



# DELITOS SEXUALES EN COLOMBIA

Proyecto Minería de datos

Profesora: Ingeniera Alexandra Pomares

Presentado por: William Forero, Nicolás Méndez, Ailín Rojas y Santiago Salamanca

## *Prefacio*

El presente documento está dirigido a todas aquellas personas que deseen conocer el proceso de minería de datos realizado por el grupo de trabajo, aplicando la metodología CRISP-DM al conjunto de datos seleccionados referentes al tema de Delitos Sexuales en Colombia. Con este documento se busca transmitir el conocimiento adquirido por el grupo durante la experiencia, evidenciando cómo se llevó a cabo cada fase de la metodología aplicada.

## Tabla de contenido

Prefacio.....	2
Tabla de contenido .....	3
Tabla de tablas .....	5
Tabla de ilustraciones .....	5
1. Introducción .....	6
2. Objetivo del proyecto.....	6
3. Infografía.....	7
4. Modificaciones en los datos .....	8
5. Vistas Minables .....	11
5.1. Clustering .....	11
5.1.1. Vista Bogotá.....	11
5.1.2. Vista Medellín.....	2
5.1.3. Vista Cali .....	2
5.1.4. Vista Otras Ciudades .....	2
5.1.5. Vista Delito contra adultos.....	2
5.2. Reglas de Asociación .....	2
5.2.1. Vista con desempleo y educación.....	2
6. Técnicas y algoritmos seleccionados para realizar el análisis .....	2
6.1. Clustering .....	2
6.2. Reglas de Asociación .....	3
6.3. Árboles de decisión .....	<b>¡Error! Marcador no definido.</b>
7. Parametrización de los algoritmos de minería de datos elegidos .....	3
7.1. CLUSTERING.....	3
7.2. Reglas de asociación .....	4
8. Resultados obtenidos. ....	5
8.1. Clustering .....	5
8.1.1. Vista Ciudad - Bogotá .....	5
8.1.2. Vista Ciudad – Medellín .....	9
8.1.3. Vista Ciudad – Cali .....	13
8.1.4. Vista otras ciudades .....	17
8.2. Reglas de asociación .....	20
9. Evaluación de calificación .....	20
9.1. Clustering .....	20
9.1.1. Vista Ciudad - Bogotá .....	21
9.1.2. Vista Ciudad - Medellín.....	21
9.1.3. Vista Ciudad - Cali .....	21
9.1.4. Vista otras ciudades .....	22

9.1.5.	Vista delito contra adultos .....	22
9.2.	Reglas de Asociación .....	23
9.3.	Árboles de decisión .....	23
9.3.1.	Vista Ciudad de Bogotá: .....	23
9.3.2.	Vista Ciudad Medellín: .....	23
9.3.3.	Vista Ciudad Cali: .....	24
9.3.4.	Vista otras ciudades: .....	24
9.3.5.	Vista delito contra adultos: .....	24
10.	VISUALIZACIÓN DE RESULTADOS OBTENIDOS .....	24
10.1.	Clustering.....	24
10.1.1.	Vista Ciudad de Bogotá: .....	25
10.1.2.	Vista Ciudad Medellín: .....	25
10.2.	Vista Ciudad Cali: .....	25
10.2.1.	Vista otras ciudades .....	26
10.2.2.	Vista delito contra adultos.....	26
10.3.	Reglas de Asociación .....	27
11.	Interpretación de resultados en términos de usuario o negocio. ....	27
12.	MÉTODO PARA HACER REPLICABLE LA APLICACIÓN DE LAS TÉCNICAS RECIENTES A DATOS EN LA ORGANIZACIÓN .....	28
12.1.	Integración con las entradas.....	29
12.2.	Integración con las salidas .....	29
12.3.	Recursos necesarios.....	29
12.4.	Material necesario .....	30
12.5.	Personal necesario y formación .....	30
13.	CONCLUSIONES Y TRABAJO FUTURO .....	31
13.1.	Conclusiones .....	31
13.2.	Trabajo futuro .....	31
14.	RENDIMIENTO TRABAJO EN GRUPO .....	32
14.1.	Porcentaje de participación.....	32
15.	REFERENCIAS .....	33

## Tabla de tablas

Tabla 1 Parametros para Clustering .....	3
Tabla 2 Parámetros para DataTools .....	4
Tabla 3 Parámetros reglas de asociación .....	5
Tabla 4 Arboles sobre clúster Bogotá .....	5
Tabla 5 Características de los clusters de Bogotá.....	6
Tabla 6 Arboles sobre clúster Medellín .....	9
Tabla 7 Características de los clusters de Medellín.....	10
Tabla 8 Características de los clusters de Cali.....	14
Tabla 9 Características de los clusters de otra ciudad .....	17
Tabla 10 Evaluacion Vista Bogotá.....	21
Tabla 11 Evaluacion Vista Medellín.....	21
Tabla 12 Evaluacion Vista Cali.....	22
Tabla 13 Evaluacion Vista Otras ciudades .....	22
Tabla 14 Evaluación Vista delito contra adultos .....	22

## Tabla de ilustraciones

Ilustración 1 Infografía Delitos Sexuales .....	7
Ilustración 2 Matriz de confusión Bogotá.....	23
Ilustración 3 Matriz de confusión Medellín.....	24
Ilustración 4 Matriz de confusión Cali.....	24
Ilustración 5 Ilustracion Cluster Bogotá .....	25
Ilustración 6 Ilustración Cluster Medellín .....	25
Ilustración 7 Ilustración Cluster Otras ciudades.....	26
Ilustración 8 Ilustración Cluster Cali .....	26
Ilustración 9 Clusters Educación y Desempleo.....	27
Ilustración 10 Reglas de asociación .....	27

## 1. Introducción

El siguiente documento contiene el resultado de nuevas ejecuciones de los procesos de minería, las cuales se realizaron con el fin de refinar las técnicas de minería empleadas, con base en los resultados obtenidos en la previa entrega.

Asimismo, contiene la descripción de cada uno de los aspectos mejorados para refinar las técnicas. entre estos se encuentran: la definición de vistas minables; en las cuales se cambió el nivel de granularidad para obtener resultados más precisos, la parametrización de los algoritmos de minería de datos elegidos; se cambiaron las parametrizaciones en las distintas herramientas utilizadas (*SQL Server Data tools* y *RapidMiner*).

Con el objetivo de obtener resultados de utilidad para la organización en cada una de las técnicas empleadas, se realizó un cruce de información con las bases de datos referentes a: La tasa de desempleo del año 2014 y Estadísticas en educación básica por municipio.

## 2. Objetivo del proyecto

Obtener información de los datos que permitan perfilar el contexto en el que ocurre un delito sexual o la víctima de dicho delito.

## 3. Infografía

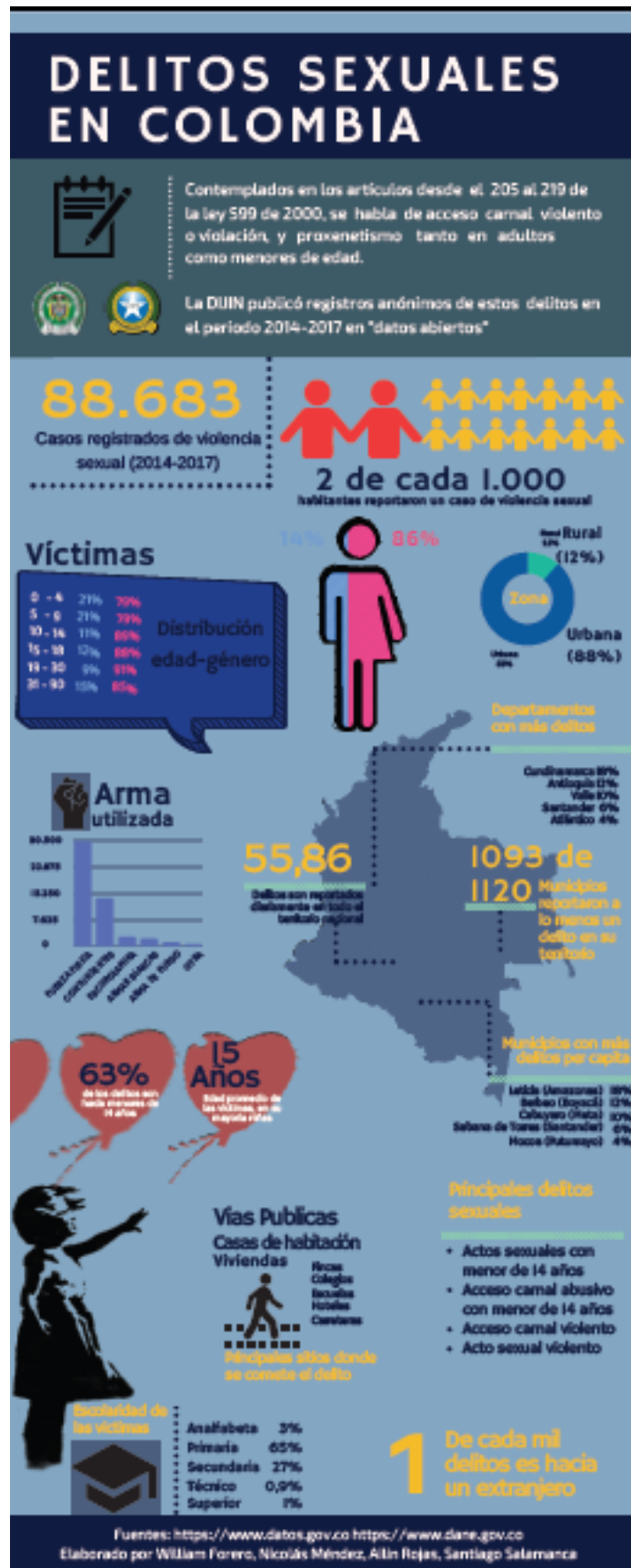


Ilustración 1 Infografía Delitos Sexuales

## 4. *Modificaciones en los datos*

La base de datos utilizada para la segunda entrega no contenía los datos necesarios para obtener información coherente y desconocida por el negocio, por lo cual se optó por buscar nuevas bases de datos que sirvieran para formar un conjunto útil y descubrir nueva información a partir de ellos. Además, se realizaron los siguientes cambios:

- Se creó el atributo “Trimestre”, para poder determinar el trimestre en el que ocurría el delito sexual. La creación de este también nos permitió usarlo para realizar una operación “Join” con otra base de datos.

Para obtener mejores resultados después de aplicar las técnicas apropiadas a los datos, se decidió unir la base de datos que se tiene actualmente, con dos bases de datos, una dedicada al desempleo en Colombia que contiene los datos de 13 municipios (Bogotá, Cali, Medellín, Barranquilla, Cúcuta, Villavicencio, Cartagena, Pereira, Bucaramanga, Pasto, Ibagué, Montería, Manizales), y otra que contiene los datos estadísticos de la educación básica:

La base de datos del desempleo en Colombia contiene los siguientes atributos:

- **Año:** Indica el año de donde se consiguieron los datos del desempleo, en este caso solo se tienen los del 2014
- **Trimestre:** Indica el trimestre del año en donde se obtuvo el porcentaje del desempleo
- **Municipio:** Indica el nombre del municipio referente a una tasa de desempleo
- **Sexo:** Indica el género de la población del municipio
- **Nivel educativo:** Referencia al nivel de educación de un grupo de personas
- **Participación global:** Porcentaje de personas de la población que recibe ingresos en relación con la población con edad de trabajar.
- **Desempleo total:** Tasa de desempleo en relación con toda la población capaz de trabajar
- **Desempleo abierto:** Porcentaje de personas desocupadas con respecto a la unión de las personas empleadas y desempleadas.
- **Desempleo oculto:** Porcentaje de personas que no estaban trabajando y no buscaron trabajo en el momento de recolección de datos porque no creen que es posible encontrarlo, por ciertas circunstancias.
- **Tasa ocupación:** Porcentaje de personas empleadas dentro del rango de edad de 14 y 64 años en relación con el total de las personas comprendidas en el mismo rango mencionado anteriormente.
- **Cesantes:** Porcentaje de personas que fueron desempleadas luego de que terminará su contrato laboral.
- **Aspirantes:** Porcentaje de población desocupada que está buscando trabajo por primera vez.

Los cambios realizados a los atributos nombrados anteriormente o a los datos en general, fueron los siguientes:

- Se utilizó la operación Map con los atributos "Municipio", "Sexo", "Nivel educativo" para que sus datos coincidan con los que se encuentran en la base de datos actual. Se



realizó este cambio ya que, en el caso de los nombres de las ciudades, en la base de datos actual eran diferentes a los que estaban en la del Desempleo.

- Se eliminaron los atributos “Año”, “Nivel educativo”, “Sexo” y “Trimestre” después de darles un uso al unirlos con la base de datos actual con los delitos sexuales. Además, se eliminó el atributo “Trimestre” de la base de datos original después de que se usó para realizar la operación “Join”.

La última base de datos que se va a usar, referente a la educación básica en Colombia tiene los siguientes atributos:

- **Año:** Año que representa el momento de cuando se obtuvieron los datos.
- **Código municipio:** Código único para representar un municipio.
- **Municipio:** Nombre del municipio de donde se registraron los datos.
- **Código departamento:** Código único para representar un departamento
- **Departamento:** Nombre del departamento al que pertenece el municipio.
- **Población:** Número de población de estudiantes del municipio al que esté relacionado
- **Tasa de matriculación:** Porcentaje de personas matriculadas según un total de personas en cierta edad determinada para realizar sus estudios hasta la educación media.
- **Cobertura neta:** Porcentaje de estudiantes matriculados en el municipio, sin contar a los que sobrepasan la edad correspondiente de cada grado.
- **Cobertura neta transición:** Porcentaje de estudiantes matriculados en el nivel de transición en el municipio, sin contar a los que sobrepasan la edad correspondiente para cursar en dicho grado.
- **Cobertura neta primaria:** Porcentaje de estudiantes matriculados en el nivel de primaria en el municipio, sin contar a los que sobrepasan la edad correspondiente para cursar en dicho grado.
- **Cobertura neta secundaria:** Porcentaje de estudiantes matriculados en el nivel de secundaria en el municipio, sin contar a los que sobrepasan la edad correspondiente para cursar en dicho grado.
- **Cobertura neta media:** Porcentaje de estudiantes matriculados en la educación media en el municipio, sin contar a los que sobrepasan la edad correspondiente para cursar en dicho grado.
- **Deserción:** Porcentaje de deserción en general (En los cursos de transición, primaria, secundaria y media) de los estudiantes en los institutos del municipio al que estén relacionados.
- **Deserción transición:** Porcentaje de deserción de los estudiantes de transición en los institutos del municipio al que estén relacionados.
- **Deserción primaria:** Porcentaje de deserción de los estudiantes de primaria en los institutos del municipio al que estén relacionados.
- **Deserción secundaria:** Porcentaje de deserción de los estudiantes de secundaria en los institutos del municipio al que estén relacionados.
- **Deserción media:** Porcentaje de deserción de los estudiantes de educación media en los institutos del municipio al que estén relacionados.



- **Aprobación:** Porcentaje de aprobación en general (En los cursos de transición, primaria, secundaria y media) de los estudiantes en los institutos del municipio al que estén relacionados.
- **Aprobación transición:** Porcentaje de aprobación de los estudiantes de transición en los institutos del municipio al que estén relacionados.
- **Aprobación primaria:** Porcentaje de aprobación de los estudiantes de primaria en los institutos del municipio al que estén relacionados.
- **Aprobación secundaria:** Porcentaje de aprobación de los estudiantes de secundaria en los institutos del municipio al que estén relacionados.
- **Aprobación media:** Porcentaje de aprobación de los estudiantes de educación media en los institutos del municipio al que estén relacionados.
- **Reprobación:** Porcentaje de reprobación en general (En los cursos de transición, primaria, secundaria y media) de los estudiantes en los institutos del municipio al que estén relacionados.
- **Reprobación transición:** Porcentaje de reprobación de los estudiantes de transición en los institutos del municipio al que estén relacionados.
- **Reprobación primaria:** Porcentaje de reprobación de los estudiantes de primaria en los institutos del municipio al que estén relacionados.
- **Reprobación secundaria:** Porcentaje de reprobación de los estudiantes de secundaria en los institutos del municipio al que estén relacionados.
- **Reprobación media:** Porcentaje de reprobación de los estudiantes de educación media en los institutos del municipio al que estén relacionados.
- **Repitencia:** Porcentaje de repitencia de un curso, sin importar cuál sea este, en general (En los cursos de transición, primaria, secundaria y media) de los estudiantes en los institutos del municipio al que estén relacionados.
- **Repitencia transición:** Porcentaje de repitencia de un curso, sin importar cuál sea este, de los estudiantes de transición en los institutos del municipio al que estén relacionados.
- **Repitencia primaria:** Porcentaje de repitencia de un curso, sin importar cuál sea este, de los estudiantes de primaria en los institutos del municipio al que estén relacionados.
- **Repitencia secundaria:** Porcentaje de repitencia de un curso, sin importar cuál sea este, de los estudiantes de secundaria en los institutos del municipio al que estén relacionados.
- **Repitencia media:** Porcentaje de repitencia de un curso, sin importar cuál sea este, de los estudiantes de educación media en los institutos del municipio al que estén relacionados.

Los cambios realizados a los atributos nombrados anteriormente o a los datos en general para que se lograra unir a la base de datos actual fueron los siguientes:

- Se realizó un filtro al Atributo “Año” para utilizar solamente los referentes al año 2014
- Se creó el atributo “percapitaEstudiantes”, el cual describe cual es la proporción de personas que pueden estudiar del total que se encuentra en la base de datos.
- Se eliminaron un total de 28 atributos, contando los que ya se tenían antes de la unión con otras bases de datos. Estos fueron:

- *APROBACIÓN*
- *AÑO*
- *Año*
- *COBERTURA\_BRUTA*
- *COBERTURA\_NETA*
- *COBERTURA\_NETA\_MEDIA*
- *COBERTURA\_NETA\_PRIMARIA*
- *COBERTURA\_NETA\_SECUNDARIA*
- *COBERTURA\_NETA\_TRANSICIÓN*
- *CÓDIGO\_DEPARTAMENTO*
- *CÓDIGO\_ETC*
- *DEPARTAMENTO*
- *DESERCIÓN*
- *ETC*
- *Hora*
- *HoraDia*
- *MUNICIPIO*
- *Móvil Agresor*
- *Móvil Victima*
- *NumDia*
- *POBLACIÓN\_5\_16*
- *REPITENCIA*
- *REPROBACIÓN*
- *SEDES\_CONECTADAS\_A\_INTERNET*
- *SituacionLaboral*
- *TAMAÑO\_PROMEDIO\_DE\_GRUPO*
- *TASA\_MATRICULACIÓN\_5\_16*
- *count(Código DANE)*

Se eliminaron dichos atributos porque se consideró que no tienen uso en las técnicas de minería que se van a utilizar, ya sea por la poca diversidad en los datos o porque estos no tienen importancia con el objetivo de la organización.

## 5. Vistas Minables

### 5.1. Clustering

Para aplicar la técnica de Clustering, se utilizaron las herramientas RapidMiner y Visual Studio - SQL Server Data Tools, en el caso de RapidMiner antes de aplicar el algoritmo k-means para Clustering, se numerizaron los atributos no ordinales mediante la numerización 1 a n, y los ordinales se numerizaron a través de la numerización 1 a 1. Por otro lado, SQL Server Data Tools al aplicar Clustering, realiza por debajo la modificación de los parámetros discretos para que estos puedan ser tomados y procesados por el algoritmo k-means para generar los distintos clusters de la vista otorgada.

Se realizaron 5 vistas minables diferentes para aplicar Clustering, 1 de las vistas generadas se enfoca en el caso atípico de los delitos contra adultos, con el fin de obtener información sobre las características de estos casos particulares. Además, se generaron 3 vistas particulares sobre las ciudades en las que más ocurren delitos sexuales con lo cual se busca caracterizar y comprender la fenomenología del crimen en estas ciudades.

En efecto, se generaron diferentes modelos con los cuales se pueden analizar y diferenciar los resultados, para llegar a una mejor conclusión.

Las vistas que se generaron fueron:

#### 5.1.1. Vista Bogotá

Esta vista contiene únicamente los registros referentes a la ciudad de Bogotá. Con esta vista se pretenden encontrar las características del contexto en el que ocurren estos delitos en esta ciudad, ya que, es en la que más suceden estos crímenes.

## 5.1.2. *Vista Medellín*

Esta vista contiene únicamente los registros referentes a la ciudad de Medellín. Con esta vista se pretenden encontrar las características del contexto en el que ocurren estos delitos en esta ciudad.

## 5.1.3. *Vista Cali*

Esta vista contiene únicamente los registros referentes a la ciudad de Cali. Con esta vista se pretenden encontrar las características del contexto en el que ocurren estos delitos en esta ciudad.

## 5.1.4. *Vista Otras Ciudades*

Esta vista contiene los registros referentes a las ciudades de: Barranquilla, Cartagena, Villavicencio, Pereira, Bucaramanga, Ibagué, Cúcuta, Pasto, Montería y Manizales. Las cuales corresponden a las ciudades restantes, después de quitar las más peligrosas. Esta vista se realizó con el fin de perfilar a las víctimas y delitos en estas ciudades, para después comparar los resultados obtenidos con los de las vistas previamente mencionadas y, asimismo, descubrir las diferencias y comprender la fenomenología de esos crímenes en las distintas ciudades.

## 5.1.5. *Vista Delito contra adultos*

Esta última vista contiene únicamente los datos de las víctimas de los delitos sexuales contra adultos. Se pretende descubrir las características que tienen estas víctimas, y el contexto en este tipo de delito.

## 5.2. *Reglas de Asociación*

### 1.1.1. *Vista con desempleo y educación*

Para crear las Reglas de Asociación se utilizará Rapid Miner. Rapid Miner hace uso de dos algoritmos, FP-Growth para generar los itemset más frecuentes y un segundo algoritmo para crear las reglas y calcular el soporte, la confianza el soporte y la mejora. Se utilizaron las vistas generadas a partir del cruce con la base de datos general, con una base de desempleo y otra de educación.

## 6. *Técnicas y algoritmos seleccionados para realizar el análisis*

### 6.1. *Clustering*

- **Algoritmo:** k-means
- **Métricas de validación interna:** Índice Davies Bouldin
- **Métricas de validación externa:** Cross Validation (Árboles sobre clusters)

## 6.2. Reglas de Asociación

**Algoritmo:** FP-Growth

**Métricas de validación:** El índice de soporte generado para cada itemset

**Algoritmo:** Generador Reglas de asociación

**Métricas de validación:** El índice de confianza generado por cada regla

## 7. Parametrización de los algoritmos de minería de datos elegidos

### 7.1. CLUSTERING

- Parámetros generales aplicados a todas las vistas - RapidMiner

Tabla 1 Parametros para Clustering

Operación	Operador	Parámetros
Normalización	Normalize	<ul style="list-style-type: none"> <li>• method: range transformation</li> <li>• min: 0.0</li> <li>• max: 1.0</li> </ul>
Numerización 1 a n	Nominal to Numerical	<ul style="list-style-type: none"> <li>• attribute filter type: all</li> <li>• coding type: dummy coding</li> <li>• unexpected value handling: all 0 and warning</li> </ul>
Numerización 1 a 1	Map	<ul style="list-style-type: none"> <li>• Attribute filter type: single</li> <li>• attribute: Escolaridad</li> </ul>
Validación del modelo	Distance Cluster Performance	<ul style="list-style-type: none"> <li>• main criterion: Davies Bouldin</li> </ul>
Algoritmo de clustering	<b>Clustering(k-means)</b>	<ul style="list-style-type: none"> <li>• k: (Varia según la vista)</li> <li>• max runs: 10</li> <li>• measure types: NumericalMeasures</li> <li>• numerical measure: EuclideanDistance</li> <li>• max optimizations steps: 100</li> </ul>

- **Parámetros generales aplicados a todas las vistas - Visual Studio SQL Server Data Tools**

Tabla 2 Parámetros para DataTools

Parámetro	Descripción	Valor	Predeterminado	Rango
Cluster_Count	Especifica la cantidad aproximada de clústeres que generará el algoritmo. Se estableció el valor en 0, Lo cual hace que el algoritmo use la heurística para determinar mejor el número de clústeres que se deben construir.	0	10	[0,...)
Cluster_Seed	Especifica el número inicial utilizado para generar clústeres aleatoriamente para la etapa inicial de la creación del modelo.		0	[0,...)
Clustering_Method	El método de Clustering que utiliza el algoritmo puede ser: EM escalable (1), EM no escalable (2), K-means escalables (3) o K-means no escalables (4). Se estableció el valor en 4, para que trabajara con el total de los registros.	4	1	1,2,3,4
Maximum_Input_Attributes	Especifica la cantidad máxima de atributos de entrada que el algoritmo puede manejar antes de invocar la selección de características. Establecer este valor en 0 especifica que no hay un número máximo de atributos.		255	[0,65535]
Maximum_States	Especifica la cantidad máxima de estados de atributo que admite el algoritmo.		100	0,[2,65535]
Minimum_Support	Este parámetro especifica el número mínimo de casos en cada grupo.		1	(0,...)
Modelling_Cardinality	Este parámetro especifica la cantidad de modelos de muestra construidos durante el proceso de agrupamiento.		10	[1,50]
Sample_Size	Especifica el número de casos que el algoritmo usa en cada pasada si el parámetro CLUSTERING_METHOD se establece en uno de los métodos de clúster escalables.	0	50000	0,[100,...)
Stopping_Tolerance	Especifica el valor utilizado para determinar cuándo se alcanza la convergencia y el algoritmo finaliza la construcción del modelo.		10	[0,...)

## 7.2. Reglas de asociación

### 7.2.1. FP-Growth

Min Support: 0.1

Max Items: 3

### 7.2.2. Generador Reglas de Asociación

Min Confidence: 0.7

Tabla 3 Parámetros reglas de asociación

Parámetro	Valor	Predeterminado	Rango
MAXIMM_ITEMSET_COUNT	0	200000	[0,...)
MAXIMUMI_ITEMSET_SIZE		3	[0,500]
MAXIMUM_SUPPORT		1.0	(0.0,...)
MINIMUM_IMPORTANCE		-999999999	(...,...)
MINIMUM_ITEMSET_SIZE		1	[1,500]
MINIMUM_PROBABILITY		0.4	[0.1,1.0]
MINIMUM_SUPPORT		0.0	[0.0,...)

## 8. Resultados obtenidos.

### 8.1. Clustering

Para la ejecución del algoritmo de clustering con las diferentes vistas, se omitieron los atributos que tienen una similitud en valores dentro de todos los clusters. Se dejaron únicamente los que se creían que podían influir en la obtención de mejores resultados. Los atributos que se removieron son:

- **Para las 3 vistas de las ciudades de Bogotá, Medellín y Cali**

Aprobación Transición, Primaria, Media y Secundaria, Barrio, Cobertura Bruta Transición, Primaria, Media y Secundaria, Código DANE, Delito, Departamento, Deserción Transición, Primaria, Media y Secundaria, Día Semana, Municipio, Nacionalidad, Repitencia, Reprobación, Zona.

- **Para la vista de las otras ciudades**

Día Semana, Zona, Código DANE, Delito, Departamento.

- **Para la vista de delito contra adultos**

Barrio, Código DANE, Nacionalidad, País de nacimiento, Tipo de Delito, Zona.

#### 8.1.1. Vista Ciudad - Bogotá

Tabla 4 Árboles sobre clúster Bogotá

Cluster	Descripción
Cluster 1	participacionglobal $\leq$ 0.614 and Día = Domingo = false

## DELITOS SEXUALES EN COLOMBIA



<b>Clúster 2</b>	participacionglobal $\leq$ 0.614
<b>Clúster 3</b>	aspirantes $\leq$ 0.207
<b>Clúster 4</b>	participacionglobal $\leq$ 0.614 and TipoDelito = DELITO CONTRA MENOR DE 14 AÑOS = true and Día = Domingo = false
<b>Clúster 5</b>	Sexo = FEMENINO = false and Clase de sitio = VIAS PUBLICAS = true
<b>Clúster 6</b>	Barrio = PARQUE NORTE C-4 = true and Escolaridad = PRIMARIA = false TipoDelito = DELITO CONTRA ADULTO
<b>Clúster 7</b>	desempleoabierto $>$ 0.565 and Clase de empleado = ESTUDIANTE = true Clase de sitio = VIAS PUBLICAS
<b>Clúster 8</b>	COLOMBIANO/A = true and participacionglobal $\leq$ 0.522

*Tabla 5 Características de los clusters de Bogotá*

Cluster	Características		
Cluster_1	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILIT Y</b>
	Cesantes	0,078473169	1
	Desempleoabierto	0,075504895	1
	Desempleooculto	0,006063736	1
	Desempleototal	0,081165694	1
	Edad Correcto	15	1
	Tasaocupacion	0,554804086	1
	Sexo	FEMENINO	0,8595368
	Tipo Sitio	AFUERA	0,8286575
	Arma Empleada	SIN EMPLEO DE ARMAS	0,794389
	Clase De Empleado	EMPLEADO PARTICULAR	0,7899642
	Escolaridad	PRIMARIA	0,7686876
	Aspirantes	0	0,7104124
Cluster_2	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILIT Y</b>
	Aspirantes	0	1
	Cesantes	0,076	1
	Desempleoabierto	0,073	1
	Desempleooculto	0,004	1
	Desempleototal	0,076	1
	Edad Correcto	8-9	1
	Escolaridad	PRIMARIA	1
	Participacionglobal	0,552	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,51	1
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	1
	Arma Empleada	SIN EMPLEO DE ARMAS	0,9208048
	Tipo Sitio	AFUERA	0,8672945
	Clase De Empleado	EMPLEADO PARTICULAR	0,8416096



# DELITOS SEXUALES EN COLOMBIA



Cluster_3			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILIT Y
	Aspirantes	0	1
	Cesantes	0,076	1
	Desempleoabierto	0,073	1
	Desempleooculto	0,004	1
	Desempleototal	0,076	1
	Edad Correcto	9	1
	Escolaridad	PRIMARIA	1
	Participacionglobal	0,552	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,51	1
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	1
	Clase De Empleado	EMPLEADO PARTICULAR	0,998045
	Tipo Sitio	AFUERA	0,9912023
	Arma Empleada	SIN EMPLEO DE ARMAS	0,9824047
	Nacionalidad	EXTERIOR	0,8035191
Cluster_4			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY
	Aspirantes	0,011	1
	Cesantes	0,105	1
	Desempleoabierto	0,109	1
	Desempleooculto	0,007	1
	Desempleototal	0,116	1
	Edad Correcto	23-24	1
	Escolaridad	SECUNDARIA	1
	Participacionglobal	0,751	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,664	1
	Nacionalidad	COLOMBIANO/A	0,7811159
	Tipo Delito	DELITO CONTRA ADULTO	0,6609442
	Tipo Sitio	AFUERA	0,6330472
Cluster_5			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILIT Y
	Cesantes	0,1407205	1
	Desempleoabierto	0,1418012	1
	Desempleooculto	0,0144494	1
	Desempleototal	0,1562687	1
	Edad Correcto	22-23	1
	Tasaocupacion	0,5916735	1
	Sexo	FEMENINO	0,9819277
	Aspirantes	0,017	0,9012048
	Escolaridad	SECUNDARIA	0,8807229
	Participacionglobal	0,729	0,8807229
	Nacionalidad	COLOMBIANO/A	0,8096386
Cluster_6			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY
	Aspirantes	0,002	1
	Cesantes	0,043	1

## DELITOS SEXUALES EN COLOMBIA



	Desempleoabierto	0,034	1
	Desempleooculto	0,01	1
	Desempleototal	0,045	1
	Edad Correcto	11-12	1
	Escolaridad	PRIMARIA	1
	Participacionglobal	0,74	1
	Sexo	MASCULINO	1
	Tasaocupacion	0,707	1
	Arma Empleada	SIN EMPLEO DE ARMAS	0,9354839
	Tipo Sitio	AFUERA	0,9119107
	Clase De Empleado	EMPLEADO PARTICULAR	0,8970223
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	0,7593052
Cluster_7			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY
	Aspirantes	0,011	1
	Cesantes	0,105	1
	Desempleoabierto	0,109	1
	Desempleooculto	0,007	1
	Desempleototal	0,116	1
	Edad Correcto	12-13	1
	Escolaridad	SECUNDARIA	1
	Participacionglobal	0,751	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,664	1
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	1
	Clase De Empleado	ESTUDIANTE	0,9285714
	Nacionalidad	COLOMBIANO/A	0,9166667
	Tipo Sitio	SITIO BAJO TECHO	0,6547619
Arma Empleada	SIN EMPLEO DE ARMAS	0,6488095	
Cluster_8			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY
	Aspirantes	0,002	1
	Cesantes	0,043	1
	Desempleoabierto	0,034	1
	Desempleooculto	0,01	1
	Desempleototal	0,045	1
	Edad Correcto	3-4	1
	Escolaridad	PRIMARIA	1
	Participacionglobal	0,74	1
	Sexo	MASCULINO	1
	Tasaocupacion	0,707	1
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	1
	Arma Empleada	SIN EMPLEO DE ARMAS	0,9777778
	Tipo Sitio	AFUERA	0,9555556
	Clase De Empleado	EMPLEADO PARTICULAR	0,9444444

### Descripción de los modelos resultantes de la vista minable de la ciudad de Bogotá:

- **Cluster\_1:** Se caracteriza por agrupar a las víctimas con edad de 15 años, en una comunidad en donde el 55% de las personas son empleadas.

- **Cluster\_2:** Se caracteriza por agrupar a las víctimas de sexo femenino con una edad entre 8 y 9 años en donde 51% de la comunidad son empleados.
- **Cluster\_3:** Se caracteriza por agrupar a las víctimas de sexo femenino de 9 años, con nacionalidad extranjera. En una comunidad en donde hay un 51% de personas empleadas.
- **Cluster\_4:** Se caracteriza por agrupar a las víctimas de sexo femenino entre 23 y 24 años, en donde el 66% son personas empleadas y donde hay una tasa de desempleo del 11%.
- **Cluster\_5:** Se caracteriza por agrupar a las víctimas de sexo femenino entre 22 y 23 años, en donde aproximadamente el 72% hacen parte de la participación global(reciben ingresos), poseen un nivel de escolaridad de Secundaria, y forman parte del 59% de la Tasa de Ocupación. Y el 15% son desempleadas.
- **Cluster\_6:** Se caracteriza por agrupar a las víctimas de sexo masculino con una edad entre 11 y 12 año, en una comunidad en donde el 70% son empleados.
- **Cluster\_7:** Se caracteriza por agrupar a las víctimas con una edad entre 12 y 13 años, de sexo femenino, que sufrieron un delito en sitio bajo techo, en una comunidad en donde el 66% son empleados.
- **Cluster\_8:** Se caracteriza por agrupar a las víctimas de sexo masculino, con una edad entre 3 y 4 años, en una comunidad donde hay un 70% de personas

## 8.1.2. Vista Ciudad – Medellín

Tabla 6 Arboles sobre clúster Medellín

Cluster	Descripción
Cluster 1	Barrio = PARQUE NORTE C-4 = true and Escolaridad = PRIMARIA = false TipoDelito = DELITO CONTRA ADULTO
Clúster 2	participacionglobal $\leq$ 0.614
Clúster 3	aspirantes $\leq$ 0.207
Clúster 4	participacionglobal $\leq$ 0.614 and TipoDelito = DELITO CONTRA MENOR DE 14 AÑOS = true and Día = Domingo = false
Clúster 5	Sexo = FEMENINO = false and Clase de sitio = VIAS PUBLICAS = true
Clúster 6	participacionglobal $\leq$ 0.614 and Día = Domingo = false
Clúster 7	desempleoabierto $>$ 0.565 and Clase de empleado = ESTUDIANTE = true Clase de sitio = VIAS PUBLICAS
Clúster 8	COLOMBIANO/A = true and participacionglobal $\leq$ 0.522

# DELITOS SEXUALES EN COLOMBIA



Tabla 7 Características de los clusters de Medellín

Cluster	Características		
Cluster_1	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILITY</b>
	Aspirantes	0,007	1
	Cesantes	0,074	1
	Desempleoabierto	0,082	1
	Desempleooculto	0	1
	Desempleototal	0,082	1
	Edad Correcto	8-9	1
	Escolaridad	PRIMARIA	1
	Participacionglobal	0,474	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,435	1
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	1
	Estado Civil	SOLTERO	0,9993017
	Zona	URBANA	0,9958101
	Arma Empleada	SIN EMPLEO DE ARMAS	0,9664804
	Clase De Empleado	EMPLEADO PARTICULAR	0,9203911
	Tipo Sitio	AFUERA	0,9057263
	Clase De Sitio	VIAS PUBLICAS	0,9022346
Cluster_2	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILIT Y</b>
	Aspirantes	0,002	1
	Cesantes	0,103678	1
	Edad Correcto	8-9	1
	Escolaridad	PRIMARIA	1
	Estado Civil	SOLTERO	1
	Zona	URBANA	0,9954002
	Arma Empleada	SIN EMPLEO DE ARMAS	0,9549218
	Clase De Empleado	EMPLEADO PARTICULAR	0,9356026
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	0,9070837
	Tipo Sitio	AFUERA	0,8436063
	Clase De Sitio	VIAS PUBLICAS	0,8408464
	Desempleoabierto	0,084	0,6513339
	Desempleooculto	0,029	0,6513339
	Desempleototal	0,113	0,6513339
	Participacionglobal	0,481	0,6513339
	Sexo	FEMENINO	0,6513339
	Tasaocupacion	0,426	0,6513339
Cluster_3	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILIT Y</b>
	Aspirantes	0,0156219	1
	Cesantes	0,1059249	1
	Edad Correcto	24	1
	Zona	URBANA	0,9876847
	Estado Civil	SOLTERO	0,8928571
	Sexo	FEMENINO	0,8189655
	Tipo Sitio	AFUERA	0,7672414

# DELITOS SEXUALES EN COLOMBIA



	<table><tr><td>Tipo Delito</td><td>DELITO CONTRA ADULTO</td><td>0,7586207</td></tr><tr><td>Clase De Sitio</td><td>VIAS PUBLICAS</td><td>0,7536946</td></tr><tr><td>Clase De Empleado</td><td>EMPLEADO PARTICULAR</td><td>0,7389163</td></tr><tr><td>Arma Empleada</td><td>SIN EMPLEO DE ARMAS</td><td>0,703202</td></tr><tr><td>Escolaridad</td><td>SECUNDARIA</td><td>0,6810345</td></tr><tr><td>País De Nacimiento</td><td>COLOMBIA</td><td>0,6699507</td></tr></table>	Tipo Delito	DELITO CONTRA ADULTO	0,7586207	Clase De Sitio	VIAS PUBLICAS	0,7536946	Clase De Empleado	EMPLEADO PARTICULAR	0,7389163	Arma Empleada	SIN EMPLEO DE ARMAS	0,703202	Escolaridad	SECUNDARIA	0,6810345	País De Nacimiento	COLOMBIA	0,6699507																																				
Tipo Delito	DELITO CONTRA ADULTO	0,7586207																																																					
Clase De Sitio	VIAS PUBLICAS	0,7536946																																																					
Clase De Empleado	EMPLEADO PARTICULAR	0,7389163																																																					
Arma Empleada	SIN EMPLEO DE ARMAS	0,703202																																																					
Escolaridad	SECUNDARIA	0,6810345																																																					
País De Nacimiento	COLOMBIA	0,6699507																																																					
Cluster_4	<table><tr><th>ATTRIBUTE_NAME</th><th>ATTRIBUTE_VALUE</th><th>PROBABILITY</th></tr><tr><td>Aspirantes</td><td>0,007</td><td>1</td></tr><tr><td>Cesantes</td><td>0,074</td><td>1</td></tr><tr><td>Desempleoabierto</td><td>0,082</td><td>1</td></tr><tr><td>Desempleooculto</td><td>0</td><td>1</td></tr><tr><td>Desempleototal</td><td>0,082</td><td>1</td></tr><tr><td>Edad Correcto</td><td>26</td><td>1</td></tr><tr><td>Escolaridad</td><td>PRIMARIA</td><td>1</td></tr><tr><td>Participacionglobal</td><td>0,474</td><td>1</td></tr><tr><td>Sexo</td><td>FEMENINO</td><td>1</td></tr><tr><td>Tasaocupacion</td><td>0,435</td><td>1</td></tr><tr><td>Zona</td><td>URBANA</td><td>1</td></tr><tr><td>Clase De Empleado</td><td>EMPLEADO PARTICULAR</td><td>0,9878049</td></tr><tr><td>Clase De Sitio</td><td>VIAS PUBLICAS</td><td>0,9463415</td></tr><tr><td>Tipo Sitio</td><td>AFUERA</td><td>0,9463415</td></tr><tr><td>Tipo Delito</td><td>DELITO CONTRA ADULTO</td><td>0,9439024</td></tr><tr><td>Estado Civil</td><td>SOLTERO</td><td>0,8902439</td></tr><tr><td>Arma Empleada</td><td>SIN EMPLEO DE ARMAS</td><td>0,8170732</td></tr></table>	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY	Aspirantes	0,007	1	Cesantes	0,074	1	Desempleoabierto	0,082	1	Desempleooculto	0	1	Desempleototal	0,082	1	Edad Correcto	26	1	Escolaridad	PRIMARIA	1	Participacionglobal	0,474	1	Sexo	FEMENINO	1	Tasaocupacion	0,435	1	Zona	URBANA	1	Clase De Empleado	EMPLEADO PARTICULAR	0,9878049	Clase De Sitio	VIAS PUBLICAS	0,9463415	Tipo Sitio	AFUERA	0,9463415	Tipo Delito	DELITO CONTRA ADULTO	0,9439024	Estado Civil	SOLTERO	0,8902439	Arma Empleada	SIN EMPLEO DE ARMAS	0,8170732
ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY																																																					
Aspirantes	0,007	1																																																					
Cesantes	0,074	1																																																					
Desempleoabierto	0,082	1																																																					
Desempleooculto	0	1																																																					
Desempleototal	0,082	1																																																					
Edad Correcto	26	1																																																					
Escolaridad	PRIMARIA	1																																																					
Participacionglobal	0,474	1																																																					
Sexo	FEMENINO	1																																																					
Tasaocupacion	0,435	1																																																					
Zona	URBANA	1																																																					
Clase De Empleado	EMPLEADO PARTICULAR	0,9878049																																																					
Clase De Sitio	VIAS PUBLICAS	0,9463415																																																					
Tipo Sitio	AFUERA	0,9463415																																																					
Tipo Delito	DELITO CONTRA ADULTO	0,9439024																																																					
Estado Civil	SOLTERO	0,8902439																																																					
Arma Empleada	SIN EMPLEO DE ARMAS	0,8170732																																																					
Cluster_5	<table><tr><th>ATTRIBUTE_NAME</th><th>ATTRIBUTE_VALUE</th><th>PROBABILITY</th></tr><tr><td>Aspirantes</td><td>0,007</td><td>1</td></tr><tr><td>Cesantes</td><td>0,074</td><td>1</td></tr><tr><td>Desempleoabierto</td><td>0,082</td><td>1</td></tr><tr><td>Desempleooculto</td><td>0</td><td>1</td></tr><tr><td>Desempleototal</td><td>0,082</td><td>1</td></tr><tr><td>Edad Correcto</td><td>14-15</td><td>1</td></tr><tr><td>Escolaridad</td><td>PRIMARIA</td><td>1</td></tr><tr><td>Participacionglobal</td><td>0,474</td><td>1</td></tr><tr><td>Sexo</td><td>FEMENINO</td><td>1</td></tr><tr><td>Tasaocupacion</td><td>0,435</td><td>1</td></tr><tr><td>Tipo Delito</td><td>DELITO CONTRA MENOR DE EDAD</td><td>1</td></tr><tr><td>Zona</td><td>URBANA</td><td>1</td></tr><tr><td>Estado Civil</td><td>SOLTERO</td><td>0,995098</td></tr><tr><td>Clase De Empleado</td><td>EMPLEADO PARTICULAR</td><td>0,9901961</td></tr><tr><td>Clase De Sitio</td><td>VIAS PUBLICAS</td><td>0,9656863</td></tr><tr><td>Tipo Sitio</td><td>AFUERA</td><td>0,9656863</td></tr><tr><td>Arma Empleada</td><td>SIN EMPLEO DE ARMAS</td><td>0,8088235</td></tr></table>	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY	Aspirantes	0,007	1	Cesantes	0,074	1	Desempleoabierto	0,082	1	Desempleooculto	0	1	Desempleototal	0,082	1	Edad Correcto	14-15	1	Escolaridad	PRIMARIA	1	Participacionglobal	0,474	1	Sexo	FEMENINO	1	Tasaocupacion	0,435	1	Tipo Delito	DELITO CONTRA MENOR DE EDAD	1	Zona	URBANA	1	Estado Civil	SOLTERO	0,995098	Clase De Empleado	EMPLEADO PARTICULAR	0,9901961	Clase De Sitio	VIAS PUBLICAS	0,9656863	Tipo Sitio	AFUERA	0,9656863	Arma Empleada	SIN EMPLEO DE ARMAS	0,8088235
ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY																																																					
Aspirantes	0,007	1																																																					
Cesantes	0,074	1																																																					
Desempleoabierto	0,082	1																																																					
Desempleooculto	0	1																																																					
Desempleototal	0,082	1																																																					
Edad Correcto	14-15	1																																																					
Escolaridad	PRIMARIA	1																																																					
Participacionglobal	0,474	1																																																					
Sexo	FEMENINO	1																																																					
Tasaocupacion	0,435	1																																																					
Tipo Delito	DELITO CONTRA MENOR DE EDAD	1																																																					
Zona	URBANA	1																																																					
Estado Civil	SOLTERO	0,995098																																																					
Clase De Empleado	EMPLEADO PARTICULAR	0,9901961																																																					
Clase De Sitio	VIAS PUBLICAS	0,9656863																																																					
Tipo Sitio	AFUERA	0,9656863																																																					
Arma Empleada	SIN EMPLEO DE ARMAS	0,8088235																																																					
Cluster_6	<table><tr><th>ATTRIBUTE_NAME</th><th>ATTRIBUTE_VALUE</th><th>PROBABILITY</th></tr><tr><td>Aspirantes</td><td>0,003</td><td>1</td></tr><tr><td>Cesantes</td><td>0,081</td><td>1</td></tr><tr><td>Desempleoabierto</td><td>0,081</td><td>1</td></tr><tr><td>Desempleooculto</td><td>0,003</td><td>1</td></tr><tr><td>Desempleototal</td><td>0,084</td><td>1</td></tr><tr><td>Edad Correcto</td><td>8</td><td>1</td></tr><tr><td>Escolaridad</td><td>PRIMARIA</td><td>1</td></tr><tr><td>Estado Civil</td><td>SOLTERO</td><td>1</td></tr></table>	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY	Aspirantes	0,003	1	Cesantes	0,081	1	Desempleoabierto	0,081	1	Desempleooculto	0,003	1	Desempleototal	0,084	1	Edad Correcto	8	1	Escolaridad	PRIMARIA	1	Estado Civil	SOLTERO	1																											
ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY																																																					
Aspirantes	0,003	1																																																					
Cesantes	0,081	1																																																					
Desempleoabierto	0,081	1																																																					
Desempleooculto	0,003	1																																																					
Desempleototal	0,084	1																																																					
Edad Correcto	8	1																																																					
Escolaridad	PRIMARIA	1																																																					
Estado Civil	SOLTERO	1																																																					

# DELITOS SEXUALES EN COLOMBIA



	Participacionglobal	0,68	1
	Sexo	MASCULINO	1
	Tasaocupacion	0,623	1
	Zona	URBANA	0,9936709
	Arma Empleada	SIN EMPLEO DE ARMAS	0,9683544
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	0,9113924
	Clase De Empleado	EMPLEADO PARTICULAR	0,9050633
	Tipo Sitio	AFUERA	0,8291139
	Clase De Sitio	VIAS PUBLICAS	0,8227848
	País De Nacimiento	COLOMBIA	0,7088608
<b>Cluster_7</b>			
	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILIT Y</b>
	Aspirantes	0,019	1
	Cesantes	0,103	1
	Desempleoabierto	0,113	1
	Desempleooculto	0,01	1
	Desempleototal	0,123	1
	Edad Correcto	13	1
	Escolaridad	SECUNDARIA	1
	Estado Civil	SOLTERO	1
	Participacionglobal	0,665	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,583	1
	País De Nacimiento	COLOMBIA	0,983871
	Zona	URBANA	0,9758065
	Clase De Empleado	ESTUDIANTE	0,9112903
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	0,6854839
<b>Cluster_8</b>			
	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILITY</b>
	Aspirantes	0,007	1
	Cesantes	0,074	1
	Clase De Empleado	EMPLEADO PARTICULAR	1
	Clase De Sitio	VIAS PUBLICAS	1
	Desempleoabierto	0,082	1
	Desempleooculto	0	1
	Desempleototal	0,082	1
	Edad Correcto	1-3	1
	Escolaridad	PRIMARIA	1
	Estado Civil	SOLTERO	1
	Participacionglobal	0,474	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,435	1
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	1
	Tipo Sitio	AFUERA	1
	Zona	URBANA	1
	Arma Empleada	SIN EMPLEO DE ARMAS	0,9473684
	País De Nacimiento	NO REPORTADO	0,6315789

**Descripción de los modelos resultantes de la vista minable de la ciudad de Medellín:**

- **Cluster 1:** Se caracteriza por agrupar a las víctimas del género femenino con educación primaria en donde la población tiene un porcentaje de desempleo mayor al 50%.
- **Cluster 2:** Se caracteriza por agrupar a las víctimas del género femenino con una edad promedio entre 8 a 9 años, con educación primaria en donde la minoría de gente entre 14 y 64 años de la población trabaja.
- **Cluster 3:** Se caracteriza por agrupar a las víctimas del género femenino con una edad promedio de 24 años con educación secundaria.
- **Cluster 4:** Se caracteriza por agrupar a las víctimas del género femenino con una edad promedio de 26 años que con una educación primaria.
- **Cluster 5:** Se caracteriza por agrupar a la victimas del género femenino con una edad promedio de 14 a 15 años.
- **Cluster 6:** Se caracteriza por agrupar a las víctimas del género masculino con una edad promedio de 8 años, con una comunidad empleada de más del 60%.
- **Cluster 7:** Se caracteriza por agrupar a las víctimas del género femenino con una escolaridad secundaria.
- **Cluster 8:** Se caracteriza por agrupar a las víctimas del género femenino con una edad promedio de 1 año a las cuales no se reportó su país de nacimiento.

## 8.1.3. Vista Ciudad – Cali

Cluster	Descripción
<b>Cluster 1</b>	TipoSitio = AFUERA = false and participacionglobal > 0.738
<b>Clúster 2</b>	Clase de empleado = ESTUDIANTE = true and Día = Domingo = false and Escolaridad = SECUNDARIA = true
<b>Clúster 3</b>	AFUERA = true and Escolaridad = PRIMARIA = true and Arma empleada = SIN EMPLEO DE ARMAS = true and Barrio = VALLE GRANDE E21 = false
<b>Clúster 4</b>	Escolaridad = ANALFABETA = true
<b>Clúster 5</b>	Clase de empleado = EMPLEADO PARTICULAR = true
<b>Clúster 6</b>	TipoDelito = DELITO CONTRA ADULTO = true
<b>Clúster 7</b>	Escolaridad = PRIMARIA = true and TipoDelito = DELITO CONTRA ADULTO = false
<b>Clúster 8</b>	TipoSitio = AFUERA = true and TipoDelito = DELITO CONTRA ADULTO = true and Clase de empleado = EMPLEADO PARTICULAR = true and Día = Domingo = false

# DELITOS SEXUALES EN COLOMBIA



Tabla 8 Características de los clusters de Cali

Cluster	Características		
Cluster_1	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILITY</b>
	Aspirantes	0	1
	Cesantes	0,1451476	1
	Desempleoabierto	0,1263164	1
	Desempleooculto	0,0188312	1
	Desempleototal	0,1451476	1
	Edad Correcto	12-13	1
	Escolaridad	PRIMARIA	1
	Participacionglobal	0,4647721	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,3971983	1
	Zona	URBANA	0,972255
	Estado Civil	SOLTERO	0,9710744
	País De Nacimiento	COLOMBIA	0,825856
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	0,7945691
	Arma Empleada	SIN EMPLEO DE ARMAS	0,7744982
Cluster_2	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILITY</b>
	Aspirantes	0,0274747	1
	Cesantes	0,1959812	1
	Desempleoabierto	0,2058799	1
	Desempleooculto	0,0176266	1
	Desempleototal	0,2235065	1
	Edad Correcto	22-47	1
	Escolaridad	SECUNDARIA	1
	Participacionglobal	0,702152	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,5452721	1
	Zona	URBANA	0,9609262
	País De Nacimiento	COLOMBIA	0,9334298
	Estado Civil	SOLTERO	0,8060781
	Tipo Delito	DELITO CONTRA ADULTO	0,7308249
Cluster_3	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILITY</b>
	Aspirantes	0,028	1
	Cesantes	0,186	1
	Desempleoabierto	0,198	1
	Desempleooculto	0,015	1
	Desempleototal	0,213	1
	Edad Correcto	14-15	1
	Escolaridad	SECUNDARIA	1
	Participacionglobal	0,699	1
	Sexo	FEMENINO	1
	Tasaocupacion	0,55	1
	Estado Civil	SOLTERO	0,9755302
	Zona	URBANA	0,9706362
	País De Nacimiento	COLOMBIA	0,9624796
	Clase De Empleado	ESTUDIANTE	0,7536705



# DELITOS SEXUALES EN COLOMBIA



Cluster_4			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILIT Y
	Aspirantes	0,0089946	1
	Cesantes	0,0991509	1
	Desempleoabierto	0,1048221	1
	Desempleooculto	0,0033747	1
	Desempleototal	0,1081375	1
	Edad Correcto	21-22	1
	Participacionglobal	0,8180647	1
	Tasaocupacion	0,7290027	1
	Zona	URBANA	0,9622642
	País De Nacimiento	COLOMBIA	0,9245283
	Estado Civil	SOLTERO	0,8490566
	Sexo	MASCULINO	0,7304582
Cluster_5			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY
	Aspirantes	0	1
	Cesantes	0,084	1
	Desempleoabierto	0,081	1
	Desempleooculto	0,003	1
	Desempleototal	0,084	1
	Edad Correcto	9-10	1
	Escolaridad	PRIMARIA	1
	Participacionglobal	0,62	1
	Sexo	MASCULINO	1
	Tasaocupacion	0,568	1
	Estado Civil	SOLTERO	0,9885057
	Zona	URBANA	0,9501916
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	0,862069
	País De Nacimiento	COLOMBIA	0,8237548
	Arma Empleada	SIN EMPLEO DE ARMAS	0,7854406
Cluster_6			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY
	Edad Correcto	15-22	1
	País De Nacimiento	COLOMBIA	0,9439252
	Tipo Delito	DELITO CONTRA ADULTO	0,9376947
	Sexo	FEMENINO	0,9034268
	Estado Civil	SOLTERO	0,8909657
	Escolaridad	SECUNDARIA	0,7757009
Cluster_7			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILIT Y
	Edad Correcto	12-13	1
	Escolaridad	SECUNDARIA	1
	Estado Civil	SOLTERO	0,9883041
	País De Nacimiento	COLOMBIA	0,9707602
	Clase De Empleado	ESTUDIANTE	0,9298246
	Tipo Delito	DELITO CONTRA MENOR DE 14 AÑOS	0,9005848
	Arma Empleada	SIN EMPLEO DE ARMAS	0,6959064
	Aspirantes	0,027	0,6900585
	Cesantes	0,205	0,6900585

	Desempleoabierto	0,213	0,6900585
	Desempleooculto	0,02	0,6900585
	Desempleototal	0,233	0,6900585
	Participacionglobal	0,705	0,6900585
	Sexo	FEMENINO	0,6900585
	Tasaocupacion	0,541	0,6900585
	Tipo Sitio	SITIO BAJO TECHO	0,6081871
<b>Cluster_8</b>			
	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILIT Y</b>
	Edad Correcto	22-47	1
	Tipo Delito	DELITO CONTRA ADULTO	1
	Aspirantes	0	0,9370079
	Estado Civil	SOLTERO	0,9370079
	Escolaridad	PRIMARIA	0,8740157
	Clase De Empleado	EMPLEADO PARTICULAR	0,8661417
	Sexo	FEMENINO	0,8110236
	Arma Empleada	SIN EMPLEO DE ARMAS	0,7480315
	Tipo Sitio	AFUERA	0,7244094
	Clase De Sitio	VIAS PUBLICAS	0,7086614
	País De Nacimiento	NO REPORTADO	0,6299213

## Descripción de los modelos resultantes de la vista minable de la ciudad de Cali:

- **Cluster\_1:** Se caracteriza por agrupar a las víctimas de sexo femenino, con una edad entre 12 y 13 años, en donde el 39% de la comunidad corresponde a personas empleadas.
- **Cluster\_2:** Se caracteriza por agrupar a las víctimas de sexo femenino, con una edad entre 26 y 27 años, en una comunidad en donde el 54% son personas empleadas y hay una participación global del 70%.
- **Cluster\_3:** Se caracteriza por agrupar a las víctimas de sexo femenino, con una edad entre 14 y 15 años, en donde el 55% de la población son personas empleadas y hay una participación global del 70%.
- **Cluster\_4:** Se caracteriza por agrupar a las víctimas de sexo masculino entre 21 y 22 años, en una comunidad en donde el 72% son personas empleadas y la participación global es del 81%.
- **Cluster\_5:** Se caracteriza por agrupar a las víctimas de sexo masculino entre 9 y 10 años, en una comunidad donde el 56% son personas empleadas y hay una participación total del 62%.
- **Cluster\_6:** Se caracteriza por agrupar a las víctimas con una edad entre 19 y 20 años.
- **Cluster\_7:** Se caracteriza por agrupar a las víctimas de sexo femenino entre 12 y 13 años, que sufrieron el delito en un sitio bajo techo y en donde el 70% de la comunidad son personas empleadas.

- **Cluster\_8:** Se caracteriza por agrupar a las víctimas de sexo femenino, con una edad de 22-47 años, que sufrieron un delito en vías públicas, no reportan su país de nacimiento, tienen un nivel de escolaridad de primaria y son empleadas particulares.

## 8.1.4. Vista otras ciudades

Tabla 9 Características de los clusters de otra ciudad

Cluster	Características		
Cluster_1	<b>ATTRIBUTE_NAME</b>	<b>ATTRIBUTE_VALUE</b>	<b>PROBABILITY</b>
	Aspirantes	0,0130663	1
	Cesantes	0,1170587	1
	Clase De Empleado	EMPLEADO PARTICULAR	0,6560292
	COBERTURA BRUTA MEDIA	71,02589	1
	COBERTURA BRUTA PRIMARIA	67,719505	1
	COBERTURA BRUTA SECUNDARIA	106,82856	1
	COBERTURA BRUTA TRANSICIÓN	86,722923	1
	Desempleoabierto	0,1210962	1
	Desempleooculto	0,008915	1
	Desempleototal	0,1300563	1
	DESERCIÓN MEDIA	1,9949013	1
	DESERCIÓN PRIMARIA	2,461961	1
	DESERCIÓN SECUNDARIA	3,1953374	1
	DESERCIÓN TRANSICIÓN	2,491581	1
	Dia Semana	ENTRE SEMANA	0,8209501
	Edad Correcto	27-28	1
	Estado Civil	SOLTERO	0,7909866
	Participacionglobal	0,6056982	1
	Percapita Correcto	1,8955566	1
	Percapita Estudiantes	1,0863849	1
	Porcentaje Municipio	0,0018956	1
	REPITENCIA MEDIA	0,9763386	1
	REPITENCIA PRIMARIA	1,3297929	1
	REPITENCIA SECUNDARIA	1,8710816	1
	REPITENCIA TRANSICIÓN	0,5869208	1
	REPROBACIÓN MEDIA	2,9383581	1
	REPROBACIÓN PRIMARIA	2,0931206	1
	REPROBACIÓN SECUNDARIA	4,5275566	1
	REPROBACIÓN TRANSICIÓN	0,2350792	1
	Sexo	FEMENINO	0,8823386
	Tasaocupacion	0,5235199	1
	Tipo Sitio	AFUERA	0,6387333

# DELITOS SEXUALES EN COLOMBIA



Cluster_2			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY
	Arma Empleada	SIN EMPLEO DE ARMAS	0,6334661
	Aspirantes	0,0055068	1
	Cesantes	0,0892627	1
	Clase De Empleado	EMPLEADO PARTICULAR	0,8289698
	Clase De Sitio	VIAS PUBLICAS	0,7746158
	COBERTURA BRUTA MEDIA	79,547957	1
	COBERTURA BRUTA PRIMARIA	87,248014	1
	COBERTURA BRUTA SECUNDARIA	89,103028	1
	COBERTURA BRUTA TRANSICIÓN	74,314815	1
	Departamento	CUNDINAMARCA	0,9430848
	Desempleoabierto	0,0880569	1
	Desempleooculto	0,0070139	1
	Desempleototal	0,0947168	1
	DESERCIÓN MEDIA	2,4731702	1
	DESERCIÓN PRIMARIA	2,0003244	1
	DESERCIÓN SECUNDARIA	3,0236596	1
	DESERCIÓN TRANSICIÓN	0,4088731	1
	Dia Semana	ENTRE SEMANA	0,8258395
	Edad Correcto	27-28	1
	Estado Civil	SOLTERO	0,8591349
	Municipio	BOGOTÁ D.C. (CT)	0,9430848
	Participacionglobal	0,6385737	1
	Percapita Correcto	1,2399203	1
	Percapita Estudiantes	0,7498069	1
	Porcentaje Municipio	0,0012399	1
	REPITENCIA MEDIA	3,9940581	1
	REPITENCIA PRIMARIA	3,5587763	1
	REPITENCIA SECUNDARIA	8,2029994	1
	REPITENCIA TRANSICIÓN	0,2110586	1
	REPROBACIÓN MEDIA	0,2468355	1
	REPROBACIÓN PRIMARIA	0,0810586	1
	REPROBACIÓN SECUNDARIA	0,3126067	1
	REPROBACIÓN TRANSICIÓN	0,0017075	1
	Sexo	FEMENINO	0,8989755
	Tasaocupacion	0,5760592	1
	Tipo Sitio	AFUERA	0,7899829
Cluster_3			
	ATTRIBUTE_NAME	ATTRIBUTE_VALUE	PROBABILITY
	Arma Empleada	SIN EMPLEO DE ARMAS	0,6935065
	Aspirantes	0,0128312	1
	Cesantes	0,1016455	1
	Clase De Empleado	EMPLEADO PARTICULAR	0,8577922
	Clase De Sitio	VIAS PUBLICAS	0,8227273
	COBERTURA BRUTA MEDIA	103,43506	1

## DELITOS SEXUALES EN COLOMBIA



COBERTURA BRUTA PRIMARIA	113,1426	1
COBERTURA BRUTA SECUNDARIA	123,73882	1
COBERTURA BRUTA TRANSICIÓN	96,347701	1
Departamento	ANTIOQUIA	0,6532468
Desempleoabierto	0,1050734	1
Desempleooculto	0,0097149	1
Desempleototal	0,1147974	1
DESERCIÓN MEDIA	1,4787273	1
DESERCIÓN PRIMARIA	0,8865584	1
DESERCIÓN SECUNDARIA	2,0947013	1
DESERCIÓN TRANSICIÓN	0,1291948	1
Dia Semana	ENTRE SEMANA	0,8298701
Edad Correcto	34-56	1
Estado Civil	SOLTERO	0,8649351
Municipio	MEDELLÍN (CT)	0,6532468
Participacionglobal	0,5978182	1
Percapita Correcto	1,7019221	1
Percapita Estudiantes	1,1744072	1
REPITENCIA MEDIA	0,8913636	1
REPITENCIA PRIMARIA	1,2203117	1
REPITENCIA SECUNDARIA	3,1448052	1
REPITENCIA TRANSICIÓN	0,2187013	1
REPROBACIÓN MEDIA	4,0714026	1
REPROBACIÓN PRIMARIA	2,5658052	1
REPROBACIÓN SECUNDARIA	6,4473636	1
REPROBACIÓN TRANSICIÓN	0,0718571	1
Sexo	FEMENINO	0,9077922
Tasaocupacion	0,527413	1
Tipo Sitio	AFUERA	0,85

### Descripción de los modelos resultantes de la vista minable de los delitos contra adultos:

- **Cluster 1:** Se caracteriza por agrupar a las víctimas de género femenino, con una edad promedio entre 27 y 28 años, en una comunidad donde la deserción de los estudiantes en general (transición, primaria, secundaria y media) es alta y cada 1,8 personas de cada 1000 son víctimas de delitos sexuales.
- **Cluster 2:** Se caracteriza por agrupar a las víctimas de la ciudad de Bogotá, de género femenino, con una edad entre 27 y 28 años, en una comunidad en donde la repitencia y deserción en secundaria poseen un índice alto. Adicionalmente, el 57% de las personas se encuentran empleadas .
- **Cluster 3:** Se caracteriza por agrupar a las víctimas de la ciudad de Medellín, con una edad entre 34 y 56 años, en donde la mayoría son mujeres. En una comunidad en donde el índice de reprobación y deserción en secundaria son bastante altos y en donde el 52% son personas empleadas.

## 8.2. Reglas de asociación

A continuación, se listan las reglas obtenidas utilizando Visual Studio junto con su soporte e importancia:

Pr...	Importance	Rule
0,630	0,608	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, Sexo = FEMENINO -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,623	0,598	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, Clase De Empleado = EMPLEADO PARTICULAR -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,621	0,596	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, Situación Laboral = EMPLEADO -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,616	0,598	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, Estado Civil = SOLTERO -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,616	0,595	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, Día Semana = ENTRE SEMANA -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,615	0,600	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,612	0,589	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, Clase De Sitio = VIAS PUBLICAS -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,612	0,589	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, Nacionalidad = COLOMBIANO/A -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,612	0,589	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, País De Nacimiento = COLOMBIA -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,612	0,597	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, Zona = URBANA -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,612	0,589	Arma Empleada = ARMA BLANCA / CORTOPUNZANTE, Tipo Sitio = AFUERA -> DELITO = ARTÍCULO 205. ACCESO CARNAL VIOLENTO
0,987	0,550	Escolaridad = TECNOLOGO -> Tipo Delito = DELITO CONTRA ADULTO
0,986	0,550	Escolaridad = TECNOLOGO, Nacionalidad = COLOMBIANO/A -> Tipo Delito = DELITO CONTRA ADULTO
0,986	0,550	Escolaridad = TECNOLOGO, País De Nacimiento = COLOMBIA -> Tipo Delito = DELITO CONTRA ADULTO
0,986	0,550	Escolaridad = TECNOLOGO, Zona = URBANA -> Tipo Delito = DELITO CONTRA ADULTO
0,985	0,552	Estado Civil = CASADO, Situación Laboral = EMPLEADO -> Tipo Delito = DELITO CONTRA ADULTO
0,985	0,549	Escolaridad = TECNOLOGO, Sexo = FEMENINO -> Tipo Delito = DELITO CONTRA ADULTO
0,984	0,550	Estado Civil = CASADO, Clase De Empleado = EMPLEADO PARTICULAR -> Tipo Delito = DELITO CONTRA ADULTO
0,975	0,544	Estado Civil = CASADO, Arma Empleada = SIN EMPLEO DE ARMAS -> Tipo Delito = DELITO CONTRA ADULTO
0,968	0,548	Estado Civil = CASADO, Sexo = FEMENINO -> Tipo Delito = DELITO CONTRA ADULTO
0,967	0,547	Estado Civil = CASADO, Día Semana = ENTRE SEMANA -> Tipo Delito = DELITO CONTRA ADULTO
0,965	0,547	Estado Civil = CASADO, País De Nacimiento = COLOMBIA -> Tipo Delito = DELITO CONTRA ADULTO
0,965	0,547	Estado Civil = CASADO, Nacionalidad = COLOMBIANO/A -> Tipo Delito = DELITO CONTRA ADULTO
0,964	0,540	Escolaridad = SUPERIOR, Tipo Sitio = SITIO BAJO TECHO -> Tipo Delito = DELITO CONTRA ADULTO
0,963	0,548	Estado Civil = CASADO -> Tipo Delito = DELITO CONTRA ADULTO
0,962	0,547	Estado Civil = CASADO, Zona = URBANA -> Tipo Delito = DELITO CONTRA ADULTO
0,961	0,538	Estado Civil = CASADO, Percapita Correcto = 1,3713782234 - 2,0947835428 -> Tipo Delito = DELITO CONTRA ADULTO
0,954	0,537	Clase De Empleado = AMA DE CASA, Arma Empleada = CONTUNDENTES -> Tipo Delito = DELITO CONTRA ADULTO
0,953	0,535	Estado Civil = CASADO, Tipo Sitio = AFUERA -> Tipo Delito = DELITO CONTRA ADULTO
0,950	0,533	Estado Civil = CASADO, Escolaridad = SECUNDARIA -> Tipo Delito = DELITO CONTRA ADULTO
0,949	0,533	Estado Civil = CASADO, Clase De Sitio = VIAS PUBLICAS -> Tipo Delito = DELITO CONTRA ADULTO
0,946	0,537	Escolaridad = SUPERIOR -> Tipo Delito = DELITO CONTRA ADULTO
0,945	0,537	Clase De Empleado = AMA DE CASA, Tipo Sitio = SITIO BAJO TECHO -> Tipo Delito = DELITO CONTRA ADULTO
0,945	0,537	Escolaridad = SUPERIOR, Tasaocupacion >= 0,7011371911 -> Tipo Delito = DELITO CONTRA ADULTO
0,945	0,536	Escolaridad = SUPERIOR, Sexo = FEMENINO -> Tipo Delito = DELITO CONTRA ADULTO
0,944	0,536	Escolaridad = SUPERIOR, Nacionalidad = COLOMBIANO/A -> Tipo Delito = DELITO CONTRA ADULTO
0,944	0,536	Escolaridad = SUPERIOR, País De Nacimiento = COLOMBIA -> Tipo Delito = DELITO CONTRA ADULTO
0,944	0,536	Escolaridad = SUPERIOR, Zona = URBANA -> Tipo Delito = DELITO CONTRA ADULTO
0,941	0,535	Tasaocupacion >= 0,7011371911, Sexo = FEMENINO -> Tipo Delito = DELITO CONTRA ADULTO
0,941	0,533	Escolaridad = SUPERIOR, Día Semana = ENTRE SEMANA -> Tipo Delito = DELITO CONTRA ADULTO
0,938	0,531	Estado Civil = UNION LIBRE, Municipio = BOGOTÁ D.C. (CT) -> Tipo Delito = DELITO CONTRA ADULTO
0,938	0,531	Estado Civil = UNION LIBRE, Percapita Correcto < 1,3713782234 -> Tipo Delito = DELITO CONTRA ADULTO

## 9. Evaluación de calificación

### 9.1. Clustering

Para evaluar los modelos resultantes, es decir la calidad de los clusters generados se utilizaron el Índice de Davies Bouldin y las distancias hacia los centroides.

El índice de Davies Bouldin indica que tan compacto es un clúster, entre más pequeño sea el valor que indica el índice, significa que los elementos en los clusters son más cercanos y que sus centros se encuentran bien separados con respecto a los otros clusters. Manteniendo el equilibrio de alta cohesión(entre los elementos de un clúster) y bajo acoplamiento (entre clusters).

En general con todas las vistas, la ejecución del algoritmo generó clusters compactos y que se encuentran bajamente acoplados entre ellos. Además, se puede observar una distribución de los registros semejante en cada clúster.



## 9.1.1. Vista Ciudad - Bogotá

Tabla 10 Evaluacion Vista Bogotá

Cluster Model	Performance Vector
<b>Cluster Model</b>  Cluster 0: 992 items Cluster 1: 1204 items Cluster 2: 1610 items Cluster 3: 1665 items Cluster 4: 753 items Cluster 5: 2384 items Cluster 6: 1016 items Cluster 7: 1009 items Total number of items: 10633	<b>PerformanceVector</b>  PerformanceVector: Avg. within centroid distance: -4.213 Avg. within centroid distance_cluster_0: -5.951 Avg. within centroid distance_cluster_1: -5.607 Avg. within centroid distance_cluster_2: -3.473 Avg. within centroid distance_cluster_3: -3.836 Avg. within centroid distance_cluster_4: -6.263 Avg. within centroid distance_cluster_5: -3.165 Avg. within centroid distance_cluster_6: -3.808 Avg. within centroid distance_cluster_7: -4.001 Davies Bouldin: -2.599

## 9.1.2. Vista Ciudad - Medellín

Tabla 11 Evaluacion Vista Medellin

Cluster Model	Performance Vector
<b>Cluster Model</b>  Cluster 0: 410 items Cluster 1: 387 items Cluster 2: 433 items Cluster 3: 818 items Cluster 4: 433 items Cluster 5: 381 items Cluster 6: 277 items Cluster 7: 1113 items Total number of items: 4252	<b>PerformanceVector</b>  PerformanceVector: Avg. within centroid distance: -3.418 Avg. within centroid distance_cluster_0: -3.982 Avg. within centroid distance_cluster_1: -2.980 Avg. within centroid distance_cluster_2: -5.402 Avg. within centroid distance_cluster_3: -2.524 Avg. within centroid distance_cluster_4: -3.050 Avg. within centroid distance_cluster_5: -3.089 Avg. within centroid distance_cluster_6: -6.495 Avg. within centroid distance_cluster_7: -2.739 Davies Bouldin: -2.318

## 9.1.3. Vista Ciudad - Cali

Tabla 12 Evaluacion Vista Cali

Cluster Model	Performance Vector
<b>Cluster Model</b>  Cluster 0: 438 items Cluster 1: 300 items Cluster 2: 959 items Cluster 3: 346 items Cluster 4: 281 items Cluster 5: 440 items Cluster 6: 788 items Cluster 7: 482 items Total number of items: 4034	<b>PerformanceVector</b>  PerformanceVector: Avg. within centroid distance: -5.190 Avg. within centroid distance_cluster_0: -5.023 Avg. within centroid distance_cluster_1: -4.782 Avg. within centroid distance_cluster_2: -4.788 Avg. within centroid distance_cluster_3: -5.486 Avg. within centroid distance_cluster_4: -5.082 Avg. within centroid distance_cluster_5: -6.372 Avg. within centroid distance_cluster_6: -5.066 Avg. within centroid distance_cluster_7: -5.370 Davies Bouldin: -2.593

## 9.1.4. Vista otras ciudades

Tabla 13 Evaluacion Vista Otras ciudades

Cluster Model	Performance Vector
<b>Cluster Model</b>  Cluster 0: 2116 items Cluster 1: 1058 items Cluster 2: 3006 items Cluster 3: 1992 items Cluster 4: 2461 items Total number of items: 10633	<b>PerformanceVector</b>  PerformanceVector: Avg. within centroid distance: -7.328 Avg. within centroid distance_cluster_0: -8.343 Avg. within centroid distance_cluster_1: -6.166 Avg. within centroid distance_cluster_2: -8.536 Avg. within centroid distance_cluster_3: -6.785 Avg. within centroid distance_cluster_4: -5.918 Davies Bouldin: -2.233

## 9.1.5. Vista delito contra adultos

Tabla 14 Evaluación Vista delito contra adultos

Cluster Model	Performance Vector
---------------	--------------------



Cluster 0: 3906 items Cluster 1: 2483 items Cluster 2: 2776 items Total number of items: 9165	PerformanceVector: Avg. within centroid distance: -7.958 Avg. within centroid distance_cluster_0: -8.396 Avg. within centroid distance_cluster_1: -10.271 Avg. within centroid distance_cluster_2: -5.272 Davies Bouldin: -2.760
--	---

## 9.2. Reglas de Asociación

### 9.2.1. Algoritmo: FP-Growth

Métricas de validación: El índice de soporte generado para cada itemset

### 9.2.2. Algoritmo: Generador Reglas de asociación

Métricas de validación: El índice de confianza generado por cada regla

## 9.3. Árboles de decisión

Se realizó la validación de los árboles, por medio de la matriz de confusión arrojando los siguientes resultados

### 9.3.1. Vista Ciudad de Bogotá:

accuracy: 99.44% +/- 0.18% (mikro: 99.44%)

	true clus...	true clus...	true clus...	true clus...	true clus...	true clus...	true clus...	true clus...	true clus...	true clus...	true clus...	true clus...	class pr...
pred. clu...	1388	0	0	6	0	0	0	0	0	0	0	0	99.57%
pred. clu...	0	1846	0	0	0	0	0	0	0	0	0	0	100.00%
pred. clu...	0	0	706	0	0	0	0	0	0	0	0	0	100.00%
pred. clu...	6	0	0	826	2	2	0	5	0	3	3	3	97.18%
pred. clu...	0	0	0	1	1341	0	1	0	0	0	0	0	99.85%
pred. clu...	0	0	0	0	1	495	1	2	0	0	0	0	99.20%
pred. clu...	0	0	0	1	2	0	835	1	8	0	0	0	98.58%
pred. clu...	0	0	0	8	0	0	0	676	0	0	0	0	98.83%
pred. clu...	0	0	0	0	0	0	4	0	1394	0	0	0	99.71%
pred. clu...	0	0	0	2	0	0	1	0	0	786	0	0	99.62%
pred. clu...	0	0	0	0	0	2	0	0	0	0	942	0	99.79%
pred. clu...	0	0	0	1	0	0	0	0	0	0	0	562	99.82%
class rec...	99.57%	100.00%	100.00%	97.75%	99.63%	99.20%	99.17%	98.83%	99.43%	99.62%	99.68%	99.47%	

Ilustración 2 Matriz de confusión Bogotá

### 9.3.2. Vista Ciudad Medellín:

accuracy: 97.84% +/- 1.50% (mikro: 97.84%)

	true clust...	true clust...	true clust...	true clust...	true clust...	true clust...	true clust...	true clust...	class pre...
pred. clu...	373	0	0	0	0	0	0	0	100.00%
pred. clu...	2	416	0	0	0	3	0	3	98.11%
pred. clu...	0	0	1089	0	0	0	0	0	100.00%
pred. clu...	0	0	0	415	0	0	1	0	99.76%
pred. clu...	0	0	0	0	816	0	0	0	100.00%
pred. clu...	1	2	0	2	0	217	43	0	81.89%
pred. clu...	2	0	0	0	0	25	259	0	90.56%
pred. clu...	0	0	0	3	0	0	0	362	99.18%
class rec...	98.68%	99.52%	100.00%	98.81%	100.00%	88.57%	85.48%	99.18%	

Ilustración 3 Matriz de confusión Medellín

## 9.3.3. Vista Ciudad Cali:

accuracy: 96.03% +/- 0.88% (mikro: 96.03%)

	true cluster_4	true cluster_6	true cluster_0	true cluster_3	true cluster_2	true cluster_1	true cluster_7	true cluster_5	class precision
pred. cluster_4	201	0	0	0	0	3	0	3	97.10%
pred. cluster_6	20	785	1	2	0	0	0	0	97.15%
pred. cluster_0	0	0	432	0	0	0	0	2	99.54%
pred. cluster_3	0	2	0	340	2	1	0	5	97.14%
pred. cluster_2	0	0	0	2	956	1	25	0	97.15%
pred. cluster_1	7	0	0	1	0	275	2	0	96.49%
pred. cluster_7	44	0	0	0	1	20	455	0	87.50%
pred. cluster_5	9	1	5	1	0	0	0	430	96.41%
class recall	71.53%	99.62%	98.63%	98.27%	99.69%	91.67%	94.40%	97.73%	

Ilustración 4 Matriz de confusión Cali

## 9.3.4. Vista otras ciudades y Vista delito contra adultos:

**La validación resultó engorrosa de realizar, debido a la gran cantidad de filas en la base de datos, y al momento de realizarla en RapidMiner este se bloqueaba**

# 10. Visualización de los resultados Obtenidos

## 10.1. Clustering

## 10.1.1. Vista Ciudad de Bogotá:

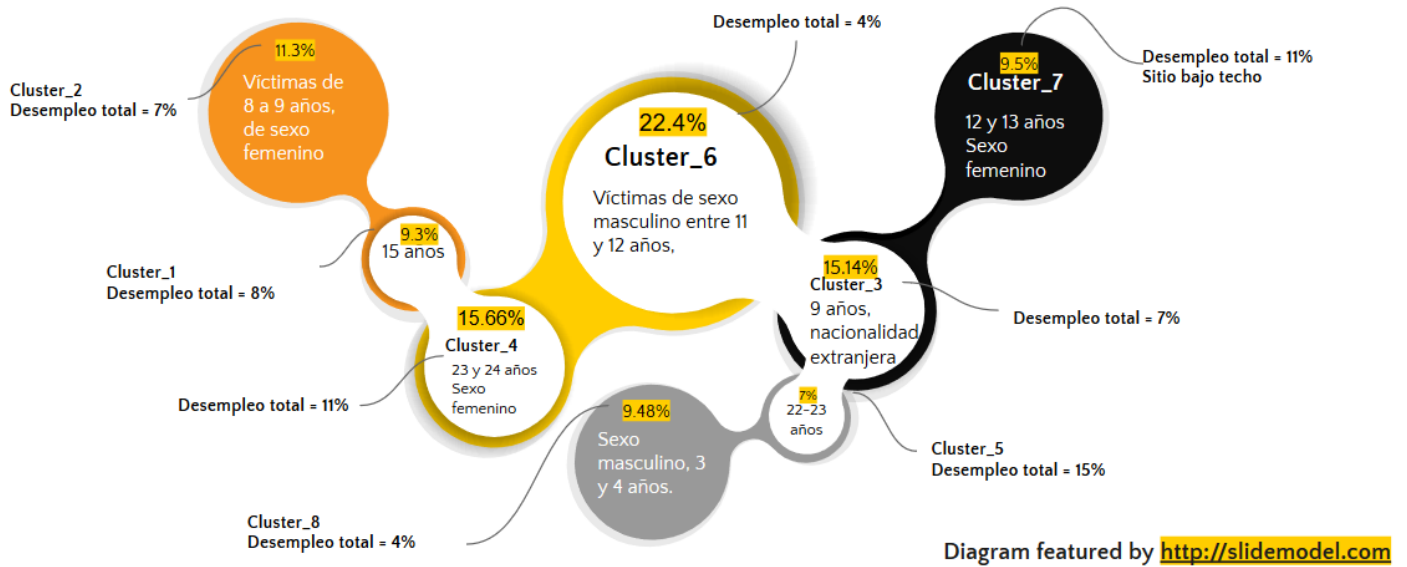


Ilustración 5 Ilustración Cluster Bogotá

## 10.1.2. Vista Ciudad Medellín:

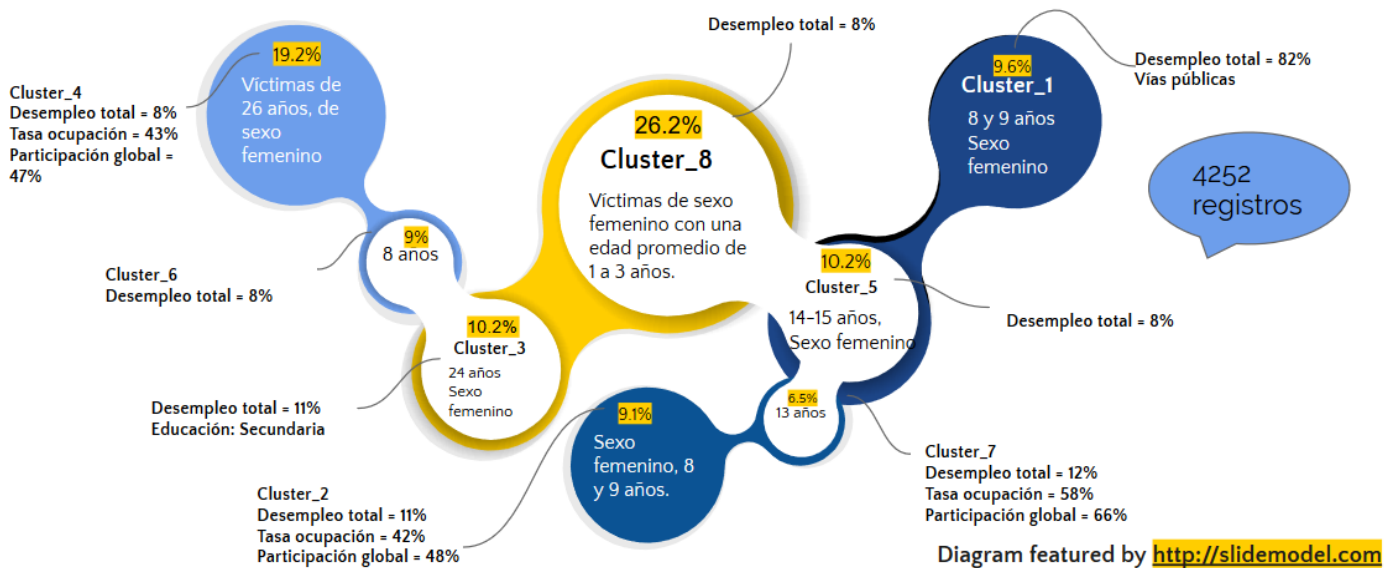


Ilustración 6 Ilustración Cluster Medellín

## 10.1.3. Vista Ciudad Cali:

# DELITOS SEXUALES EN COLOMBIA

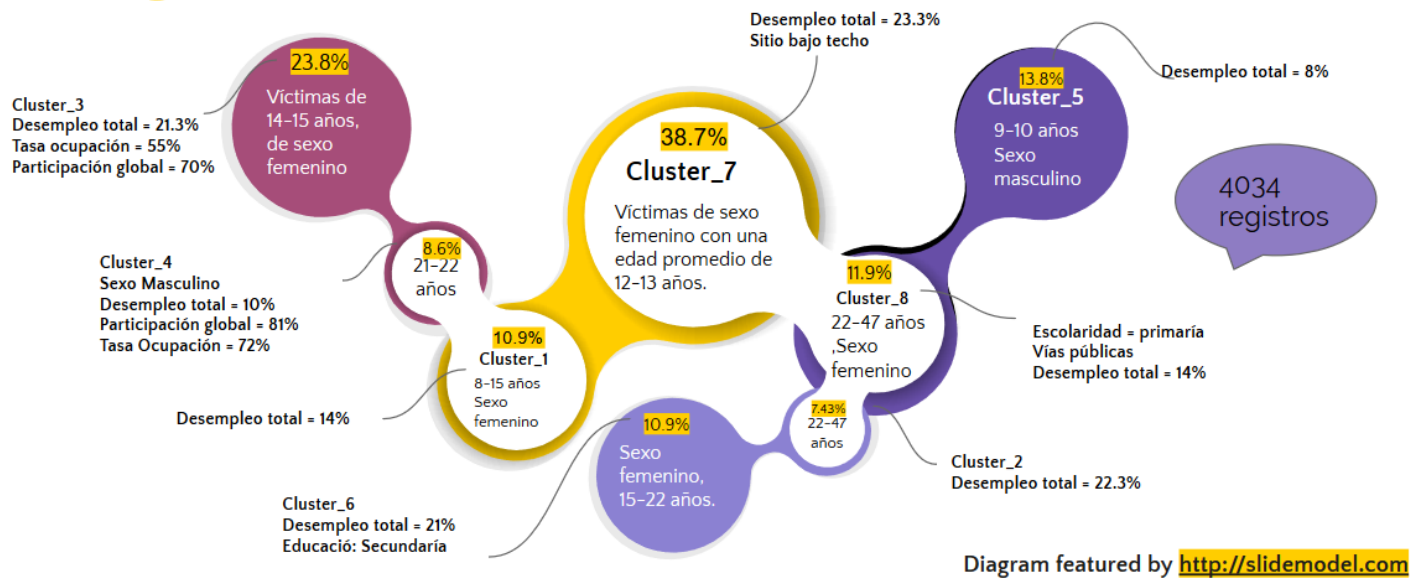


Ilustración 8 Ilustración Cluster Cali

## 10.1.4. Vista otras ciudades

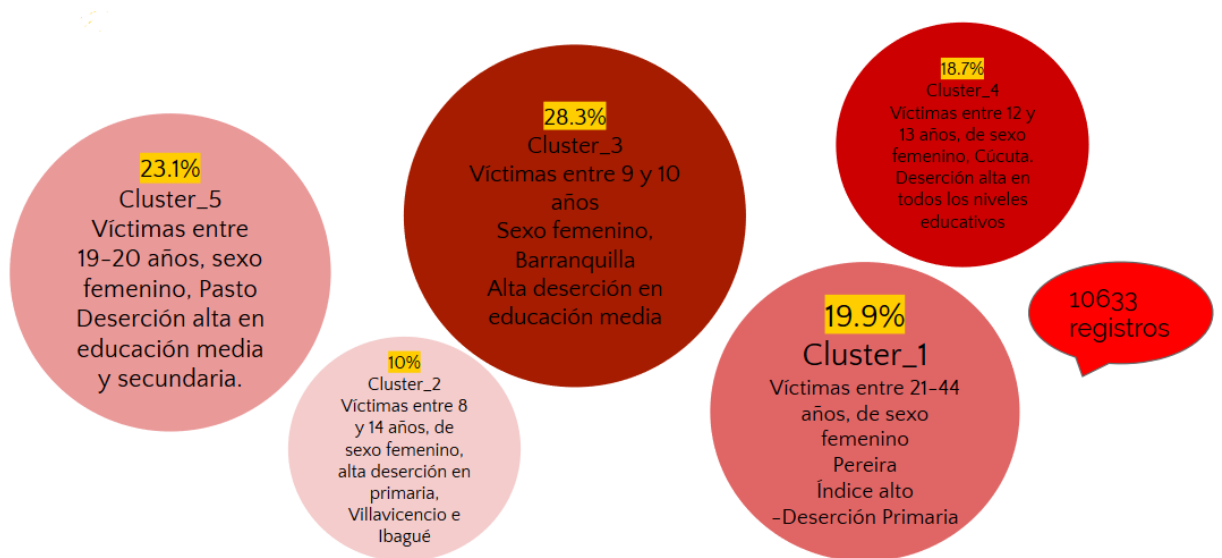


Ilustración 7 Ilustración Cluster Otras ciudades

## 10.1.5. Vista delito contra adultos

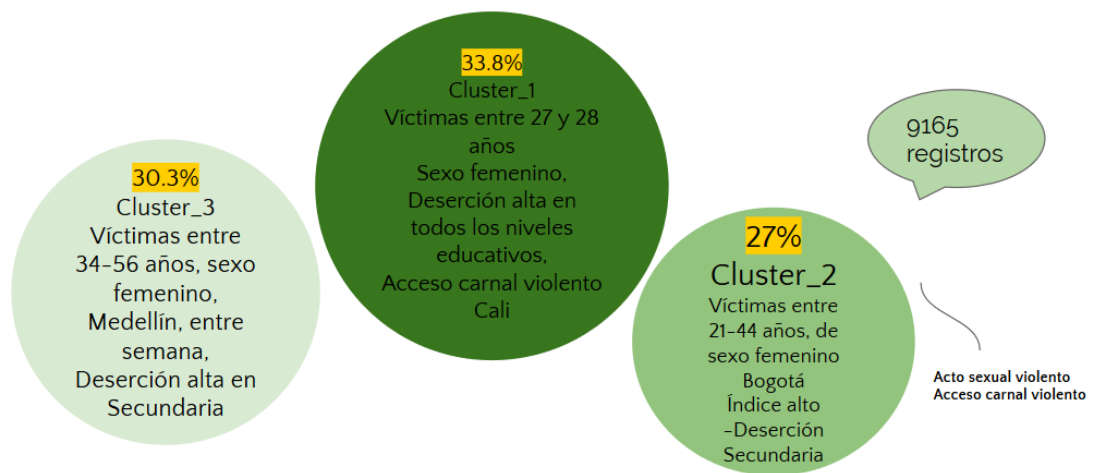


Ilustración 9 Clusters Educación y Desempleo

## 10.2. Reglas de Asociación

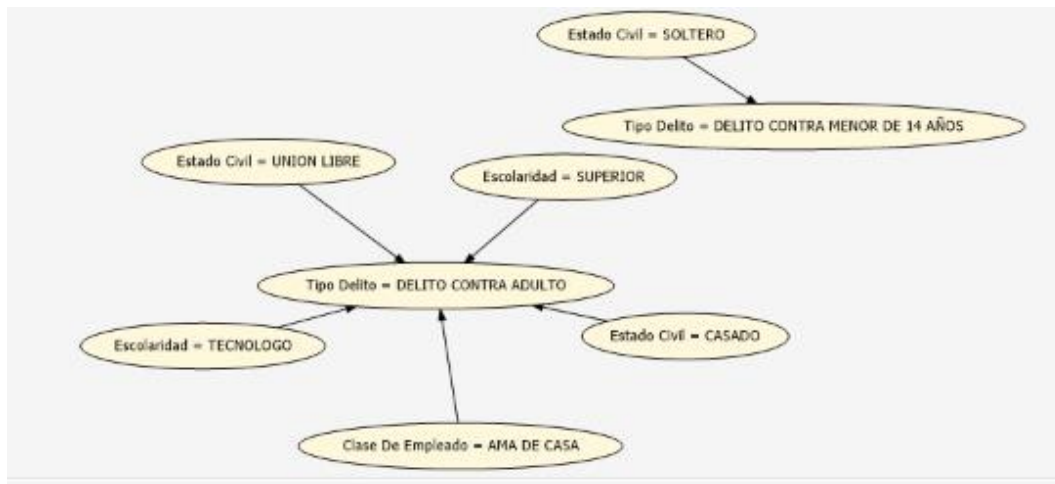


Ilustración 10 Reglas de asociación

## 11. Interpretación de resultados en términos de usuario o negocio.

La unión con las distintas bases de datos nos ayudó a lograr mejores agrupaciones en el momento de realizar clustering, dado que se logró una mejor separación de los datos distinguiendo bien un rango de edad de las víctimas, el municipio donde se concentran más delitos de cierto tipo y varias distinciones entre los datos de los atributos usados. Igualmente, se realizó validación cruzada con los resultados del clustering y por medio de árboles de

decisión, y adicionalmente de esta manera se generaron árboles de decisión para encontrar patrones que no se pueden deducir a simple vista.

No obstante, también encontramos en una gran cantidad de los clusters con información repetida que no ayuda a los resultados del negocio, por lo que hicimos al lado esta información y solo utilizamos la más relevante y la que nos ayudaba a comprender más la situación en torno a los delitos sexuales.

Con las reglas de asociación, sin embargo, podemos ver que las reglas obtenidas no generan información novedosa, sino que muestran que en realidad el nivel de educación y el estado civil covarían con la edad, y por extensión con los delitos relacionados con la edad de la víctima.

Se puede concluir de los resultados que obtuvimos de los clusters, lo siguiente:

- En la Vista independiente de Bogotá se puede observar que es el lugar donde se presentan más delitos sexuales relacionados a personas del género masculino menores de 14 años.
- En la Vista independiente de Cali, se puede resaltar que la mayoría de las víctimas mayores de edad que fueron agredidas sexualmente, son personas que no terminaron sus estudios en el colegio, además que la tasa de desempleo en dicha ciudad es la mayor de las tres ciudades que fueron analizadas.
- En la Vista que contiene las demás ciudades, se consiguió una relación muy importante. Se descubrió que entre más deserción y reprobación en los cursos de cualquier nivel educativo se presentan más casos de delitos sexuales en tal lugar relacionados con mujeres de cualquier edad.
- Para cualquiera de los resultados mencionados, se recomienda a las autoridades estar alerta a cualquier actividad sospechosa relacionada con niños de dicha aproximada edad, para evitar situaciones graves que afecten al menor de edad, además de que estos siempre deben estar en compañía de un adulto de confianza para su mayor seguridad.
- En la última vista, que contiene los datos relacionados con delitos sexuales contra adultos, se distinguieron 3 ciudades: Bogotá, Cali y Medellín. En todas estas, había un índice alto de deserción en el nivel educativo de la secundaria. Esto corrobora aún más la relación que existe entre la deserción educativa con los delitos sexuales.
- Los niveles de desempleo evidencian un factor de vulnerabilidad en las víctimas, ya sea por falta de recursos o el no tener acceso a oportunidades.

## 12. MÉTODO PARA HACER REPLICABLE LA APLICACIÓN DE LAS TÉCNICAS RECIENTES A DATOS EN LA ORGANIZACIÓN

Para que las técnicas empleadas se integren de manera eficiente tanto con el funcionamiento operacional y estratégico de la organización, es decir: La policía nacional.

Al referirnos al organigrama de la policía y transversalmente al de la DIJIN, se ha determinado que el lugar más idóneo para realizar esta integración son el ámbito operativo y administrativo de la policía nacional, de manera específica en la "Dirección de Seguridad Ciudadana" y en la "Dirección de Investigación Criminal e Interpol". Asimismo, en esta última en el Área Investigativa Especializada, en el "Grupo Investigativo Contra los Delitos Sexuales y la Familia".

No obstante, el proceso de integración debe realizarse durante todo el proceso de minería de datos. Sin embargo, son las entradas (los datos y conocimiento previo) y las salidas (los modelos y patrones), los que deben acoplarse eficazmente.

## 12.1. Integración con las entradas

Ya que las herramientas de minería de datos deben ser capaces de acceder directamente a las fuentes de datos de la organización y estas se encuentran dispersas, se propone implantar un almacén de datos centralizado para que haya una conexión directa de dichas herramientas a este. Además de un programa que verifique la calidad de los datos, el cual se centrará en la limpieza de los datos.

## 12.2. Integración con las salidas

Debido a que los modelos extraídos de la minería de datos deben aplicarse, utilizarse para tomar decisiones y difundirse en las áreas de la organización donde pueden ser útiles, es necesario integrarlos con los distintos sistemas operacionales y estratégicos que puedan existir ya implementados en esta. Ya que, continuamente están tomando decisiones y es preciso averiguar cómo la minería de datos puede afectar en estas decisiones.

Suponiendo que la policía tenga integrados sistemas propios de ayuda a la toma de decisiones (DSS, Decision Support Systems), sistemas de la administración de la información (MIS, Management Information Systems), sistemas de información ejecutiva (EIS, Executive Information Systems), así como los sistemas específicos de partes del negocio: sistemas CRM (Customer Relation Management), sistemas ERP (Enterprise Resource Planning) y sistemas SCM (Supply Chain Management). Es necesario crear interrelaciones entre estos sistemas con la minería de datos para que surjan nuevas reglas de negocio. Estas reglas obligan a revisar los procesos, adecuándolos a este nuevo conocimiento.

## 12.3. Recursos necesarios

Para llevar a cabo esta replicación, es necesario dedicar ciertos recursos, tanto materiales, como humanos. Específicamente una estación de trabajo capaz de ejecutar el software necesario y procesar la información:

### 12.3.1. Requisitos Mínimos

- Procesador de dos núcleos, 2Ghz
- 4GB RAM
- 1GB de espacio disponible
- Sistema operativo con arquitectura de 64 bits (Recomendado)
- Windows 7, Windows 8, Windows 8.1, Windows 10 (Recomendado)



- Java 8

### 12.4. *Material necesario*

Herramientas de software de minería de datos. Se recomienda RapidMiner Studio y SQL Server Data Tools. Ya que, son sistemas integrados que permiten experimentar con distintas técnicas, son fáciles de manejar y gratuitas. Los procesos utilizados para el desarrollo del proyecto fueron modelados usando estas herramientas y serán anexados para que puedan ser reutilizados.

### 12.5. *Personal necesario y formación*

Se necesita de personas que tengan experiencia con las herramientas previamente nombradas, además es primordial que dentro del personal puedan analizar los resultados de las diferentes ejecuciones que se hagan en el momento de utilizar las técnicas de minería.



## 13. CONCLUSIONES Y TRABAJO FUTURO

### 13.1. Conclusiones

Durante el desarrollo de este proyecto a través de la metodología CRISP-DM, se puede concluir que se cumplió con el objetivo correspondiente a cada fase en la implementación, lo cual nos llevó a obtener resultados de utilidad para la organización dueña del problema; La Policía Nacional.

Cabe destacar la importancia que tiene la interpretación en términos del negocio los modelos resultantes del proceso de minería, ya que no funcionan por sí solos y al ser implantado el proceso de minería en la organización, este generará nuevas reglas de negocio y afectará la toma de decisiones.

### 13.2. Trabajo futuro

Tomando en cuenta la metodología planteada por el grupo para hacer replicable la aplicación de técnicas recientes a datos en la organización, un punto a considerar como mejora a futuro es el tema del respaldo de la información. Esto no fue considerado dentro de los alcances de este trabajo, sin embargo también es importante contar con alguna medida de recuperación de datos históricos ante alguna eventualidad que ocurra en la empresa.

Asimismo, se hace una recomendación respecto a la destinación de recursos por parte de la organización ya que, es necesario invertir en la limpieza y almacenamiento de los datos, ya que de estos depende el éxito del proyecto.

Finalmente, valdría la pena evaluar los atributos que se están registrando, y si hay mas atributos que pudieran ser valiosos, como la relación entre la víctima y el agresor.

## 14. RENDIMIENTO TRABAJO EN GRUPO

### 14.1. Porcentaje de participación

El trabajo en grupo fue equitativo, cada integrante participó y opinó en cada decisión al desarrollar cada fase de la metodología aplicada. Como resultado del trabajo en equipo se obtuvo un óptimo desempeño y productividad al momento de reunirnos.

Integrante	Porcentaje%
William Forero	25
Nicolás Méndez	25
Ailín Rojas	25
Santiago Salamanca	25

## 15. REFERENCIAS

- [1] Reports-archive.adm.cs.cmu.edu. (2018). [online] Available at: <http://reports-archive.adm.cs.cmu.edu/anon/ml2007/CMU-ML-07-114.pdf> [Accessed 16 Apr. 2018].
- [2] Disi.unal.edu.co. (2018). [online] Available at: [http://www.disi.unal.edu.co/profesores/eleonguz/cursos/md/presentaciones/Sesion13\\_validation\\_Clustering.pdf](http://www.disi.unal.edu.co/profesores/eleonguz/cursos/md/presentaciones/Sesion13_validation_Clustering.pdf) [Accessed 11 Apr. 2018].
- [3] <https://mariadearteaga.files.wordpress.com/2016/05/discovery-complex-anomalous1.pdf>
- [4] [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S1794-31082011000200007](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-31082011000200007)
- [5] [https://www.policia.gov.co/sites/default/files/exploracion\\_de\\_la\\_violencia\\_sexual\\_en\\_la\\_ciudad\\_de\\_bogota\\_una\\_aplicacion\\_de\\_tecnicas\\_de\\_mineria\\_de\\_datos.html](https://www.policia.gov.co/sites/default/files/exploracion_de_la_violencia_sexual_en_la_ciudad_de_bogota_una_aplicacion_de_tecnicas_de_mineria_de_datos.html)