Proyecto Web Scraping con Scrapy

Empleando la información que se puede obtener mediante *Scraping* de la página web:

- https://www.cia.gov/the-world-factbook/countries

respondan las preguntas siguientes:

1. ¿Existe alguna relación entre el GDP (PIB) de un país, en términos de su poder adquisitivo, y su porcentaje de usuarios de Internet?
2. ¿Además, la tendencia resulta similar para países con ingreso económico bajo, mediano y alto?

Para realizar el proceso de scraping deben utilizar ***Scrapy*** y selectores ***XPath***. Además, el proyecto lo pueden realizar en equipos de entre uno y tres integrantes.

Evaluación

Para este proyecto serán evaluados dos aspectos (esto implica dos calificaciones de proyecto):

1. **Fundamentos de visualización**: que los gráficos empleados se presenten conforme a los 4 principios básicos de presentación de datos de Edward R. Tufte (maximizar la tasa de tinta de datos, minimizar el factor mentira, minimizar la *chatarra gráfica* y utilizar las escalas apropiadas y un etiquetado claro).
2. Proceso de **Web Scraping**, **Data Wrangling**, obtención de resultados y conclusiones/respuesta de las preguntas planteadas.

⋆ Fecha límite de entrega: lunes 09 de junio de 2025, 12:00 horas.

Información de los países del mundo

La información a recolectar de cada país es:

#	Atributo	Definición
1	nombre	Nombre corto del país
2	área	Superficie en km^2
3	población	Número de habitantes
4	gdp	Producto Interno Bruto (Real GDP per capita)
5	desempleo	Tasa de desempleo (Unemployment rate)
6	impuestos	Tasa de impuestos (Taxes and other revenues)
7	deuda	Deuda externa (Debt - external)
8	tasa_de_cambio	Tasa de cambio a dólares (Exchange rates in US Dollars)
9	usuarios_internet	Usuarios con acceso a Internet (Internet users - total)
10	porcentaje_internet	Porcentaje de la población con acceso a Internet (Internet users - percent of population)
11	aeropuertos	Número de aeropuertos (Airports)
12	barcos_mercantes	Barcos mercantes # Merchant marine
13	inversion_militar	Gasto militar en % del PIB (Military expenditures: % of GDP)
14	image_urls	Los URLs de las imágenes de la bandera nacional del país
15	images	Información adicional de cada imagen descargada
16	image_name	Nombre de cada imagen de la bandera nacional

> **Nota**.- El atributo `gdp` debe transformarse a dato categórico estableciendo la clasificación siguiente:
>
> - Ingreso bajo $GDP \leq 1.0\ trillions$
> - Ingreso medio $\Rightarrow 1.0 < GDP \leq 3.0\ trillions$
> - Ingreso alto $\Rightarrow GDP > 3.0\ trillions$

⋆ Una vez finalizado el proceso de scraping y obtenidos los datos, así como las imágenes de las banderas nacionales de cada país, deberán utilizar Pandas para leer el archivo `.csv` y realizar todo

el procesamiento necesario para limpiar los datos. Posteriormente, utilicen los gráficos que consideren convenientes para fundamentar las respuesta a las dos preguntas planteadas.