



EA1. Formulación de una necesidad de ingeniería de datos

JUAN ESTEBAN ATEHORTUA SANCHEZ

Proyecto Integrado V - Línea de Énfasis
PREICA2502B020074

ANDRES FELIPE CALLEJAS

MEDELLIN

2025

Análisis Económico Global (1960–2023)

- **Tema:** Observación de indicadores macroeconómicos (PIB, exportaciones, importaciones e inflación)
- **Fuente de datos:**
 - Global Imports of Goods and Services (1960–Present) — *Frederick Salazar (2023)*
 - Global Exports of Goods and Services (1960–Present) — *Frederick Salazar (2023)*
 - PIB (GDP) Global by Countries since 1960 to 2021 — *Frederick Salazar (2023)*
 - Global Inflation Rate (1960–Present) — *Frederick Salazar (2023)*

Objetivo General

Analizar la evolución económica global entre 1960 y 2023 a partir de la integración de datos públicos del Banco Mundial relacionados con el Producto Interno Bruto (PIB), las exportaciones, las importaciones y la inflación. El propósito es comprender, mediante un enfoque analítico, las dinámicas del crecimiento económico y su relación con los principales indicadores comerciales y monetarios, identificando tendencias y comportamientos relevantes a nivel internacional.

Objetivos Específicos

- Analizar la evolución histórica del PIB global para interpretar las tendencias del crecimiento económico y sus variaciones entre 1960 y 2023 desde una perspectiva descriptiva y analítica.
- Examinar las tendencias de las importaciones y exportaciones como proporción del PIB con el fin de interpretar los niveles de apertura comercial y su posible relación con los procesos de desarrollo económico global.
- Analizar el comportamiento de la inflación anual a nivel mundial para identificar patrones vinculados con periodos de estabilidad y desequilibrio económico, considerando su incidencia en el desempeño general de las economías.
- Explorar las relaciones entre el PIB, las importaciones, las exportaciones y la inflación para detectar comportamientos comunes, contrastes y dinámicas regionales que aporten a la comprensión integral del sistema económico global.
- Visualizar e interpretar la evolución de los indicadores económicos mediante representaciones analíticas que faciliten la comprensión de las tendencias, correlaciones y transformaciones a lo largo del tiempo.
- Reflexionar sobre los hallazgos obtenidos a partir del análisis de los datos, valorando su aporte para la interpretación de los fenómenos económicos internacionales y la formulación de perspectivas futuras de estudio.

Ejemplos de Aplicaciones Potenciales

Las siguientes líneas de análisis ilustran posibles extensiones o aplicaciones del modelo desarrollado. Cada una de ellas aprovecha la estructura de datos integrada para profundizar en distintos aspectos de la dinámica económica global, priorizando la interpretación analítica de los resultados.

Línea de análisis	Descripción
Relación entre PIB y comercio exterior	Analizar si los países con mayores niveles de apertura comercial (exportaciones e importaciones) mantienen un crecimiento sostenido del PIB a lo largo del tiempo.
Impacto de la inflación en el desempeño económico	Explorar cómo los niveles de inflación sostenidos o extremos inciden en el crecimiento económico y la estabilidad macroeconómica de los países.
Comparativo regional	Contrastar el comportamiento económico entre regiones (por ejemplo, América Latina, Europa y Asia) para identificar diferencias estructurales y niveles de resiliencia económica.
Monitoreo de crisis económicas	Detectar y caracterizar periodos de crisis a partir de caídas simultáneas en el PIB y el comercio exterior (como la crisis del petróleo, la crisis asiática o la pandemia de COVID-19).

Enfoque metodológico orientado al uso de datos

El desarrollo del proyecto se sustentó en un enfoque metodológico basado en el uso, tratamiento y análisis de datos económicos abiertos. Dicho enfoque privilegia la interpretación analítica de la información por encima de la aplicación de modelos estadísticos, buscando generar conocimiento a partir de la exploración estructurada y visual de los indicadores.

Las principales etapas metodológicas fueron las siguientes:

1. **Recolección y limpieza de datos:** Se recopilaron indicadores económicos públicos del Banco Mundial, asegurando la coherencia temporal y la comparabilidad entre países y regiones. El proceso de limpieza incluyó la homologación de unidades, la gestión de valores faltantes y la verificación de consistencia en las series históricas.
2. **Modelado relacional (SQLite):** Se diseñó una estructura de base de datos relacional en SQLite para organizar los indicadores y permitir la ejecución de consultas analíticas complejas. Esta estructura facilita el cruce de variables económicas, la segmentación por periodos y regiones, y la reutilización del modelo para futuros análisis.
3. **Exploración de patrones históricos:** A través de consultas SQL complementadas con herramientas de análisis en *pandas* y *matplotlib*, se exploraron tendencias, relaciones y variaciones entre los indicadores a lo largo del tiempo. Esta etapa permitió identificar comportamientos recurrentes y diferencias estructurales entre economías.
4. **Generación de insights visuales:** Los resultados se sintetizaron mediante gráficos, paneles interactivos y reportes comparativos que facilitaron la interpretación de las dinámicas económicas globales. Estas visualizaciones sirvieron como soporte para el análisis descriptivo y la comunicación clara de los hallazgos.

Estructura del Proyecto

proyecto

— README.md	← Documentación principal
— run.ipynb	← Notebook ejecutable del pipeline y consultas
— data/	
— raw_data/	← Datos originales descargados de Kaggle
— normalized_data/	← Datos estandarizados a separacion por comas
— unified_data/	← Primer resultado de integración
— unified_clean/	← Versión final limpia y normalizada (ETL)
— db/	
— project.db	← Base de datos SQLite final
— docs/	
— EA1.db	← Documentación de Evidencia 1
— gantt.md	← Planificación del proyecto

Datasets Utilizados

Indicador	Dataset	URL	Licencia
Importaciones (% PIB)	<i>Global Imports of Goods and Services (1960–Present)</i>	Kaggle	PDDL
Exportaciones (% PIB)	<i>Global Exports of Goods and Services (1960–Present)</i>	Kaggle	PDDL
PIB Global (1960–2021)	<i>PIB (GDP) Global by Countries since 1960 to 2021</i>	Kaggle	PDDL
Inflación (% anual)	<i>Global Inflation Rate (1960–Present)</i>	Kaggle	PDDL

Proceso Metodológico (ETL)

El proceso metodológico se estructuró bajo un enfoque **ETL (Extract, Transform, Load)**, orientado a garantizar la calidad, coherencia y utilidad analítica de los datos económicos utilizados. Este enfoque permitió transformar un conjunto heterogéneo de archivos provenientes del Banco Mundial en una base de datos integrada, limpia y lista para su exploración.

El ciclo ETL se diseñó con el propósito de **extraer la información relevante, depurarla y normalizarla** según criterios uniformes, y finalmente **cargarla en un modelo relacional** que facilite la realización de consultas analíticas complejas. Más que un proceso técnico aislado, este procedimiento constituyó la base metodológica del proyecto, ya que permitió consolidar datos comparables entre países, regiones y periodos, garantizando la trazabilidad y consistencia del análisis posterior.

A lo largo de este proceso se desarrollaron diversas etapas —desde la limpieza inicial y la estandarización de campos hasta la unificación y modelado relacional—, cada una orientada a optimizar la calidad del conjunto de datos y maximizar su potencial analítico.

1. Limpieza Inicial

La primera etapa del proceso consistió en la depuración y estandarización de los datos obtenidos del Banco Mundial. El objetivo fue garantizar la coherencia estructural y facilitar su posterior integración en un modelo relacional. Entre las principales acciones realizadas se incluyen:

- a. **Conversión de separadores:** transformación de los delimitadores originales para establecer una separación uniforme por comas (,), asegurando la correcta lectura de los archivos por los sistemas de análisis.

- b. **Estandarización de nombres de campos:** adopción de la convención *snake_case* para unificar la nomenclatura de variables y mejorar la legibilidad del código.
- c. **Tratamiento de valores faltantes:** detección y sustitución de vacíos (*NaN*) por la etiqueta "N/A", con el propósito de conservar la integridad del conjunto de datos y evitar errores durante el procesamiento posterior.

2. Normalización y Unificación de Datos

En esta fase se buscó consolidar la información en un formato homogéneo que permitiera su análisis transversal y temporal.

El proceso incluyó las siguientes operaciones:

- a. **Eliminación de duplicados:** mediante la combinación de las claves (*country_code*, *year*, *indicator_code*) para asegurar la unicidad de los registros.
- b. **Clasificación de registros:** identificación de agregados regionales a través del campo *is_aggregate* = 1, diferenciándolos de los datos correspondientes a países individuales.
- c. **Estructuración del modelo relacional:** creación de un conjunto de tablas normalizadas que facilitan las consultas analíticas:
 - *dim_geo*: contiene información geográfica sobre países, regiones y grupos económicos.
 - *dim_indicator*: almacena la descripción y metadatos de los indicadores económicos.
 - *fact_indicators*: tabla de hechos principal, con los valores de cada indicador por país y año.
 - *fact_wide*: versión pivotada que consolida los indicadores por país/año, útil para análisis comparativos y visualizaciones.

Base de Datos (SQLite)

- **Archivo:**

db/project.db

- **Tablas principales**

Tabla	Descripción
dim_geo	Dimensión geográfica (países y regiones)
dim_indicator	Catálogo de indicadores económicos
fact_indicators	Hechos normalizados por país/año/indicador
fact_wide	Versión pivotada para análisis rápido
vw_exports_imports	Vista SQL para comparar exportaciones/importaciones

Resultados y Análisis

El desarrollo del proyecto permitió integrar una base de datos global con más de seis décadas de información económica, abarcando el periodo comprendido entre 1960 y 2023. Esta integración facilitó el análisis conjunto de los principales indicadores del Banco Mundial —Producto Interno Bruto (PIB), inflación, exportaciones e importaciones—, ofreciendo una visión amplia y coherente de la evolución económica mundial.

El análisis de los indicadores permitió identificar relaciones consistentes entre el crecimiento del PIB, los niveles de inflación y la dinámica del comercio exterior. Estos resultados reflejan cómo los procesos de apertura comercial y las variaciones monetarias han acompañado, en distintos grados, las trayectorias de crecimiento de las economías a lo largo del tiempo.

Asimismo, el diseño de la base analítica posibilitó realizar consultas y comparaciones por país, década y región, lo que amplía las posibilidades de interpretación y permite examinar las particularidades económicas de cada contexto geográfico y temporal.

Finalmente, la estructura del modelo facilita la incorporación de nuevos indicadores del Banco Mundial, lo que permite replicar y ampliar el análisis en investigaciones futuras, manteniendo la consistencia analítica y la comparabilidad de los resultados obtenidos.

Bibliografía

Salazar, F. (2023). *Global Imports of Goods and Services (1960–Present)* [dataset].

Kaggle. <https://www.kaggle.com/datasets/fredericksalazar/global-imports-goods-and-services-1960-present>

Salazar, F. (2023). *Global Exports of Goods and Services (1960–Present)* [dataset].

Kaggle. <https://www.kaggle.com/datasets/fredericksalazar/global-exports-of-goods-and-services-1960-present>

Salazar, F. (2023). *PIB (GDP) Global by Countries since 1960 to 2021* [dataset].

Kaggle. <https://www.kaggle.com/datasets/fredericksalazar/pib-gdp-global-by-countries-since-1960-to-2021>

Salazar, F. (2023). *Global Inflation Rate (1960–Present)* [dataset].

Kaggle. <https://www.kaggle.com/datasets/fredericksalazar/global-inflation-rate-1960-present>

World Bank. (2023). *World Development Indicators*. The World Bank

Group. <https://databank.worldbank.org/source/world-development-indicators>