

## Taller 1

Muchos gobiernos alrededor del mundo han promovido la iniciativa de datos abiertos como un mecanismo de transparencia que habilita a cualquier ciudadano tener acceso a datos de dominio público. De esta forma, los ciudadanos pueden participar en la creación de políticas públicas, hacer seguimiento al gasto público y en general tener visibilidad de variados aspectos de índole público.

Colombia no es la excepción. En el portal de datos abiertos [datos.gov.co](https://datos.gov.co) los ciudadanos pueden acceder a datos oficiales y no oficiales de diferentes categorías, entidades y zonas del país. Adicionalmente, también pueden aportar a la comunidad subiendo nuevos datos o análisis sobre datos ya existentes. Al día de hoy existen casi 6 mil conjuntos de datos, más de 20 mil visualizaciones y 935 entidades.

La Policía Nacional en conjunto con el Ministerio de Defensa Nacional han publicado historiales de delitos para todo el territorio nacional. Actualmente, estos historiales se encuentran repartidos en diferentes conjuntos de datos publicados en el portal de datos abiertos. Los links de acceso a los diferentes conjuntos de datos se encuentran a continuación:

- [Delitos sexuales](#)
- [Violencia intrafamiliar](#)
- [Secuestro](#)
- [Terrorismo](#)
- [Homicidio](#)
- Hurto [1](#) [2](#) [3](#)
- [Lesiones personales](#)
- [Extorsión](#)

Puntos a desarrollar:

1. **[15%]** Seleccione al menos 4 conjuntos de datos y realice un perfilamiento de cada uno. Describa los hallazgos encontrados en cada conjunto de datos. No olvide mencionar su estructura y aspectos relevantes de calidad como campos nulos, departamentos mal escritos, formatos de fechas incorrectos, entre otros. Si no evidencia ningún problema de calidad de datos, también mencionelo.

2. **[20%]** Responda la siguientes preguntas para cada uno de los conjuntos de datos seleccionados:
  - 2.1. **[4%]** ¿Cuáles han sido los departamentos (TOP 3) más afectados en términos de cantidad de delitos cometidos en los últimos 5 años?
  - 2.2. **[4%]** Para los casos en los que aplique, ¿cuál ha sido el arma o medio más común para cometer el delito?
  - 2.3. **[4%]** Para los casos en los que aplique, ¿cómo ha sido la proporción de géneros y grupos etarios que han estado involucrados en este tipo de delito? ¿Han variado con el paso de los años?
  - 2.4. **[4%]** ¿Se evidencia alguna tendencia para cometer dicho delito en algún mes particular del año?
  - 2.5. **[4%]** Para los casos en los que se disponga del detalle del delito o de una descripción, como por ejemplo en delitos sexuales y secuestro, ¿cuáles son las descripciones o modalidades más comunes?
3. **[25%]** A partir de alguno de los conjuntos de datos seleccionados, visualice una serie de tiempo por año y mes que permita comparar la cantidad de delitos cometidos para los departamentos con mayor ocurrencia durante los últimos 5 años. Para que los resultados entre departamentos sean comparables, es importante que normalice las cantidades obtenidas por cantidad de habitantes. En [este](#) archivo puede encontrar la población por departamento para el año 2018. Asuma que la población no ha cambiado con el paso de los años.
4. **[30%]** A partir de los conjuntos de datos seleccionados, construya un único dataset que integre la totalidad de los delitos ocurridos por departamento y municipio. Muestre los valores normalizados por cantidad de habitantes realizando un proceso similar al del punto anterior. En [este](#) archivo puede encontrar proyecciones anuales de las poblaciones por departamento. Utilice la proyección para el año en curso. Considere solamente los municipios con más de 1 millón de habitantes.
5. **[10%]** Grabe un video de máximo 5 minutos en donde muestre el proceso realizado y los hallazgos más importantes de los diferentes análisis. Suba el video a YouTube. Todos los integrantes del equipo deben participar en el video.

**Aspectos a tener en cuenta:**

- Note que los reportes no se encuentran detallados a nivel de hecho particular. Presentan un nivel de agregación a partir de sus diferentes atributos como departamento, municipio, fecha, entre otros, razón por la cual en algunos registros encontrará cantidades superiores a 1. Esta cantidad corresponde a la cantidad de hechos ocurridos para ese conjunto de atributos.
- Asegúrese de eliminar registros repetidos que puedan afectar el análisis. Puede hacer uso de la función [drop\\_duplicates](#).
- De forma similar a SQL, Pandas cuenta con una función para convertir textos a fechas llamada [to\\_datetime](#).
- Para los casos en los que necesita realizar agrupaciones, puede usar las funciones [groupby](#) y [agg](#).
- Para cruzar conjuntos de datos, puede utilizar la función [merge](#).

**Mecanismo de entrega:**

Suba todo el código fuente a un repositorio público de GitHub. Utilice el archivo Readme para listar los integrantes del grupo de trabajo y dar respuesta a las diferentes preguntas planteadas. Adjunte también en este archivo el link del video.

Solo UN integrante del equipo debe formalizar la entrega a través del aula virtual adjuntando la URL del repositorio. Cualquier commit (cambio del código) que se realice posterior a la fecha de entrega, NO será tenido en cuenta al momento de asignar la calificación. NO se aceptarán documentos adicionales en formatos como Word o PDF.