conceptosclaros²

# VARIABLES. ESTRATEGIA RÁPIDA. LIMPIEZA. EDA FAST

## Tabla de las variables.

En el ejemplo que vamos a trabajar de world data:

**Link datos:**
**https://drive.google.com/file/d/1eO2yS9FCoaRJDW7dxvv8vqimsJkU8Of5/view?usp=sharing**

Te he creado la tabla de datos con las explicaciones.
Es importante clasificar el tipo de variables porque va a depender de ello la manera en que vas a describir los datos.

| Nombre Variables | Explicación | Tipo Variable |
|---|---|---|
| Country Name | Name of the Country | Nominal |
| Country Code | Globally Recognized 3 Character Country Code | Nominal |
| Continent Name | Name of continent | Nominal |
| Year | The Year of The Measurement | Ordinal, discreta |
| Agriculture (% GDP) | **Agricultural Activities (% in GDP): This metric measures the contribution of the agricultural sector to a nation's economy.** | Continua |
| Ease of Doing Business | The Evaluation of the World Bank on how easy it is to launch a new business in the country | Continua |
| Education Expenditure (% GDP) | **Education Expenditures (% in GDP): The percentage of GDP spent on education signifies a nation's investment in its human capital and future workforce.** | Continua |
| Export (% GDP) | **Export (% in GDP): Export as a percentage of GDP illustrates a country's global trade competitiveness and its reliance on international markets.** | Continua |
| GDP | **GDP (Gross Domestic Product): GDP measures the total economic output of a country, reflecting its economic health and performance.** | Continua |
| Health Expenditure (% GDP) | **Health Expenditures (% in GDP): This metric reflects the proportion of a country's GDP allocated to healthcare, highlighting its commitment to public health.** | Continua |
| Import (% GDP) | **Import (% in GDP): Import as a percentage of GDP demonstrates a country's dependence on foreign goods and services.** | Continua |
| Industry (% GDP) | **Industrial Activities (% in GDP): Industrial activities as a percentage of GDP signify the role of manufacturing and industry in a country's economy.** | Continua |
| Inflation Rate | **Inflation Rate: The inflation rate indicates the rate at which the general price level of goods and services rises, affecting a country's purchasing power.** | Continua |
| R&D | **Research And Development (R&D): R&D investment indicates a country's commitment to innovation and technological advancement.** | Continua |

| | | |
|---|---|---|
| Service (% GDP) | **Service Activities (% in GDP): The service sector's percentage of GDP shows the importance of services in economic activities.** | Continua |
| Unemployment | **Unemployment Rate: The unemployment rate measures the percentage of the labor force that is without jobs but is available for and seeking employment.** | Continua |
| Population | **Population: The total population of a country reflects its size and demographic composition.** | Continua |
| Land | **Land Area: The total land area of a country in square kilometers or square miles, which encompasses all land used for agriculture, forests, and other purposes.** | Continua |
| Export | **Exports: The total value of all goods and services sold by a country to entities in other countries over a specific period of time.** | Continua |
| Import | **Imports: The total value of all goods and services purchased by a country from entities in other countries over a specific period of time.** | Continua |
| Education Expenditure | **Education Spending: The total public and private expenditure on education, including salaries of teachers and costs for materials, over a specific period of time, usually one year.** | Continua |
| Health Expenditure | **Health Spending: The total public and private expenditure on health, including preventive and curative health services, family planning, nutrition activities, and emergency aid designated for health, over a specific period of time, usually one year.** | Continua |
| Net Trade | **Net Trade: Net trade, calculated as the difference between exports and imports, provides insights into a country's trade surplus or deficit.** | Continua |
| GDP Per Capita | **GDP Per Capita: GDP per capita calculates the average income of each individual in a country and is an indicator of living standards.** | Continua |
| Population Density | **Population Density: Population density measures the number of individuals living per unit of area, providing insights into urbanization and land usage.** | Continua |
| Net Trade Positive | **Es 1 si el net trade es positivo. Y 0 si es negativo** | Nominal |

Cuando trabajes un proyecto acuérdate de tener esta tabla de variables.
Piensa también que puedes crear nuevas variables a partir de las que tienes.

Por ejemplo, aquí hemos creado Net Trade Positive que es 1 si el net trade es positivo, y 0 si es negativo.

## Definir la estrategia con CHAT GPT

Chat GPT es una herramienta maravillosa que puede ser tu consultor de datos en tu equipo. Solo tienes que pasarle la tabla de datos y comentarle que quieres analizar los datos y que te propongo una estrategia.

Yo uso el Data Analyst de Chat GPT. Lo puedes cargar en complementos.

El prompt o pregunta que hago es simple:

**Le copio la tabla de las variables y le comento lo siguiente:**

**“Estoy analizando estos datos y quiero plantear la estrategia de análisis de datos. Puedes plantear tres posibles estrategias o problemas a solucionar con las variables que te comparto?”**

Puedes seguir manteniendo la conversación para tratar de afinar los puntos a analizar.

**Experimenta con Chat GPT y crea tu estrategia.**

En mi caso salieron estos puntos:

**1. Evaluación de la Diversificación Económica y la Resiliencia:**

Problema: Dependencia excesiva de un sector económico, aumentando la vulnerabilidad a choques externos.

Estrategia: Analizar "Agriculture (% GDP)", "Industry (% GDP)" y "Service (% GDP)" para evaluar la diversificación económica. Investigar cómo esta diversificación afecta al "GDP Per Capita" y examinar la influencia de "Export (% GDP)" y "Import (% GDP)" en la estabilidad económica.

**2. Determinación de Factores Clave para el Crecimiento Económico:**

Problema: Identificación de inversiones que impulsan el crecimiento económico sostenible.

Estrategia: Estudiar la relación entre "Ease of Doing Business" y "GDP Per Capita" para encontrar correlaciones con la prosperidad económica. Correlacionar "Education Expenditure (% GDP)" y "Health Expenditure (% GDP)" con el "GDP Per Capita" para determinar si el gasto en salud y educación se asocia con mayores ingresos per cápita.

**3. Optimización de Políticas para la Mejora del Empleo y el Comercio:**

Problema: Tasas de desempleo altas y balanzas comerciales negativas.

Estrategia: Utilizar "Unemployment" para identificar patrones asociados con el desempleo. Investigar la correlación entre inversiones en "R&D" y la "Ease of Doing Business" con las tasas de empleo. Examinar "Net Trade Positive" para entender la dinámica del comercio internacional y el impacto de las políticas comerciales en la economía local.

**4. Análisis de la Densidad Poblacional y su Impacto en la Calidad de Vida:**

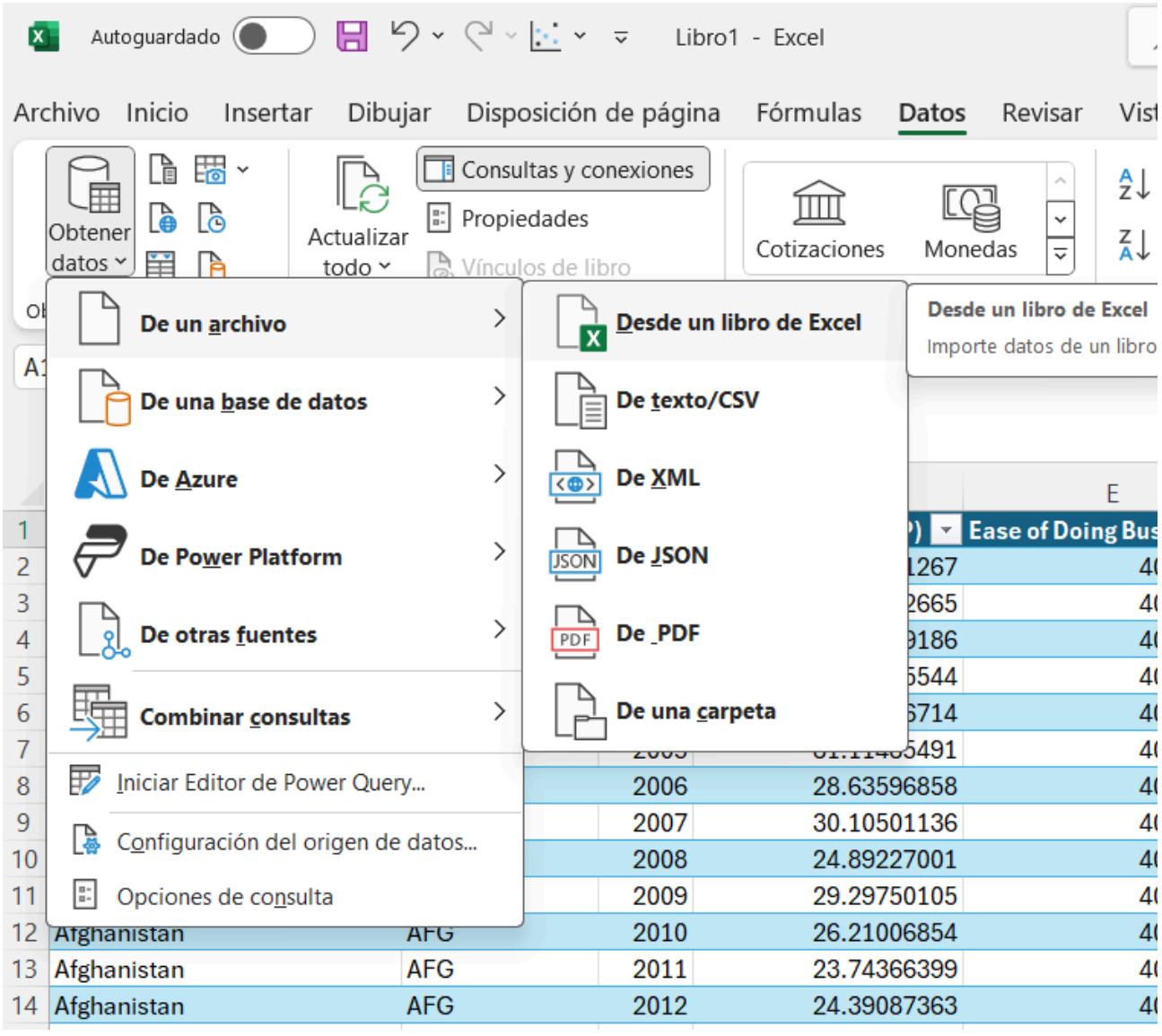
Problema: Posibles problemas asociados con la alta densidad poblacional, como la urbanización y el acceso a servicios.

Estrategia: Explorar la relación entre "Population Density" y variables como "GDP Per Capita", "Health Expenditure (% GDP)", y "Education Expenditure (% GDP)". Determinar el impacto de la densidad poblacional en la eficiencia de los servicios y su correlación con la prosperidad económica y el bienestar social. Evaluar la relación entre la densidad poblacional, el desempleo y la balanza comercial.

Se podrían añadir más, como analizar la tendencia en el tiempo de alguna indicadores para ver cómo han evolucionado los países en los últimos años.

## Leer los datos con Power Query

Para leer los datos uso Power Query de Excel.



Acuérdate de configurar en Opciones de Consulta el idioma para que las fechas y los separadores de decimales funcionen.

**Lee tus datos en tu Excel usando este proceso. Te ayudará para después poder limpiar los datos fácilmente.**

## Aplica EDA univariado para detectar errores

Aplica los siguientes gráficos a las variables numéricas:

EDA univariado:

Variables cuantitativas:

- Boxplot
- Histograma
- Media, desviación, mediana etc...

.
Variables categóricas:

- Tabla de frecuencia
- Diagrama de barras o sectores

**Aplica estos gráficos a las variables cuantitativas y cualitativas para validar que están correctas.**

Es una práctica muy sencilla y rápida que te ayuda a detectar errores en los datos.

¡A por esos primeros gráficos!

## Exploración de datos interactiva en Excel

Otra práctica muy usada es poder explorar los datos de forma interactiva con Excel usando las tablas dinámicas.

Analiza la evolución del % de agricultura, % servicio, % de industria en el tiempo de los diferentes países.

También hazlo por continentes.

Usa las tablas dinámicas y las segmentaciones como planteamos en la sesión.

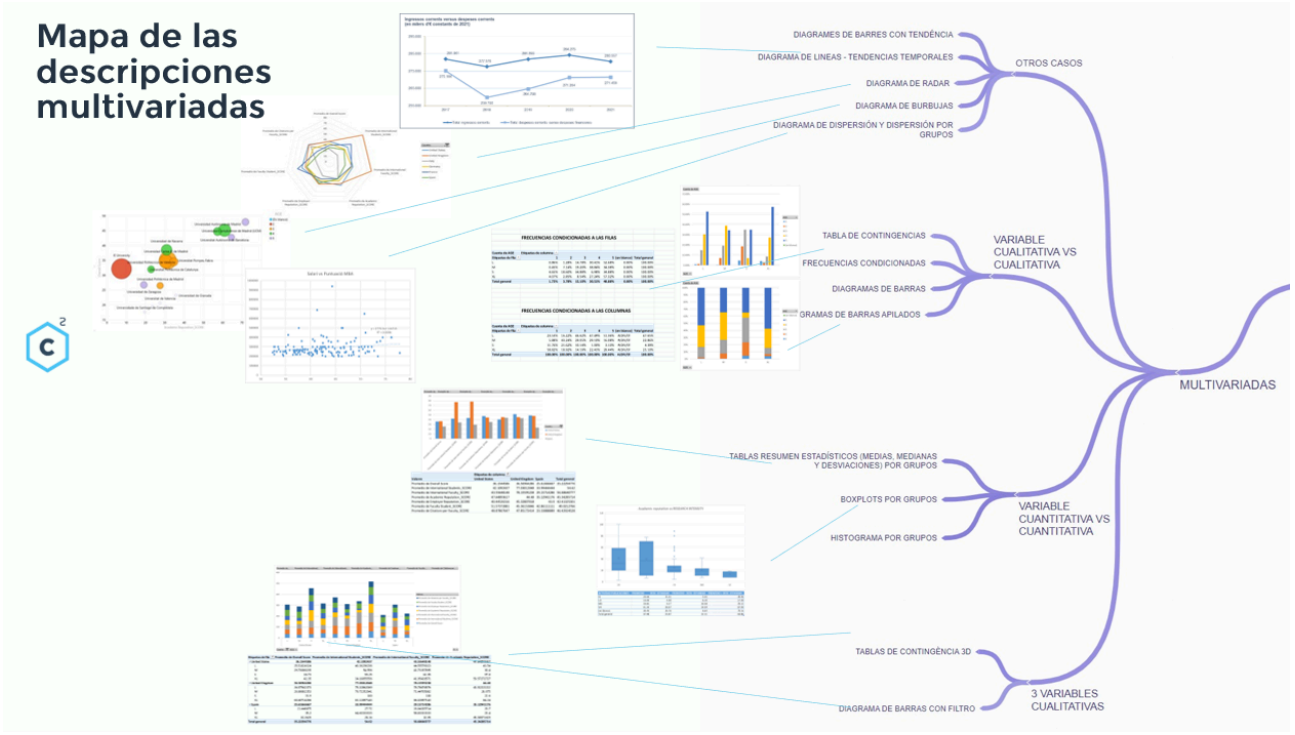
En 15 minutos puedes tener un dashboard en bruto para analizar tus datos.

# EDA multivariado

Aquí vamos a empezar a contestar algunas de las preguntas que hemos trabajado en la estrategia.

Según la pregunta/estrategia, las variables van a cambiar y por consecuencia la manera de explorar los datos también.

El mapa de la descripción multivariada es:




.

Aplica este paso a paso para empezar un proyecto y en 30 minutos puedes tener tus primeros resultados siempre y cuando la limpieza no sea muy compleja.

¡A aplicar estos pasos fast and furious!

Cuando termines tómate algo que te apetezca, un té, un zumo, una cerveza o lo que quieras.

Hay que celebrar las primeras victorias en ciencia de datos.

1 fuerte abrazo
Jordi Ollé