el proceso de Extracción, Transformación y Carga (ETL), y al establecer una base sólida para el procesamiento de la información, nos sumergimos en un análisis exploratorio detallado. Este análisis tiene como objetivo identificar patrones y tendencias preliminares entre las variables, así como posibles relaciones entre ellas, proporcionando una visión más clara del caso de estudio en cuestión.

Al obtener una comprensión profunda de la situación mediante este análisis exploratorio, logramos un panorama más claro del problema. Este conocimiento detallado y contextualizado sienta las bases para la selección adecuada de modelos en la resolución del problema. En última instancia, la combinación de la imputación de datos y el análisis exploratorio de datos facilita la toma de decisiones informadas, permitiendo abordar el problema con soluciones más precisas y eficientes.

- **Construcción de Modelos de Pronóstico:**

La construcción de modelos de pronóstico será una fase crucial, utilizando los resultados del análisis exploratorio. Se establecerá y ratificará la correcta selección de las mejores variables predictoras, así como las limitaciones encontradas en la información. Todo el proceso correspondiente a la verificación y cumplimiento de las hipótesis requeridas de los modelos a implementar se llevará a cabo, incluyendo diferentes transformaciones y técnicas necesarias. Finalmente, se entrenarán los modelos que mejor se ajusten al comportamiento de los datos, se calcularán los pronósticos y se evaluarán los resultados con criterios adecuados en series de tiempo, proporcionando una base predictiva para las tasas de mortalidad por infecciones respiratorias en los cinco países seleccionados.

- **Gestión Eficiente de Datos a Gran Escala:**

Se mantendrá un enfoque continuo en la gestión eficiente de datos a gran escala, utilizando sistemas avanzados de analítica para procesar la información de manera efectiva, garantizando la seguridad y la integridad de los datos. Además, se actualizará la información de los datos reales a medida que se obtenga de fuentes confiables.

- **Enfoque Analítico Continuo:**

Se seguirá un enfoque analítico continuo, utilizando métodos estadísticos y técnicas de modelado de bases de datos para identificar patrones, tendencias y relaciones en los datos, con énfasis en la evolución a lo largo de los años. Se realizará la evaluación de los pronósticos versus los datos reales que vayan apareciendo con el tiempo, con el fin de auditar el funcionamiento del modelo escogido para el pronóstico. Los resultados obtenidos serán sometidos a validación, ya sea mediante técnicas de validación cruzada o comparación con conjuntos de datos independientes para fortalecer la confiabilidad de los resultados.

- **Análisis de Escenarios Futuros:**

Se explorarán diversos escenarios futuros de mortalidad por infecciones respiratorias para el periodo de 2020 a 2025, considerando la influencia de variables económicas, ambientales y sociodemográficas en estas proyecciones.

- **Documentación Detallada:**

Cada paso del proyecto será documentado detalladamente, proporcionando transparencia en los procesos, metodologías y decisiones tomadas. Esta documentación facilitará la revisión por pares y la replicación del estudio.

- **Elaboración de Conclusiones y Recomendaciones:**

Con base en los hallazgos, se desarrollarán conclusiones sólidas y se proporcionarán recomendaciones para enfoques informados y estratégicos en la promoción de la salud para los países en el estudio.

- **Comunicación de Resultados:**

Finalmente, se preparará informes y presentaciones para comunicar los resultados a audiencias relevantes, incluyendo la comunidad científica, profesionales de la salud y responsables de la toma de decisiones. Estos pasos forman parte de una metodología integral que busca no solo comprender las tendencias pasadas, sino también proyectar de manera informada el futuro de la mortalidad por infecciones respiratorias a nivel global.

8 Lecciones aprendidas

La realización de un análisis exhaustivo sobre la mortalidad por infecciones respiratorias en cinco potencias globales durante el extenso periodo de 1990 a 2025 ha sido un desafío que ha llevado al equipo a una travesía de preparación, ejecución y reflexión continua. Este proyecto no solo abordó la complejidad de consolidar datos históricos y contemporáneos, sino que también buscó comprender las interrelaciones entre variables económicas y ambientales con el fin de proyectar posibles escenarios futuros.

Se comparte las lecciones aprendidas a lo largo de este proyecto, donde se aborda en tres puntos principales: Preparación, Ejecución y Recomendaciones. Desde los desafíos iniciales hasta la implementación exitosa de modelos de datos en entornos SQL y No SQL, cada paso proporcionó valiosas enseñanzas y se considera esenciales para futuros análisis y proyectos similares.

Ofrece una visión integral del proceso, y resalta la importancia de la preparación meticulosa, la ejecución precisa y las recomendaciones fundamentales para abordar proyectos de investigación de esta magnitud.

1. Preparación:

- Desafío Inicial:

Afronta el desafío de obtener una base de datos inicial que incluye información sobre países, años y diversas causas de muerte. La integración de esta base con datos económicos como el PIB, Índice de Desarrollo Humano, esperanza de vida y emisiones de CO2 fue un proceso complejo.

- Identificación de Variables:

Identificar variables cualitativas y cuantitativas fue esencial para determinar el objetivo del análisis y definir el enfoque del proyecto.

- Modelo de Datos:

La implementación del Modelo de Datos en diferentes entornos, como SQL y MongoDB, es crucial para el éxito del proyecto. Establecer el objetivo del análisis es fundamental para la selección de variables y la elección de herramientas de procesamiento de datos.

2. Ejecución:

- Modelo de Datos en SQL y No SQL:

Después de consolidar la información, se construye el Modelo de Datos en SQL y No SQL (Oracle y MongoDB), en donde la optimización de los resultados garantiza la concisión y verificabilidad de la información.

- Interpretación de Datos:

La elección de herramientas de procesamiento de datos facilitó la exploración y análisis de datos, asegura la interpretación precisa para cumplir con los objetivos del proyecto.

3. Recomendaciones:

- Calidad de los Datos:

La importancia de la calidad de los datos para un análisis efectivo e identificación de variables objetivo.

- Uso de Modelos Estadísticos:

Se recomienda la aplicación de diversos modelos estadísticos para la clasificación y comprensión de los datos.

4. Definición del Objetivo:

- La definición clara del objetivo es vital para guiar cada etapa del desarrollo y análisis.
- Planeación y Seguimiento:
La planificación y el seguimiento riguroso de las actividades son críticas para evitar retrasos en el proyecto.
- Toma de Decisiones:
Se destaca la importancia de la toma de decisiones oportuna en momentos cruciales para garantizar la continuidad del proyecto.
- Directriz del profesor:
SApoyo valioso del profesor en la resolución de dudas durante las sesiones de seguimiento.

Estas lecciones aprendidas no solo contribuyo al éxito del proyecto, sino también proporciona valiosas pautas para futuros análisis de datos y proyectos similares. El enfoque en la calidad, la planificación y la toma de decisiones informada son fundamentales para superar los desafíos y alcanzar los resultados deseados.

9 Bibliografía

Kaggle. (2022). Cause of Deaths Around the World (Historical Data).
<https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world>

PopulationPyramid. (1990). Población por país en 1990.
<https://www.populationpyramid.net/es/población-por-pais/1990/>

Datosmacro.expansion.com. (2023). Datos macroeconómicos y financieros.
<https://datosmacro.expansion.com>

Oracle Corporation. (2023). Oracle SQL Developer [Versión 23.1.0.097]. Oracle Corporation.
<https://www.oracle.com/database/sqldeveloper/>

Oracle Corporation. (2023). Oracle SQL Data Modeler [23.1.0.087]. Oracle Corporation.
<https://www.oracle.com/es/database/sqldeveloper/technologies/sql-data-modeler/>

MIRO. The Visual Workspace for Innovation. (s. f.-c).
<https://miro.com/>

MongoDB. (s. f.). MongoDB: the Developer Data platform.
<https://mongodb.com/>