

MIA—V1E4-002 | Data, Big Data and Open Data

Actividad Nro.2

Nombre:

Fecha:

1. De acuerdo a **open Data** y los repositorios que publican datos, buscar datos del último censo de población y vivienda de al menos 2 países de América del Sur y graficar la distribución poblacional entre hombres y mujeres, incluyendo la distribución (hombre/mujer) en Bolivia.

Realice una representación gráfica comparativa entre los países y su distribución (hombre/mujer) y sintetice una conclusión de acuerdo a la comparación de los países analizados.

2. De acuerdo al sitio de datos abiertos Kaggle buscar un conjunto de datos (Dataset) y mediante PySpark o bibliotecas de su preferencia en python, obtener lo siguiente:

Fuente:	(url donde obtiene el dato)
Nro de registros	(Cantidad de registros)
Tipos de estructura	(Estructurado, semi estructurado, no estructurado)
Temática de los datos	(Censo de población y vivienda / Programas y películas de Netflix, etc)
Sintetizar alguna información	(Distribución de edad, Nros desempleados, Distribución de películas y series, etc)
Concluir en un par de líneas la información obtenida	(A partir del síntesis dar una conclusión de la información sintetizada)

Se recomienda que el conjunto de datos, sea mayor a 1.000.000 registros.

Acompañar algunas gráficas realizadas en Python que pueda reflejar el análisis o conclusión de la información sintetizada.

Plazo de entrega, hasta el día martes 21 de agosto.