

Estudio y agrupación de clases de homicidios en Colombia en los últimos 8 años.

**Nombre de la organización:** Policía Nacional de Colombia.

**Autores:** Elkin Rafel Pulido Farías – Jorge Duban Bernal Cardenas

## Contenido

Resumen ejecutivo .....	3
Glosario de términos .....	3
Problemática origen del proyecto .....	3
Solución planteada (objetivo del proyecto de analítica).....	3
Evaluación de éxito propuesta .....	4
Cumplimiento del plan del proyecto por etapas.....	4
Descripción cualitativa de los aportes generados por el proyecto .....	4
Informe de analistas de datos .....	5
Orígenes de datos .....	5
Ajustes realizados a los datos.....	5
Agrupaciones generadas el entrenamiento .....	6
Análisis de las agrupaciones.....	6
Plan de mantenimiento del modelo.....	7
Lecciones y recomendaciones.....	7
Se obtuvo una tasa de éxito acorde a la naturaleza del problema .....	8

## Resumen ejecutivo

### Glosario de términos

- Homicidio: es un delito contra la vida humana que se produce cuando una persona mata a otra.
- Arma blanca: Arma constituida por una hoja metálica u otro material de características físicas semejantes, cortante o punzante.
- Arma de fuego: Arma que utiliza una materia explosiva para lanzar proyectiles
- Artefacto incendiario: es un dispositivo explosivo usado frecuentemente en la guerra no convencional o guerra asimétrica, por fuerzas comando, guerrillas y terroristas.
- Minería de datos: es el proceso de hallar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados.
- Agrupación: Conjunto de caracteres que se asocian por un factor común.
- Google Colab: Es un servicio en la nube, que nos provee de una Jupyter Notebook a la que podemos acceder con un navegador web sin importar si «en casa» usamos Windows, Linux o Mac. Permite crear libros en Python 2 o en 3. Tiene preinstaladas las librerías comunes usadas en datascience y la posibilidad de instalar otras según sea necesario.

### Problemática origen del proyecto

Actualmente en Colombia la tasa de homicidios no decae, ya que día a día muchas personas se ven involucradas en este tipo de delitos y las organizaciones encargadas de atender dichos casos, como la Policía Nacional de Colombia, no pueden evitar el infortunio de no encontrar una solución viable o un patrón que ayude a reducir los homicidios en todo el país.

Por otro lado, mucha gente es consciente de los asesinatos causados y la inseguridad en la que se ven envueltos en muchas ocasiones, y aunque están al tanto del problema e incluso detecten lugares donde hay más posibilidades de que se ocasionen un homicidio, no hay opción más que simplemente pasar esto por alto y considerar que los homicidios ocasionados por razones como el robo se pueden volver hasta comunes y creer en que ellos mismos no se verán envueltos en este tipo de delito o que la seguridad social pueda mejorar.

Por tales razones, con el desarrollo del proyecto se quiere reducir la tasa de homicidios en el país impactando en la calidad de vida de los colombianos lo máximo posible en relación con el tema tratado.

### Solución planteada (objetivo del proyecto de analítica)

Gracias a las nuevas tecnologías con las que se puede contar para mejorar la calidad de vida con el desarrollo de software, estudio de datos, entre muchas otras cosas, la solución planteada mediante minería de datos se promueve a identificar 4 clases de homicidios en los últimos 8 años con el propósito de que la organización encargada de la seguridad social (Policía Nacional de Colombia), pueda tomar medidas preventivas como aumentar la seguridad en relación con los asesinatos que sean agrupados; reduciendo así mismo la tasa de homicidios, permitiendo que los colombianos se sientan más seguros y se logre mejorar la calidad de vida.

### Evaluación de éxito propuesta

El resultado exitoso del estudio de los datos de homicidios en los últimos 8 años se determinará si el número de agrupaciones cumplen con el objetivo planteado y efectivamente cada agrupación esta determinada por algunos factores que se distingan de las otras.

### Cumplimiento del plan del proyecto por etapas

Fase	Tiempo estimado (semanas)	Tiempo real (semanas)	Costo estimado	Costo real
Comprensión del negocio	2	3	\$ 0	\$ 0
Comprensión de los datos	3	4	\$ 0	\$ 0
Preparación de los datos	5	3	\$ 0	\$ 0
Modelado	3	4	\$ 0	\$ 0
Evaluación	2	3	\$ 0	\$ 0
Despliegue	1	1	\$ 480.000	\$ 480.000
Total	16	18	\$480.000	\$480.000

### Descripción cualitativa de los aportes generados por el proyecto

Con el desarrollo del proyecto se logró encontrar un patrón en los homicidios mediante 4 agrupaciones que pueden permitir a la empresa tomar un mejor curso de acción en la seguridad social en el país.

Por otro lado, se sufrieron algunos retrasos debido a que en un principio en la declaración de los objetivos se proponía como solución hacer una predicción con los datos proporcionados por la empresa y se trabajó en línea a eso, lamentablemente en la creación de modelos que permitieran lograr lo planteado se notó que la mayoría de datos que contenían los atributos estaban sesgados por lo cual después de un estudio más a fondo se determinó que no había una solución factible que permitiera arreglar los datos y seguir con el objetivo de predicción. Por lo tanto, se propuso una solución que se adaptaba mejor a los datos de estudio y se planteó un nuevo objetivo de agrupamiento de homicidios.

Debido a lo anterior, se permitió un mejor uso de los datos que proporcionó la empresa favoreciendo a la misma, ya que, aunque el sesgo de datos no es manejable por la empresa puesto que los homicidios son determinados por muchos factores, se logró plantear una solución que además de uso de los datos con los que se trabajó se aplica también para los datos de asesinatos futuros.

## Informe de analistas de datos

### Orígenes de datos

Nombre del origen:		Estadística Delictiva
Cantidad de datos:	102305	
Hipótesis relacionada con los objetivos del proyecto	Se identifica 4 clases de homicidios en los últimos 8 años que permiten tomar decisiones para reducir la tasa de homicidios en el año 2019.	
Fuente / url:	<a href="https://www.policia.gov.co/grupo-información-criminalidad/estadistica-delictiva">https://www.policia.gov.co/grupo-información-criminalidad/estadistica-delictiva</a>	
Principales atributos		
Atributo	Tipo	Descripción
Departamento	Varchar	Proporciona el departamento en el que se llevó a cabo el homicidio.
Día	int	Proporciona el día de la semana en el que se llevó a cabo el homicidio.
Hora	int	Proporciona la hora en el que se llevó a cabo el homicidio.
Zona	Varchar	Proporciona la zona en el que se llevó a cabo el homicidio, es decir si es Rural o Urbana.
Modalidad o arma	Varchar	Proporciona el tipo de modalidad con la que se cometió el homicidio, por ejemplo, si fue con un arma de fuego, arma blanca, etc.
Movilidad del Agresor	Varchar	Proporciona el tipo de movilidad que tenía el asesino para cometer el homicidio, por ejemplo, si fue a pie, en bicicleta, conduciendo, etc.
Movilidad de la Víctima	Varchar	Proporciona el tipo de movilidad que tenía la víctima, por ejemplo, si fue a pie, en bicicleta, conduciendo, etc.
Edad de la víctima	int	Proporciona la edad de la víctima.
Sexo de la víctima	int	Proporciona el género de la víctima.

### Ajustes realizados a los datos

Nombre del origen:	
Decisiones sobre la calidad de los datos.	Debido a que los datos no son completamente de calidad para el debido proceso de analítica de datos se modifican los campos necesarios y en relación con el objetivo se extraen datos de el atributo fecha para nuevos atributos.
Descripción de otros ajustes hechos a los datos en general	Se llenan campos vacíos. Se quitan los caracteres como tildes. Se normalizan los datos.

Ajustes particulares	
Atributo	Descripción del ajuste
<b>Departamento</b>	Se quitan tildes y se normalizan asignando a cada departamento un número desde 0 a 31.
<b>Día</b>	Se quitan tildes y se normalizan asignando a cada día un número desde 0 a 6.
<b>Hora</b>	Se normalizan asignando a cada intervalo entre 60 minutos un número desde 0 a 23.
<b>Zona</b>	Se normalizan asignando a cada tipo de zona un numero de 0 a 1.
<b>Modalidad o arma</b>	Se quitan tildes y se normalizan asignando a cada modalidad de arma un número desde 0 a 36.
<b>Movilidad del Agresor</b>	Se quitan tildes y se normalizan asignando a cada movilidad un número desde 0 a 12.
<b>Movilidad de la Víctima</b>	Se quitan tildes y se normalizan asignando a cada movilidad un número desde 0 a 12.
<b>Sexo de la víctima</b>	Se normalizan asignando a cada tipo de género un número de 0 a 1.
<b>Mes</b>	Este atributo se obtiene del atributo fecha.
<b>Día del año</b>	Este atributo se obtiene del atributo fecha.
<b>Semana del año</b>	Este atributo se obtiene del atributo fecha.

### Agrupaciones generadas el entrenamiento

Agrupaciones	Número de homicidios
01	40119
02	15284
03	33968
04	12934

### Análisis de las agrupaciones

- **Agrupación 01: Relación entre Zona Urbana y los primeros 7 meses:**

En esta primera agrupación los homicidios se relacionan en base a la zona urbana y lo primeros 7 meses (enero, febrero, marzo, abril, mayo, junio y julio), cuenta con la mayor cantidad de homicidios en las agrupaciones (40119). Solo podemos deducir que es debido a que como en las zonas urbanas, que es donde hay mayor cantidad de población, están mas activas por que se presentan mayores eventos como lo es el inicio de año nuevo, la celebración de los reyes magos, semana santa, y en mitad de año las perdonas se ven más influenciadas por recesos de trabajo u otros, hay mas movimiento y muchos de los homicidios suceden a causa de atracos o relacionados.

- **Agrupación 02: Relación entre Zona Rural y los primeros 6 meses:**

En esta segunda agrupación los homicidios se relacionan en base a la zona rural y lo primeros 6 meses (enero, febrero, marzo, abril, mayo, junio), la cantidad de homicidios en esta agrupación es de 15284. Al igual que el anterior, deducimos que hay más actividad por

que se presentan mayores eventos como lo es el inicio de año nuevo, la celebración de los reyes magos, semana santa, y en mitad de año las perdonas se ven más influenciadas por recesos de trabajo u otros. A comparación de la primera agrupación estos homicidios son hechos en las zonas rurales del país, por eso su cantidad de víctimas es mucho menor que en las zonas urbanas, pues hay menos población.

- **Agrupación 03: Relación entre Zona Urbana y los últimos 6 meses:**

En esta tercera agrupación los homicidios se relacionan en base a la zona urbana y los últimos 6 meses (julio, agosto, septiembre, octubre, noviembre y diciembre), cuenta con una cantidad de homicidios de 33968. En este deducimos que como ya mencionamos, en las zonas urbanas es donde hay mayor población, pero acá ya hay más reducción, pues en este tiempo las actividades más importantes hacen que haya más actividad son el día de las brujas y la celebración de fin de año, en dichas ocasiones lo más influyente es el fin de año en diciembre ya que es muy conocido que la cantidad de muertes en dicho mes es más alta.

- **Agrupación 04: Relación entre Zona Rural y los últimos 6 meses:**

En esta cuarta agrupación los homicidios se relacionan en base a la zona rural y los últimos 6 meses (julio, agosto, septiembre, octubre, noviembre y diciembre), cuenta con la menor cantidad de homicidios en las agrupaciones (12934). Al igual que la anterior agrupación lo que más afecta son las mismas mencionadas actividades solo que esta agrupación tiene menos homicidios por ser donde hay menos población.

## Plan de mantenimiento del modelo

En el plan de mantenimiento se llevará a cabo cada año. Esto es una vez se tenga los datos de los homicidios ocurridos en el año que ya ha pasado, de tal manera que a estos datos se le realizara el proceso de la etapa de preparación de los datos para los atributos con los que trabaja el modelo para luego agregarlos a los datos que ya se tienen y ejecutar dicho modelo para identificar nuevamente las 4 clases de homicidios.

## Lecciones y recomendaciones

Responda a las siguientes preguntas

- **¿Existen formas de simplificar o mejorar alguna de las fases u operación particular?, ¿cuáles?**

En un principio, como se habían planteado objetivos diferentes al actual, notamos que mientras comenzábamos a prepararnos para hacer los respectivos modelos para una predicción, algunos de los objetivos planteados en la primera fase (compresión del negocio) no eran muy específicos o no se podían utilizar para hacer un respectivo modelo idóneo y, por lo tanto, teniendo lo anterior presente, los homicidios para el año 2019. Por otro lado, se llegó a la conclusión de se debe hacer un mejor estudio y revisión de los datos para determinar si algunos de estos están sesgados en el entendimiento de los datos.

- **¿Cuáles fueron los fallos o errores de cometieron?, ¿Cómo se pueden evitar la próxima vez?**

Al igual que en la pregunta anterior, el principal fallo fue no identificar que los datos estaban muy sesgados ya que esto solo se notó cuando estábamos creando los modelos con el propósito de hacer una predicción y debido a que no se pudo generar una solución el proyecto tuvo que plantear un nuevo objetivo y dirigirse por el método de agrupamiento, pero de igual forma, apuntado a reducir la tasa de homicidios para el año 2019.

- **Alguna recomendación para el uso o mejoramiento de lo generado en el proyecto.**

Se recomienda que los análisis sean lo más precisos posibles de acuerdo a los resultados de los agrupamientos.

- **Alguna recomendación para el cliente sobre la forma de abordar el proyecto.**

En este caso, como los análisis son algo tediosos, se recomienda activar el filtro en las columnas que corresponde a cada atributo. Luego de esto, fijarse en aquellos atributos que en sus opciones de filtro muestran muy pocas opciones (en general de 1 a 6) ya que la mayoría de atributos siempre tendrán mas de 6 opciones de filtro; basado en esto se puede identificar el agrupamiento por el que esta hecho la clase de homicidio.

- **Propuesta de posible nuevo proyecto en la misma línea.**

Si la organización proporciona datos que no abarquen tantas cualidades en un solo atributo y determine que los datos cuenten con la información necesaria, evitando el sesgo de los mismos, se puede crear un sistema que prediga los homicidios en el próximo año correspondiente.

## Se obtuvo una tasa de éxito acorde a la naturaleza del problema

El modelo cumple con el objetivo planteado y se identifican las 4 clases de homicidios que se pueden detectar en los últimos 8 años.