

# Ciencia de Datos: Visualización de la Información

UNAM-IIMAS, Semestre 2023-2

## Tarea-06: Proporciones, distribuciones y cantidades

### Antecedentes

La renta petrolera (% del Producto Interno Bruto, PIB) es la diferencia entre el valor de la producción de petróleo crudo a precios regionales y los costos totales de producción. La contabilidad de la contribución de los recursos naturales a la producción económica es importante en la construcción de un marco analítico para el desarrollo sostenible. En algunos países, las ganancias de los recursos naturales, especialmente de los combustibles fósiles y los minerales, representan una parte considerable del PIB, y gran parte de estas ganancias vienen en forma de rentas económicas: ingresos por encima del costo de extracción de los recursos. Las rentas de los recursos no renovables (combustibles fósiles y minerales), así como las rentas de la sobreexplotación de los bosques, indican la liquidación del stock de capital de un país. Cuando los países utilizan dichas rentas para respaldar el consumo actual en lugar de invertir en nuevo capital para reemplazar el que se está agotando, en realidad están tomando prestado contra su futuro.

### ***Concepto y Metodología Estadística:***

Las estimaciones de las rentas de los recursos naturales se calculan como la diferencia entre el precio de un bien y el costo promedio de producirlo. Esto se hace estimando el precio de las unidades de productos básicos específicos y restando las estimaciones de los costos unitarios promedio de extracción o costos de cosecha. Estas rentas unitarias luego se multiplican por las cantidades físicas que los países extraen o cosechan para determinar las rentas de cada producto básico como parte del producto interno bruto (PIB).

### Actividad 1

Utilice el conjunto de datos [GDP\\_PETR.xls](#) para crear visualizaciones correspondientes a proporciones, distribuciones y cantidades de la renta petrolera (% of GDP) a nivel mundial. Tome en cuenta las siguientes consideraciones:

- Realice la limpieza de los datos para facilitar la creación de las visualizaciones, por ejemplo, puede *remover* algunas columnas, definir *nombres más adecuados* a las columnas, *adicionar* una columna que represente una suma o promedio de variables.
- Realice una *grafica de línea* donde se compare cuál fue la tendencia de los 5 países con la mayor renta petrolera (% of GDP) en el periodo comprendido del año 1970 al 2020.

Realice un *diagrama acumulativo*, también conocido como *diagrama de líneas apiladas* ó *diagrama de áreas* donde se visualicen los 5 principales países y su renta petrolera. Para cualquier valor NaN establezca de forma predeterminada en *cero*.

Represente el diagrama de áreas anterior como *gráfica de porcentaje de proporciones*.

## Actividad 2

Analice el comportamiento de la renta petrolera en el caso de *México*. Para ello realice un *diagrama de barras* en el periodo comprendido del año 1970 al 2020.

Utilice el método `annotate` de la capa de secuencias de comandos **scripting layer** o la interfaz `pyplot` (**pyplot interface**) para resaltar alguna característica de interés en el comportamiento de la renta petrolera.

Ahora analice el comportamiento de la renta petrolera mexicana considerando los siguientes periodos de gobierno:

- P1: 1970-1976
- P2: 1976-1982
- P3: 1982-1988
- P4: 1988-1994
- P5: 1994-2000
- P6: 2000-2006
- P7: 2006-2012
- P8: 2012-2018
- P9: 2018-2020

Realice un *gráfico circular* considerando los 5 periodos de gobierno con mayor renta petrolera. De acuerdo a las graficas obtenidas, mencione una **conclusion** de los resultados mostrados.

### NOTA:

El **sesgo de informe** (también llamado informe selectivo) afecta qué estudios “salen a la luz” y cuáles no. Los estudios pueden seleccionarse (o no seleccionarse) por varias razones, incluidas agendas personales, falta de conocimiento o métodos de análisis descuidados.

*Este tipo de sesgo es un problema importante en los informes científicos y los informes de ensayos clínicos.*

El sesgo de informe significa que **solo se incluye una selección de resultados en cualquier análisis**, que generalmente cubre solo una fracción de la evidencia relevante. Esto puede conducir a decisiones inapropiadas, desperdicio de recursos e investigaciones futuras equivocadas.