

Índice

- 1. Introducción
- 2. Entendimiento del negocio
 - 2.1. Determinar objetivos del negocio
 - 2.1.1. Contexto
 - 2.1.2. Objetivos del negocio
 - 2.1.3. Criterios de éxito del negocio
 - 2.2. Análisis de la situación
 - 2.2.1. Inventario de recursos
 - 2.2.2. Requerimientos
 - 2.2.3. Presunciones
 - 2.2.4. Restricciones
 - 2.2.5. Riesgos y contingencias
 - 2.2.6. Costos y beneficios
 - 2.3. Objetivos de la minería de datos
- 3. Comprensión de datos
 - 3.1. Recolección inicial de datos
 - 3.2. Descripción de los datos
 - 3.3. Exploración de datos
 - 3.4. Verificación de la calidad de los datos
- 4. Preparación de datos
 - 4.1. Selección de datos
 - 4.2. Limpieza de datos
 - 4.3. Construcción de datos
 - 4.4. Integración de datos
 - 4.5. Formato de datos
- 5. Modelado
 - 5.1. Selección de técnicas de modelado
 - 5.2. Parámetros de ajuste
 - 5.3. Descripción de modelos
 - 5.3.1. Relevancia de atributos
 - 5.3.3. Reglas para la clasificación
 - 5.3.4. Resultado de red bayesiana
- 6. Conclusiones

1. Introducción

Cada año se cometen millones de actos criminales en todo el mundo. La inseguridad es un tema diario en muchos lugares del mundo. Estados Unidos es uno de los países con mayor tasa de criminalidad entre las potencias mundiales. Es importante que se pueda lograr la prevención y disminución de los crímenes. En este trabajo se analizarán una serie de datos de los Estados Unidos sobre diferentes crímenes y actos delictivos producidos en distintos contextos poblacionales, pudiendo analizar cuales de estos factores poblacionales son determinantes en la ocurrencia de crímenes.

2. Entendimiento del negocio

2.1. Determinar objetivos del negocio

2.1.1. Contexto

Los datos contienen una combinación de información socio-económica recolectada por Michael Redmond, en la universidad La Salle, de Filadelfia en el año 2009. Las fuentes de información son el censo de 1990 de los Estados Unidos, estadísticas policiales a partir de la encuesta US LEMAS de 1990 y el registro de crímenes de 1995. Todas estas publicaciones son ofrecidas al público y mantenidas por el Departamento de Comercio y el Departamento de Justicia de los Estados Unidos, más específicamente la Oficina de Censos y la Oficina de Estadísticas de Justicia respectivamente dentro de dichos departamentos.

2.1.2. Objetivos del negocio

En base a los datos recolectados de crímenes y sus características, poder prevenir futuros crímenes por medio de diferentes medios como mayores controles policiales o campañas de concientización.

2.1.3. Criterios de éxito del negocio

Se espera poder disminuir la cantidad de delitos ocurridos, mediante una mejor prevención y gestión de los recursos policiales disponibles.

2.2. Análisis de la situación

2.2.1. Inventario de recursos

Los recursos con los que se cuentan son:

- Personal especializado en minería de datos.
- Material digital con la información socio-económica y el historial de crímenes distribuidos por zonas.
- Hardware necesario para correr los algoritmos.
- Herramientas de minería de datos:
 - NNClass para clasificación.
 - CTree para inducción.
 - Elvira para algoritmos de red bayesiana.

2.2.2 Requerimientos

- Acceso irrestricto a los datos del censo.
- Acceso irrestricto al historial de crímenes.

2.2.3 Presunciones

- Se suponen datos sin margen de error
- Se supone que no hubo cambios en la forma de clasificar

2.2.4 Restricciones

- Se presenta una limitación a raíz de la información obtenida de la encuesta LEMAS del departamento de la policía, que consta con únicamente un registro de 100 oficiales más un recuento de pequeños departamentos.
- Se debieron dejar afuera numerosas comunidades puesto que no se contaba con toda la información necesaria.
- Algunas de las estadísticas sobre violaciones presentaron controversias, por lo que el recuento de las mismas no es exacto en algunas comunidades. Esto causa que algunos de estos valores no sean del todo confiables debido a la falta de denuncias y por lo tanto no se pueden incluir en el conjunto de datos.
- Se cuenta con los datos del censo de 1990 únicamente.

2.2.5 Riesgos y contingencias

Riesgo	Contingencia
La información obtenida no se encuadra con la	Si la información no es concluyente el proyecto deberá ser retomado con un historial depurado envergadura para poder construir información en base a más registros.

realidad.	
La información obtenida es de baja calidad.	Las conclusiones a las cuales la realidad pero tienen una calidad tan mediocre que no son utilizables. De nuevo, proyecto con un historial de casos depurado y en lo posible con más casos.
La información obtenida está desactualizada debido a la antigüedad de los datos	

2.2.6 Costos y beneficios

Los costos del presente proyecto corresponden tanto a los recursos materiales como humanos. Las herramientas de software son gratuitas, por lo que únicamente se requiere una computadora de prestaciones normales. Además, se requiere el trabajo de especialistas en minería de datos para poder realizar todas las tareas necesarias para el análisis y conclusión de los datos.

Teniendo en cuenta que los resultados del presente estudio podrán ayudar en la prevención y disminución de crímenes, el beneficio es muy alto, respecto del bajo costo.

2.3 Objetivos de la minería de datos

Determinar los atributos socio-económicos más relevantes para poder clasificar las causas de los crímenes y así poder llevar adelante campañas y medidas policiales que ayuden a prevenir el delito en distintas zonas urbanas de características similares.

3. Comprensión de datos

3.1. Recolección inicial de datos

El conjunto de datos como ya se ha mencionado anteriormente fue recolectado en el año 2009 y proviene de las bases de datos del Departamento de Justicia y de Comercio de los Estados Unidos¹.

3.2. Descripción de los datos

El conjunto de datos consta de 2215 registros, con 147 atributos:

Propiedad	Descripción
communityname	Nombre de la comunidad
state	Estado de EEUU (2 letras de la abreviación postal)
countyCode	Código numérico del condado
communityCode	Código numérico de la comunidad
fold	Número de veces para validaciones cruzadas no aleatorias, potencialmente útil para depurar pruebas de pareo
population	Población de la comunidad
householdsize	Personas promedio por hogar
racepctblack	Porcentaje de personas afroamericanas
racePctWhite	Porcentaje de personas caucásicas
racePctAsian	Porcentaje de personas de descendencia asiática
racePctHisp	Porcentaje de personas de descendencia hispana
agePct12t21	Porcentaje de personas entre los 12-21 años
agePct12t29	Porcentaje de personas entre los 12-29 años

¹ <https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime+Unnormalized>

agePct16t24	Porcentaje de personas entre los 16-24 años
agePct65up	Porcentaje de personas con 65 o más años
numbUrban	Número de personas viviendo en áreas urbanas
pctUrban	Porcentaje de personas viviendo en áreas urbanas
medIncome	Ingreso medio por hogar
pctWWage	Porcentaje de hogares con ingresos por sueldos o salarios en 1989
pctWFarmSelf	Porcentaje de hogares con ingresos por granja o autoempleados en 1989
pctWInvInc	Porcentaje de hogares con ingresos por inversiones o alquiler en 1989
pctWSocSec	Porcentaje de hogares con ingresos por seguridad social en 1989
pctWPubAsst	Porcentaje de hogares con ingresos por asistencia pública en 1989
pctWRetire	Porcentaje de hogares con ingresos por jubilación en 1989
medFamInc	Ingreso familiar medio
perCapInc	Ingreso per capita
whitePerCap	Ingreso per capita para personas caucásicas
blackPerCap	Ingreso per capita para afroamericanos
indianPerCap	Ingreso per capita para nativos americanos
AsianPerCap	Ingreso per capita para personas de descendencia asiática
OtherPerCap	Ingreso per capita para personas de otra descendencia
HispPerCap	Ingreso per capita para personas de descendencia hispana
NumUnderPov	Número de personas bajo el nivel de pobreza

PctPopUnderPov	Porcentaje de personas bajo el nivel de pobreza
PctLess9thGrade	Porcentaje de personas de 25 años o más, con un nivel de educación inferior al 9° grado
PctNotHSGrad	Porcentaje de personas de 25 años o más, que no se graduaron de la secundaria
PctBSorMore	Porcentaje de personas de 25 años o más, con una licenciatura o superior
PctUnemployed	Porcentaje de personas de 16 años o más desempleados
PctEmploy	Porcentaje de personas de 16 años o más empleados
PctEmplManu	Porcentaje de personas de 16 años o más, empleados en la industria manufacturera
PctEmplProfServ	Porcentaje de personas de 16 años o más, empleados en servicios profesionales
PctOccupManu	Porcentaje de personas de 16 años o más, empleados en la industria manufacturera
PctOccupMgmtProf	Porcentaje de personas de 16 años o más, empleados en gestión u ocupaciones profesionales
MalePctDivorce	Porcentaje de hombres divorciados
MalePctNevMarr	Porcentaje de hombres que nunca se casaron
FemalePctDiv	Porcentaje de mujeres divorciadas
TotalPctDiv	Porcentaje de población divorciada
PersPerFam	Personas promedio por familia
PctFam2Par	Porcentaje de familias (con hijos) encabezadas por 2 padres
PctKids2Par	Porcentaje de niños en casa de familia con 2 padres
PctYoungKids2Par	Porcentaje de niños con 4 años o menos con 2 casas de los padres
PctTeen2Par	Porcentaje de chicos entre 12-17 años con 2 casas de los padres

PctWorkMomYoungKids	Porcentaje de madres trabajadoras con hijos de 6 años o menos
PctWorkMom	Porcentaje de madres trabajadoras con chicos de 6 años o menos
NumKidsBornNeverMar	Porcentaje de madres trabajadoras con chicos de 18 años o menos
PctKidsBornNeverMar	Porcentaje de niños nacidos de padres no casados
NumImmig	Porcentaje de niños nacidos en el extranjero
PctImmigRecent	Porcentaje de inmigrantes que inmigraron hace menos de 3 años
PctImmigRec5	Porcentaje de inmigrantes que inmigraron hace menos de 5 años
PctImmigRec8	Porcentaje de inmigrantes que inmigraron hace menos de 8 años
PctImmigRec10	Porcentaje de inmigrantes que inmigraron hace menos de 10 años
PctRecentImmig	Porcentaje de la población que inmigraron hace menos de 3 años
PctReclImmig5	Porcentaje de la población que inmigraron hace menos de 5 años
PctReclImmig8	Porcentaje de la población que inmigraron hace menos de 8 años
PctReclImmig10	Porcentaje de la población que inmigraron hace menos de 10 años
PctSpeakEnglOnly	Porcentaje de personas que hablan solamente ingles
PctNotSpeakEnglWell	Porcentaje de personas que no hablan ingles bien
PctLargHouseFam	Porcentaje de hogares familiares grandes (6 o mas)
PctLargHouseOccup	Porcentaje de hogares ocupados grandes (6 o mas)
PersPerOccupHous	Personas promedio por hogar
PersPerOwnOccHous	Personas promedio por hogares ocupados por el

	propietario
PersPerRentOccHous	Personas promedio por hogares ocupados por inquilinos
PctPersOwnOccup	Porcentaje de personas en hogares ocupados por el propietario
PctPersDenseHous	Porcentaje de personas en hogares densamente ocupados (más de 1 persona por cuarto)
PctHousLess3BR	Porcentaje de viviendas con menos de 3 dormitorios
MedNumBR	Número medio de dormitorios
HousVacant	Número de hogares vacantes
PctHousOccup	Porcentaje de viviendas ocupadas
PctHousOwnOcc	Porcentaje de viviendas ocupadas por el propietario
PctVacantBoarded	Porcentaje de viviendas desocupadas tapiadas
PctVacMore6Mos	Porcentaje de viviendas desocupadas que no estuvieron ocupadas por más de 6 meses
MedYrHousBuilt	Media anual de viviendas construídas
PctHousNoPhone	Porcentaje de casas ocupadas sin telefono
PctWOFullPlumb	Porcentaje de casas ocupadas sin instalaciones de fontanería completas
OwnOccLowQuart	Viviendas ocupadas por el propietario - valor cuartil inferior
OwnOccMedVal	Viviendas ocupadas por el propietario - valor medio
OwnOccHiQuart	Viviendas ocupadas por el propietario - valor cuartil superior
OwnOccQrange	Viviendas ocupadas por el propietario - diferencia entre valores del cuartil inferior y el cuartil superior
RentLowQ	Viviendas rentadas - valor cuartil inferior
RentMedian	Viviendas rentadas - valor medio
RentHighQ	Viviendas rentadas - valor cuartil superior

RentQrange	Viviendas rentadas - diferencia entre valores del cuartil inferior y el cuartil superior
MedRent	Renta bruta media
MedRentPctHousInc	Renta bruta media como porcentaje de ingreso del hogar
MedOwnCostPctInc	Renta bruta media como porcentaje de ingreso del hogar - para dueños con hipoteca
MedOwnCostPctIncNoMtg	Costo medio de los propietarios como porcentaje de ingreso del hogar - para dueños sin hipoteca
NumInShelters	Número de personas en refugios para desamparados
NumStreet	Número de personas sin hogar contadas en las calles
PctForeignBorn	Porcentaje de personas extranjeras
PctBornSameState	Porcentaje de personas nacidas en el mismo estado de residencia actual
PctSameHouse85	Porcentaje de personas viviendo en mismo hogar desde 1985
PctSameCity85	Porcentaje de personas viviendo en la misma ciudad desde 1985
PctSameState85	Porcentaje de personas viviendo en el mismo estado desde 1985
LemasSwornFT	Número de policías a tiempo completo
LemasSwFTPerPop	Policías a tiempo completo por cada 100 mil habitantes
LemasSwFTFieldOps	Número de policías a tiempo completo en operaciones de campo (en la calle, no administrativos)
LemasSwFTFieldPerPop	Número de policías a tiempo completo en operaciones de campo (en la calle, no administrativos) por cada 100 mil habitantes
LemasTotalReq	Cantidad de solicitudes para policía
LemasTotReqPerPop	Cantidad de solicitudes para policía por cada 100 mil habitantes
PolicReqPerOffic	Cantidad de solicitudes para policía por cada oficial de

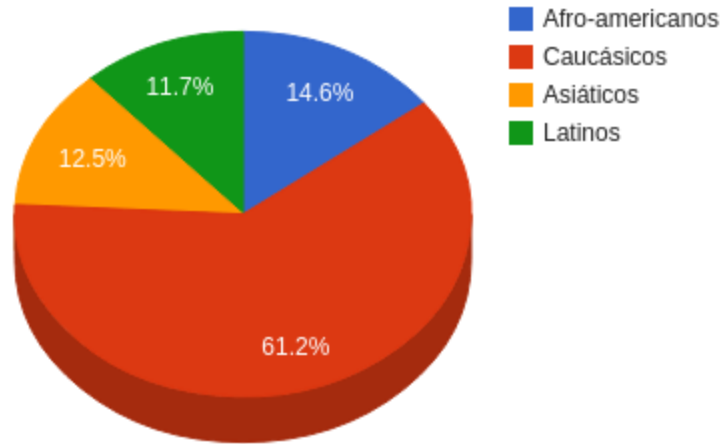
	policía
PolicPerPop	Cantidad de policías per cápita
RacialMatchCommPol	Medida de la paridad racial entre la comunidad y la fuerza policial. Valores altos indican proporciones similares
PctPolicWhite	Porcentaje de policías caucásicos
PctPolicBlack	Porcentaje de policías afro-americanos
PctPolicHisp	Porcentaje de policías latinos
PctPolicAsian	Porcentaje de policías asiáticos
PctPolicMinor	Porcentaje de policías que son minoría (de cualquier tipo)
OfficAssgnDrugUnits	Cantidad de oficiales asignados a la unidad anti-drogas
NumKindsDrugsSeiz	Número de diferentes tipos de drogas incautadas
PolicAveOTWorked	Promedio de tiempo extra trabajo por los policías
LandArea	Superficie total de la comunidad
PopDens	Densidad de población promedio
PctUsePubTrans	Porcentaje de uso del transporte público
PolicCars	Cantidad de coches policías
PolicOperBudg	Presupuesto policial anual para operaciones de campo
LemasPctPolicOnPatr	Policías a tiempo completo en patrullaje
LemasGangUnitDeploy	Unidad de pandillas desplegadas
LemasPctOfficDrugUn	Porcentaje de policías asignados a la unidad anti-drogas
PolicBudgPerPop	Presupuesto policial per cápita
murders	Cantidad de asesinatos registrados
murdPerPop	Cantidad de asesinatos registrados per cápita
rapes	Cantidad de violaciones registrados

rapesPerPop	Cantidad de violaciones registrados per cápita
robberies	Cantidad de robos con violencia registrados
robbPerPop	Cantidad de robos con violencia registrados per cápita
assaults	Cantidad de asaltos registrados
assaultPerPop	Cantidad de asaltos registrados per cápita
burglaries	Cantidad de robos con violación a la propiedad privada registrados
burglPerPop	Cantidad de robos con violación a la propiedad privada registrados per cápita
larcenies	Cantidad de hurtos registrados
larcPerPop	Cantidad de hurtos registrados per cápita
autoTheft	Cantidad de robos de automóviles registrados
autoTheftPerPop	Cantidad de robos de automóviles registrados per cápita
arsons	Cantidad de incendios provocados registrados
arsonsPerPop	Cantidad de incendios provocados registrados per cápita
ViolentCrimesPerPop	Cantidad de crímenes de carácter violento per cápita
nonViolPerPop	Cantidad de crímenes de carácter no violento per cápita

3.3. Exploración de datos

Estudiando los datos obtenidos durante la recolección de datos, se puede ver que existen algunas tendencias. En primer lugar, si observamos la distribución según el origen de las personas, vemos que existe una clara mayoría de caucásicos, mientras que el porcentaje de afroamericanos, latinos y asiáticos es menor, pero igualmente distribuido.

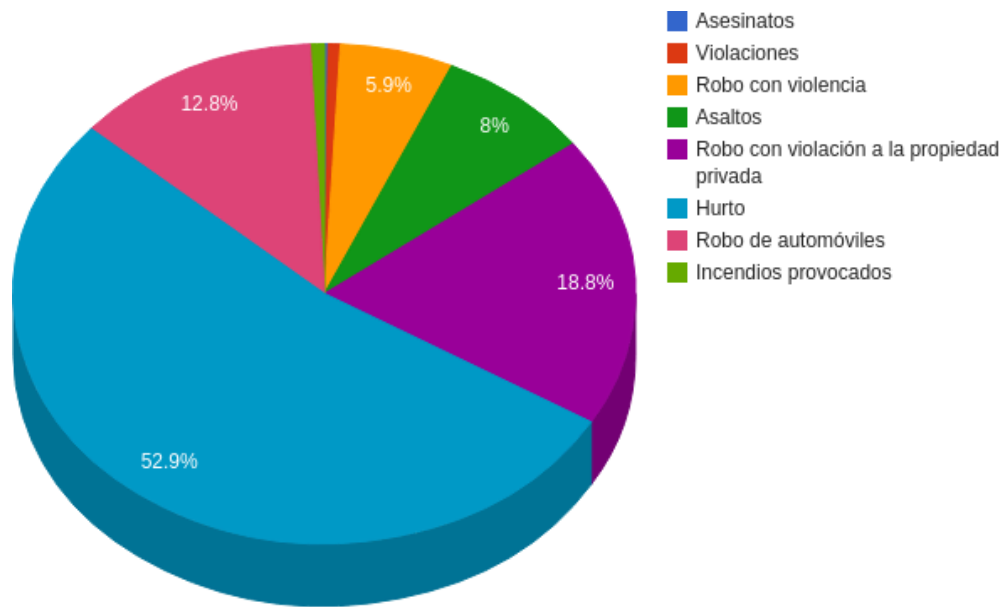
Distribución de población según su raza



Esto, nos condicionará más adelante para poder armar las categorías correspondientes para los distintos atributos.

Por otro lado, los registros muestran los delitos clasificados según distintos tipos:

Tipo de crimen	Cantidad
Asesinatos	17.199
Violaciones	56.289
Robo con violencia	526.826
Asaltos	719.015
Robo con violación a la propiedad privada	1.683.856
Hurto	4.728.436
Robo de automóviles	1.142.924
Incendios provocados	65.648



Lo que se puede observar de esta distribución, es que aproximadamente la mitad de los delitos corresponden a hurtos simples (sin violencia), otro cuarto corresponden a otros crímenes no violentos y únicamente el último cuarto corresponde a los crímenes que nos interesa estudiar.

3.4. Verificación de la calidad de los datos

Teniendo en cuenta las fuentes de las cuales fueron obtenidas los datos, se supone que los mismo se presentan sin errores. Sin embargo, existen numerosos registros incompletos, puesto que el conjunto de datos se obtuvo a partir de la cruza de diversos conjuntos de datos que no contenían toda la información necesaria para obtener toda la información.

4.Preparación de datos

4.1. Selección de datos

A partir de los datos disponibles, se decidió utilizar únicamente los siguientes atributos para poder aplicar las herramientas de forma adecuada, y cumplir los objetivos de negocios:

- Population
- PctLess9thGrade

- Householdsize
- RacePctBlack
- RacePctWhite
- RacePctAsian
- RacePctHisp
- MedIncome
- PctPopUnderPov
- PctNotHSGrad
- PctUnemployed
- PctHousOccup
- PctHousOccup
- PctForeignBorn
- PopDens
- ViolentCrimesPerPop

4.2. Limpieza de datos

Se eliminaron del set de datos original aquellos registros cuyos atributos característicos (seleccionados en el paso anterior) se encontraban incompletos.

4.3. Construcción de datos

Se hizo una redistribución de los valores de los atributos seleccionados para poder tener una mejor distribución de los datos. Los atributos y sus valores quedaron de la siguiente manera:

Propiedad	Categorías
population	1: menos de 10.000 habitantes 2: entre 10.000 y 100.000 habitantes 3: entre 100.000 y 200.000 habitantes 4: más 200.000 habitantes
householdsize	1: menos de 2.5 personas por hogar promedio. 2: más de 2.5 personas por hogar promedio.
racepctblack	1: 0% - 3% de personas de raza negra 2: 3% - 9% de personas de raza negra 3: 9% - 100% de personas de raza negra
racePctWhite	1: 0% - 7% de personas de raza caucásica 2: 7% - 85% de personas de raza caucásica 3: 8% - 95% de personas de raza caucásica 4: 9% - 100% de personas de raza caucásica



racePctAsian	1: 0% - 1% de personas de origen asiatico 2: 1% - 3% de personas de origen asiatico 3: 3% - 100% de personas de origen asiatico
racePctHisp	1: 0% - 2% de personas de origen latino 2: 2% - 6% de personas de origen latino 3: 6% - 100% de personas de origen latino
medIncome	1: menos de 20.000 U\$S de promedio ingresos anuales 2: 20.000 - 33.000 U\$S de promedio ingresos anuales 3: 33.000 - 46.000 U\$S de promedio ingresos anuales 4: más 45.000 U\$S de promedio ingresos anuales
PctPopUnderPov	1: menos del 2% de población bajo la línea de pobreza 2: 2% - 11% de población bajo la línea de pobreza 3: 11% - 18% de población bajo la línea de pobreza 4: más del 18% de población bajo la línea de pobreza
PctLess9thGrade	1: menos del 3% de personas analfabetas 2: entre el 3% y 9% de personas analfabetas 3: entre el 9% y 15% de personas analfabetas 4: más del 15% de personas analfabetas
PctNotHSGrad	1: menos del 10% de personas semi-analfabetas 2: entre el 10% y 22% de personas semi-analfabetas 3: entre el 22% y 32% de personas semi-analfabetas 4: más 32% de personas semi-analfabetas
PctUnemployed	1: menos del 5% de desocupación 2: entre el 5% y 7.5% de desocupación 3: más 7.5% de desocupación
PctHousOccup	1: menos del 89.5% de propiedades ocupadas 2: entre 89.5% y 95% de propiedades ocupadas 3: más 95% de propiedades ocupadas
PctForeignBorn	1: menos del 3% de extranjeros 2: 3% - 9% de extranjeros 3: más del 9 de extranjeros
PopDens	1: menos de 1300 habitantes por km2 2: entre 1300 y 4100 habitantes por km2 3: más 4100 habitantes por km2
ViolentCrimesPerPop	1: Tasa de crímenes violentos BAJA 2: Tasa de crímenes violentos MEDIA

3: Tasa de crímenes violentos ALTA

4.4. Integración de datos

Debido a que está toda la información disponible ya fue integrada previamente, no hace falta integrar datos de distintas fuentes.

4.5. Formato de datos

No fue necesario hacer una modificación del formato ya que el set de datos original fue descargado en un archivo de Excel, el cual es el formato necesario para utilizar las herramientas de modelado.

5. Modelado

5.1. Selección de técnicas de modelado

En el proyecto se utilizarán tres tipos de modelo: una red backpropagation para determinar la relevancia de los atributos para clasificar, un algoritmo TDIDT para generar reglas de inducción para saber cómo se clasifican los candidatos y un clasificador Naive Bayes. La red de backpropagation será implementada usando la herramienta NNclass para poder analizar la significancia de los atributos a la hora de clasificar. El algoritmo TDIDT elegido es el C4.5 y se la implementará con la herramienta CTree para poder generar reglas mediante la inducción. La red bayesiana será implementada con la aplicación Elvira.

5.2. Parámetros de ajuste

La red de backpropagation se usó con 2 capas ocultas de neuronas, 200 ciclos para entrenar la red, el parámetro de aprendizaje con 0,9, un momentum de 0,1 y una variación absoluta del peso inicial en 0,7. Además, la red tiene los siguientes valores fijos establecidos:

Propiedad	Valor fijo
population	2
householdsize	2

Propiedad	Valor fijo
PctPopUnderPov	1
PctLess9thGrad	1

racepctblack	1
racePctWhite	1
racePctAsian	1
racePctHisp	1
medIncome	1

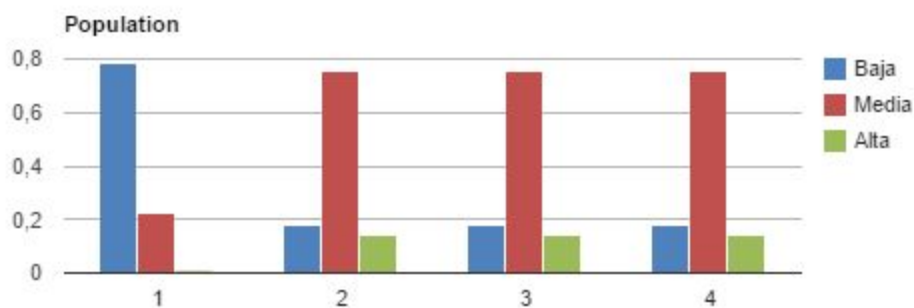
e	
PctNotHSGrad	1
PctUnemployed	1
PctHousOccup	1
PctForeignBorn	1
PopDens	2

5.3. Descripción de modelos

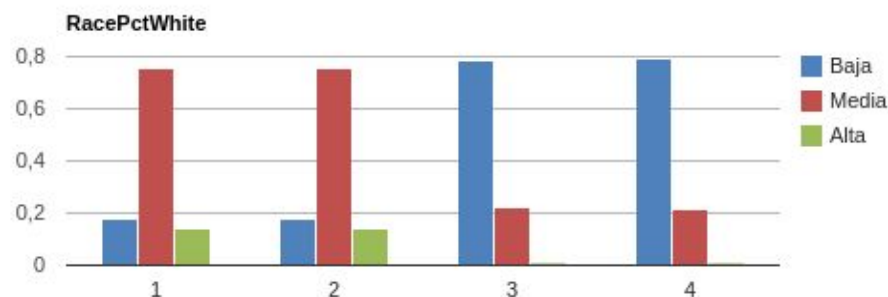
A continuación se mostrarán los resultados para cada uno de los modelos.

5.3.1 Relevancia de atributos

A continuación se separaron los atributos que en primera instancia parecieran ser relevantes para determinar el grado de crímenes de carácter violento:

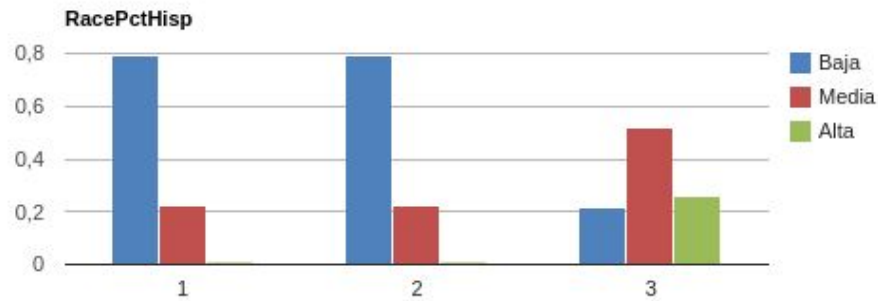


La población muestra ser relevante para el salto de categorías de 1 a 2 que casualmente corresponde al paso de un pueblo a una ciudad. Luego, no presenta cambios.

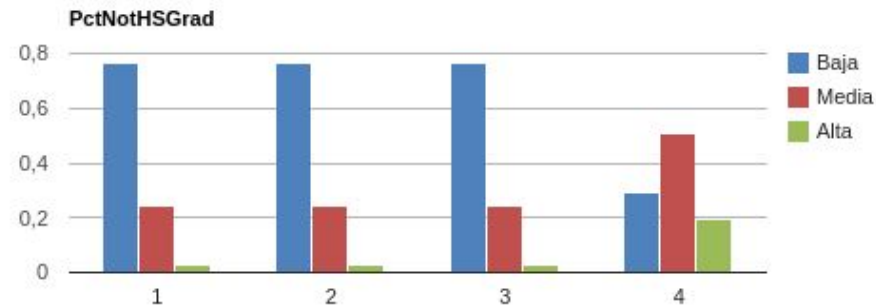


Se puede observar cómo a medida que el porcentaje de caucásicos aumenta respecto

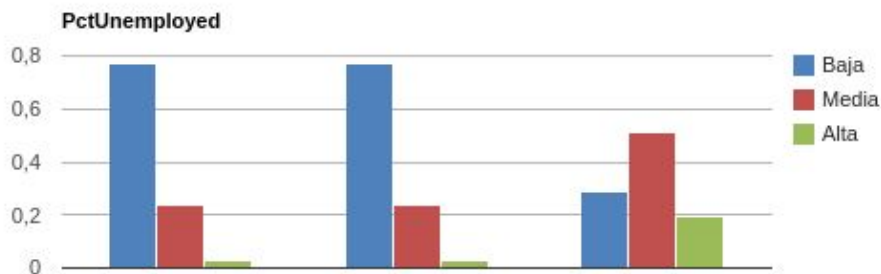
del total de la población, la tasa de delitos violentos disminuye.



Por el contrario, se puede ver estudiando los resultados para el atributo predictor RacePctHisp, que cuanto mayor es el porcentaje de población latina, tiende a aumentar la el nivel de violencia en los crímenes registrados.

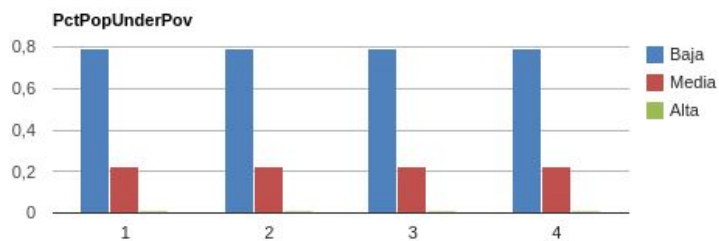
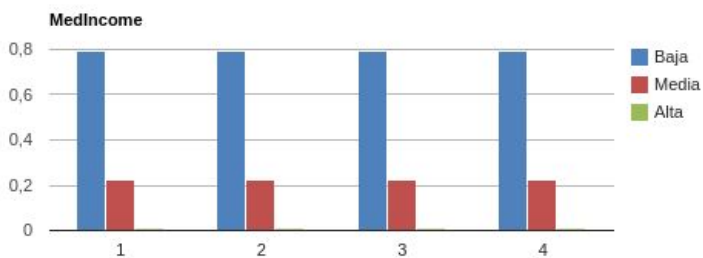
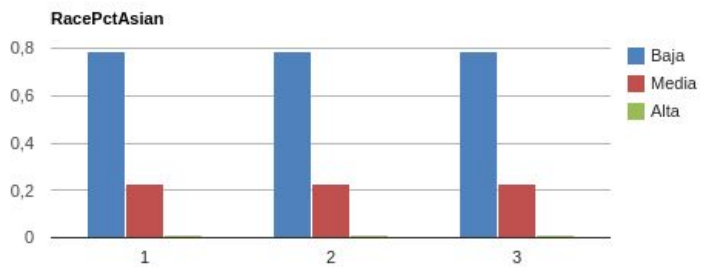
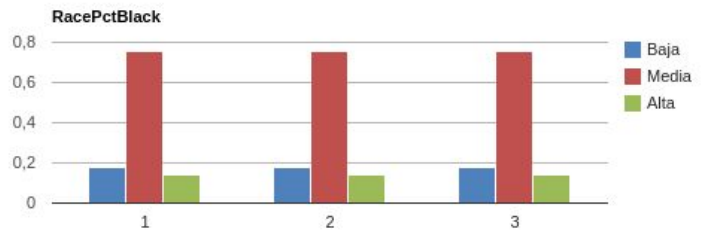


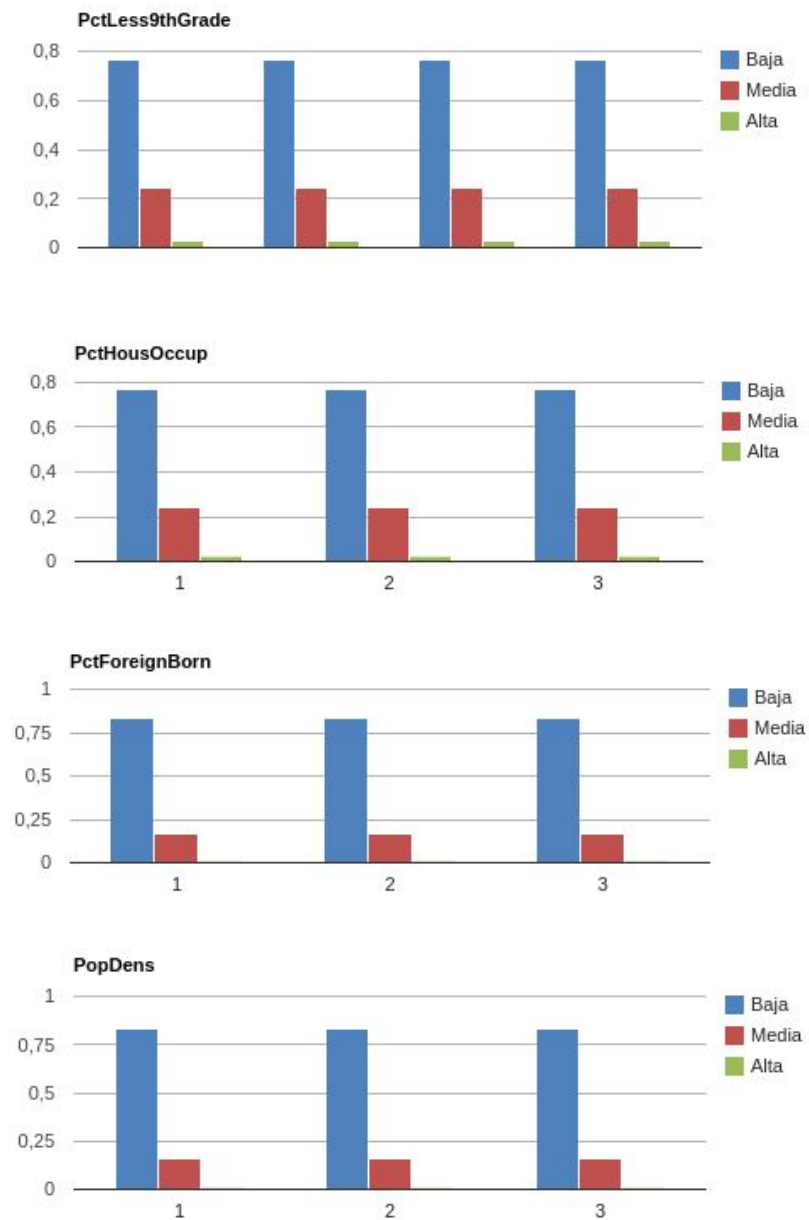
Otro factor decisivo es el nivel de población semi-analfabeta. Podemos ver que si bien para niveles de semi-analfabetismo inferiores a 32%, el nivel de crímenes violentos no parece verse afectado, cuando este lo supera los valores cambian tendiendo a niveles medios y altos de violencia.



Finalmente, el desempleo también es una de las causas por la cual se ve afectado el nivel de violencia en los delitos registrados. Al igual que en el caso anterior, para niveles de desempleo superiores al 7.5%, aumenta la violencia en los delitos.

Por otro lado, los siguientes atributos mostraron no ser relevantes para determinar por sí solos el nivel de violencia en los delitos:





5.3.3 Reglas para la clasificación

A continuación se mostrarán los resultados para la inducción de reglas:

Regla nro.	Definición
0	ViolentCrimesPerPop = baja
1	IF PctHousOccup = 3 AND racePctWhite = 4

	THEN ViolentCrimesPerPop = baja
2	IF racePctWhite = 1 THEN ViolentCrimesPerPop = alta
3	IF PctNotHSGrad = 2 THEN ViolentCrimesPerPop = baja
4	IF racePctWhite = 4 THEN ViolentCrimesPerPop = baja
5	IF medIncome = 4 THEN ViolentCrimesPerPop = baja
6	IF racePctHisp = 3 THEN ViolentCrimesPerPop = media
7	IF racePctWhite = 2 THEN ViolentCrimesPerPop = media
8	IF racePctHisp = 1 THEN ViolentCrimesPerPop = baja
9	IF PctUnemployed = 1 THEN ViolentCrimesPerPop = baja
10	IF medIncome = 3 THEN ViolentCrimesPerPop = baja
11	IF racePctBlack = 3 THEN ViolentCrimesPerPop = alta
12	IF PctLess9thGrade = 2 AND population = 1 THEN ViolentCrimesPerPop = baja
13	IF PctUnemployed = 1 AND racePctHisp = 1 THEN ViolentCrimesPerPop = baja

Regla nro.	Class	Lenght	Support	Confidence	Capture
0	baja	0	100,0%	41,5%	100,0%
1	baja	2	17,1%	87,0%	35,9%
2	alta	1	18,5%	64,2%	55,6%
3	baja	1	39,9%	53,4%	51,3%
4	baja	1	30,0%	78,4%	56,7%
5	baja	1	17,8%	75,9%	32,5%
6	media	1	30,7%	51,0%	42,1%
7	media	1	21,3%	54,6%	31,3%
8	baja	1	45,8%	54,6%	60,2%

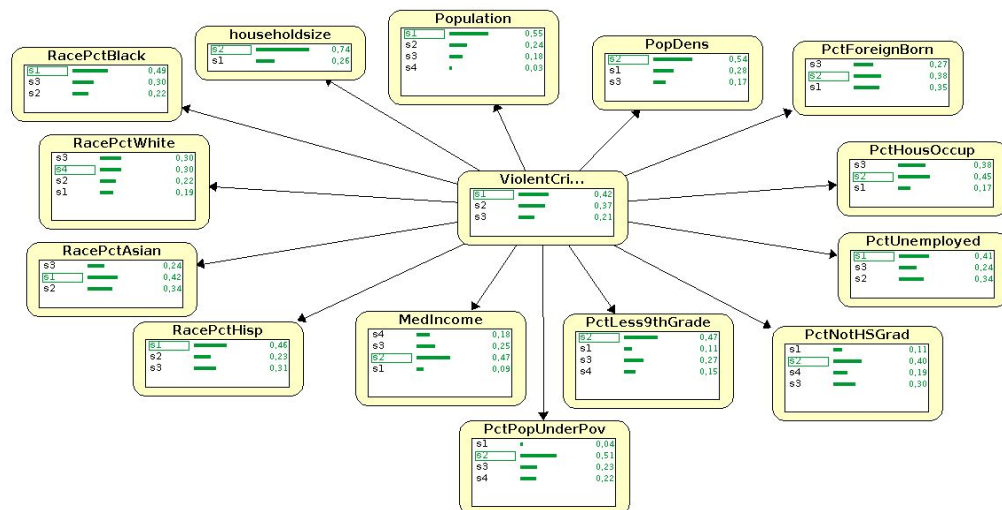
9	baja	1	41,8%	64,4%	64,8%
10	baja	1	25,4%	51,4%	31,5%
11	alta	1	29,9%	51,2%	71,8%
12	baja	2	25,1%	66,0%	39,9%
13	baja	2	20,7%	78,0%	39,0%

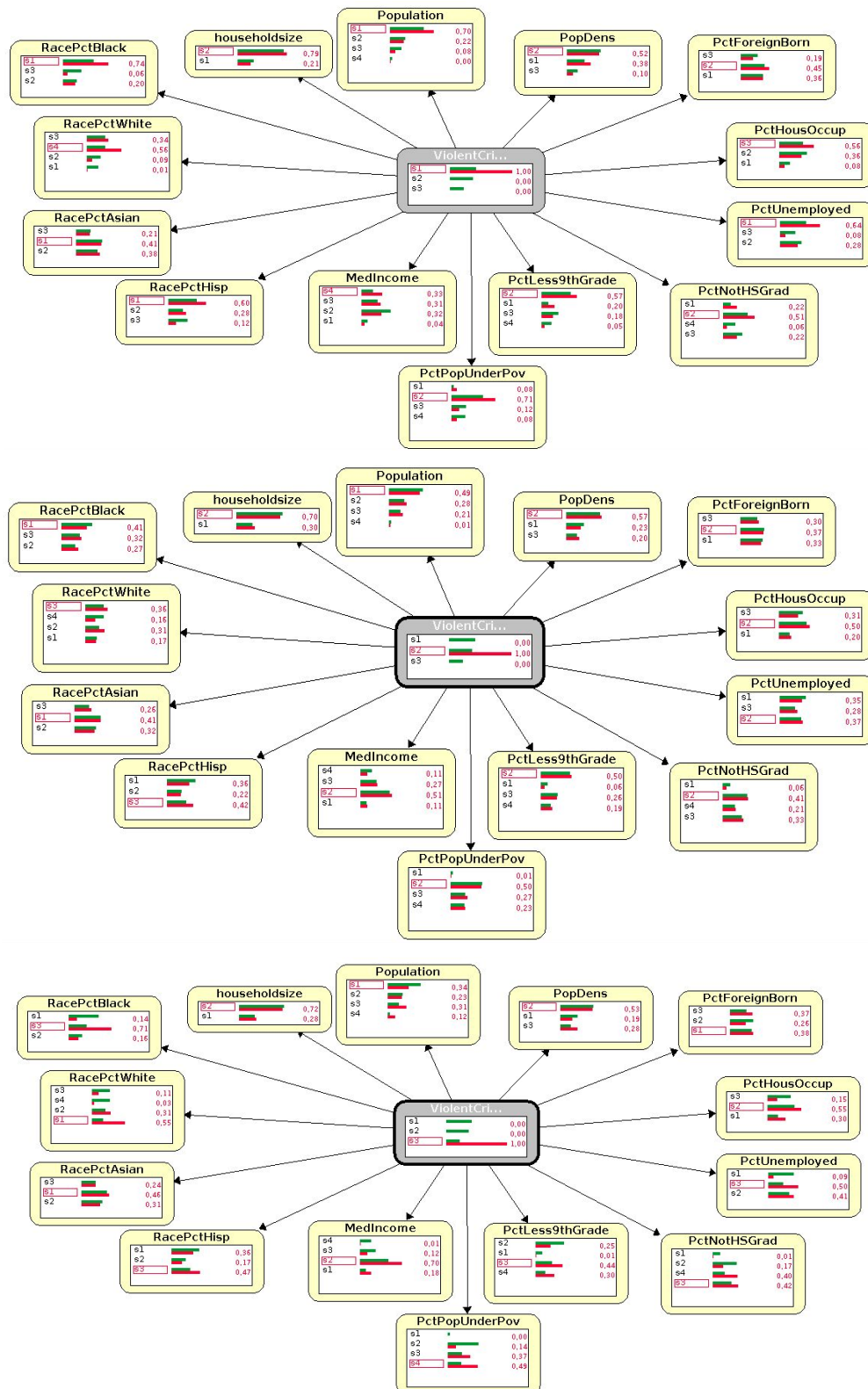
Se puede ver como la mayoría de las reglas determinan niveles bajos de violencia, y esto está fuertemente relacionado con que existe una mayor tendencia a dichos niveles en los registros que se tienen. Por otro lado, vemos una clara concordancia con los valores obtenidos en el primer modelo, con las redes neuronales, puesto que son los mismos atributos (los de mayor importancia) que se identificaron como decisivos los que condicionan las distintas reglas.

Tanto el porcentaje de individuos caucásicos como latinos definen el nivel de violencia dependiendo de la categoría de los mismos.

Otro atributo que aparece es el de los ingresos medios, y muestra que para zonas con altos ingresos los niveles de violencia son bajos, al igual que para los niveles bajos de desempleo.

5.3.4 Resultado de red bayesiana





Mediante esta herramienta, se nos permite ver la influencia de cada atributo en cada

una de las categorías de nuestra variable a predecir.

Tras una evaluación de los resultados vemos como algunos atributos no muestran variaciones en la influencia para determinar el nivel de violencia, mientras que otros sí lo hacen para alguna/s categoría/s del nivel de violencia.

Entre estos últimos, vemos que la población de la zona tiende a oponerse al aumento de la violencia en los crímenes, sin embargo debe tenerse cuidado al analizar dicho atributo puesto que la última categoría del mismo no presenta casi registros. Además se puede ver que una baja población influye fuertemente en un resultado bajo en el porcentaje de delitos violentos.

Otro atributo que muestra tener un impacto interesante es el nivel de desempleo, al igual que el anterior, mantiene una relación inversa con ViolentCrimePerCpt. También podemos ver como se invierte la relación entre la primera de sus categorías y las otras dos para las categorías Media y Alta.

El porcentaje de personas analfabetas, semi-analfabetas y el nivel de pobreza muestran el mismo tipo de comportamientos, con una relación directa con respecto a la cantidad de crímenes violentos que se registran.

Con respecto a los ingresos medios percibidos, se evidencia el mismo comportamiento que se obtuvo con el uso de las herramientas previas.

Por último las otras variables que presentan interés son las del porcentaje de personas según su raza. A diferencia de los otros modelos, en este además podemos observar que el incremento de la población afroamericana suele estar acompañada de un aumento en los delitos violentos por persona registrados.

6. Conclusiones

Finalmente, con toda la información obtenida a partir del análisis de los datos por medio de las distintas herramientas, se pudieron obtener varias y diferentes conclusiones acerca de qué factores son relevantes y deben ser tenidos en cuenta a la hora de prevenir e intentar disminuir los delitos de carácter violento que ocurren en una zona.

Si bien los datos corresponden a los Estados Unidos y su censo, los patrones de comportamiento según el origen de las personas suele mantenerse debido a la tradición y el traspaso de cultura, por lo que los resultados pueden ser tomados como

referencia para otras naciones, teniendo en cuenta siempre las variaciones que puedan existir en cuanto a la población y la forma de vida.

En primer lugar, es interesante destacar que de los atributos seleccionados para llevar adelante el análisis, sólo una pequeña parte mostró tener una influencia real sobre la cantidad de crímenes violentos por persona. Y que los mismos no siempre eran determinantes por sí solos, sino que la combinación lo era.

En cuanto a los resultados concretos que se pudieron obtener, se vé claramente que la cantidad de habitantes (población), sólo es un factor en el salto de pueblo a ciudad, donde en los primeros la cantidad de delitos violentos es muy inferior a las ciudades. Sin embargo, el nivel de ingresos medios de la población (para los distintos tamaños de ciudad) mostró tener importancia ya que mayores los ingresos, menor la cantidad de delitos violentos por persona.

Desde el punto de vista educativo se observó un resultado no tan previsto desde un principio, que es que el nivel de semi-analfabetismo mostró tener más influencia para determinar los niveles altos de violencia en los delitos que el nivel de analfabetismo y el de pobreza. Este mismo comportamiento corresponde también al nivel de desempleo.

Con respecto al porcentaje de personas caucásicas, latinas, afroamericanas y asiáticas se llegó a la conclusión de que cuanto mayor es el porcentaje de personas latinas y/o afroamericanas, mayor es la tendencia a tener mucha cantidad de delitos violentos, contrario a la población caucásica. Por otro lado, la población asiática no mostró influencia alguna. Sin embargo, como se vió en la exploración de los datos, la distribución de personas de cada etnia no era semejante, por lo que no es posible decir que estos últimos resultados son concluyentes, aunque sí es posible afirmar que el porcentaje de latinos y afroamericanos tienen una mayor influencia sobre la cantidad de delitos violentos que el porcentaje de personas de origen asiático.

Por último, para lograr los objetivos que se propusieron al principio de este trabajo, para prevenir y disminuir los delitos de carácter violento de una población, es necesario prestar principal atención y aumentar el control policial en ciudades y particularmente en zonas cuyos ingresos promedio no sean muy elevados y el nivel de educación general sea bajo. Y además en zonas donde haya una distribución de personas de diferentes razas bastante heterogénea.