**Ingeniería de Sistemas**

**Computación Científica y Analítica**

**Taller 1**

Enlace de entrega: [aquí](https://aulavirtual.unbosque.edu.co/mod/assign/view.php?id=319016)

Fecha máxima de entrega: **Marzo 8, 2022 11:59 PM**

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 y que de esta forma, si lo desea, pueda aportar al desarrollo de políticas públicas, económicas y de cualquier otra índole, siendo ya no solo una labor exclusiva de los funcionarios públicos.

Colombia no es la excepción. En el portal [datos.gov.co](https://www.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 los datos ya existentes. Al día de hoy existen casi 6 mil conjuntos de datos, más de 12 mil visualizaciones y 960 entidades.

La Policía Nacional ha publicado un histórico desde el 2010 a la fecha de delitos de diferente tipo dividido en 10 reportes:

* [Delitos sexuales](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Delitos-sexuales-Polic-a-Nacional/fpe5-yrmw)
* [Violencia intrafamiliar](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Delito-Violencia-Intrafamiliar-Polic-a-Nac/vuyt-mqpw)
* [Secuestro](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Delito-Secuestro-Polic-a-Nacional/9jhz-f3a7)
* [Terrorismo](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Delito-Terrorismo-Polic-a-Nacional/37p5-impc)
* [Homicidio](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Homicidios-y-Homicidios-accidente-de-tr-ns/ha6j-pa2r)
* Hurto [1](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Hurto-por-Modalidades-Polic-a-Nacional/6sqw-8cg5) [2](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Hurto-por-Modalidades-Polic-a-Nacional/9vha-vh9n) [3](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Hurto-por-Modalidades-Polic-a-Nacional/d4fr-sbn2)
* [Lesiones personales](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Lesiones-Personales-y-en-Accidente-de-Tr-n/72sg-cybi)
* [Extorsión](https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Delito-Extorsi-n-Polic-a-Nacional/cekg-4xti)

Realice un análisis exploratorio sobre al menos 3 de los anteriores reportes y responda la siguientes preguntas para cada uno de los tipos de delitos:

1. ¿Cuáles han sido los departamentos (TOP 3) más afectados a nivel histórico y en el último año?
2. ¿Cuál ha sido históricamente el arma o medio más común para cometer el delito?
3. ¿Cómo es la proporción de géneros y grupos etarios que han estado involucrados en este tipo de delito? ¿Han variado con el paso de los años?
4. ¿Se evidencia alguna tendencia para cometer dicho delito en algún mes particular del año?
5. 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 los delitos, descripciones o modalidades más comunes?
6. (BONO) Para tener una mayor significancia estadística de los resultados obtenidos en el punto 1, es importante que normalice las cantidades obtenidas por cantidad de habitantes en el municipio. En [este](https://drive.google.com/file/d/11eD13quFrHrFC6lkODydOgZZNaZbeVak/view?usp=sharing) CSV puede obtener la población por departamento para el año 2018. Asuma que la población no ha cambiado con el paso de los años.

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 cuál 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](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.drop_duplicates.html) de Pandas.
* De forma similar a SQL, Pandas cuenta con una función para convertir textos a fechas: [to\_datetime](https://pandas.pydata.org/docs/reference/api/pandas.to_datetime.html).
* Para los casos en los que necesita realizar agrupaciones, puede usar la función [groupby](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html).
* Reporte cualquier problema de calidad de datos que evidencie durante su análisis: campos nulos, departamentos mal escritos, formatos de fechas incorrectos, rangos de fechas inexistentes, entre otros. Si no evidencia ningún problema de calidad de datos, también mencionelo.
* Para realizar el bono, necesitará hacer un join de los dos datasets. Puede hacerlo utilizando la función [merge](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.merge.html).
* Evidencie todos los pasos que realizó para el análisis exploratorio de datos, así como para responder las diferentes preguntas planteadas, para cada uno de los conjuntos de datos que seleccionó.

**Mecanismo de entrega:** [aquí](https://datascience-unbosque.github.io/index.html#faq)

**Rúbrica:**

* 80% - Cumplimiento de todos los requerimientos.
* 20% - Calidad y documentación de código, uso de GitHub y contribución equitativa del grupo de trabajo. Video subido a YouTube.