

# UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

Departamento Académico de Estadística e  
Informática



Facultad de Economía y Planificación

**Curso:**

- Base de datos II

**Profesor:**

- Soto Rodríguez, Iván

**Alumnos:**

- Waldo Jose Gomez Servan
- Alvaro Cesar Tauma Salvador
- Nayra Liz Limaymanta Jorge
- Aron Salvador Uriarte Zaga
- Nestor lopez lazaro

2021

## **1. RESUMEN.**

El enfoque de Seguridad Humana se relaciona con la teoría del desarrollo humano. Ambas temáticas la seguridad humana y el desarrollo humano- tienen al Programa de las Naciones Unidas (PNUD) como uno de sus principales promotores. Lo que se resalta en el enfoque de la Seguridad Humana es entender la problemática de la seguridad desde diversas dimensiones en que se desarrolla la vida de las personas y como estas dimensiones son obstáculos u oportunidades para dicha seguridad. En la Seguridad Humana, la novedad consiste en entender los problemas de la seguridad a partir de las limitaciones que la persona o la comunidad tiene para vivir sin temor.

Este trabajo tiene como finalidad identificar en que distritos del país hay mayor cantidad delincuencia y también ver de que manera se puede ayudar a disminuir este problema de seguridad, definiendo estrategias para el objetivo que es reducir la inseguridad ciudadana a favor de la población.

## **2. INTRODUCCIÓN:**

### **2.1. PROBLEMA**

El concepto de seguridad está orientado en función de la protección frente a riesgos y amenazas, y no solo en función de riesgos-amenazas específicos; el término ciudadana no se ha producido inicialmente desde el concepto de lo urbano-ciudad, sino desde la idea del ciudadano, entendido como sujeto de derechos y en un enfoque de desarrollo humano.

La seguridad ciudadana se entiende como la protección de la vida, integridad, el patrimonio de las personas frente a los riesgos y las amenazas que las puedan poner en peligro.

La inseguridad ciudadana es hoy uno de los principales problemas que preocupan al país. Los índices de delitos cometidos en las calles, que habían bajado drásticamente producto de las medidas de confinamiento dispuestas para frenar la expansión del covid-19, han comenzado de nuevo a subir, lo que causa inquietud entre los peruanos.

Según los datos estadísticos obtenidos del INEI (2021), en el semestre noviembre 2020 – abril 2021, el 18.6% de la población de 15 a más años de edad fueron víctimas de algún hecho delictivo; esto se dio a conocer en el informe técnico Estadísticas de Seguridad ciudadana, elaborado con los resultados de la Encuesta Nacional de Programas Presupuestales.

En el semestre de análisis, el 21.8% de la población de 15 a 29 años de edad, padeció de algún hecho delictivo, la población de 30 a 44 años de edad el

22.4% también fue víctima, la población de 45 a 64 años de edad el 17.6% y los que tienen de 65 años a más el 7.6%;

En el semestre de estudio, de 100 habitantes del área urbana del país 9 fueron víctimas de robo, dinero, cartera o celular. En las ciudades de 20 000 a más habitantes, fueron 11 de cada 100 habitantes; en Lima Metropolitana y la Provincia Constitucional del Callao a 12 de cada 100 habitantes y en centros poblados urbanos de entre 2000 y menos de 20 000 habitantes, afectó a 5 de cada 100 habitantes.

El 82,3% de la población del área urbana a nivel nacional percibe que en los próximos 12 meses puede ser víctima de algún hecho delictivo.

En el semestre noviembre 2020-abril 2021, el 82,3% de la población de 15 y más años de edad percibió que en los próximos 12 meses puede ser víctima de algún hecho delictivo. En ciudades de 20 mil a más habitantes 86,1% y en centros poblados urbanos entre 2 mil y menos de 20 mil habitantes 72,5%. En Lima Metropolitana y la Provincia Constitucional del Callao ascendió a 90,3%.

## **2.2.JUSTIFICACIÓN DEL TRABAJO**

La seguridad ciudadana representa uno de los aspectos más importantes que determina nuestra calidad de vida, reduce la vulnerabilidad de poblaciones de riesgo y fomenta un entorno más favorable a la inversión que permita generar condiciones para el desarrollo sostenible.

Por esta razón se va a elaborar un análisis BI sobre el problema de seguridad en donde se analizarán los factores de riesgo causas y medidas preventivas necesarias

## **2.3.FORMULACIÓN DEL PROBLEMA**

- ¿Qué percepción de inseguridad tienen las familias en su comunidad y como ha afectado sus hábitos de comportamiento, Como afecta la calidad de vida en el Perú?
- ¿Cómo podríamos devolver la confiabilidad en la seguridad peruana con la ayuda de los datos obtenidos en encuestas a nivel nacional?

## **2.4.OBJETIVOS GENERALES Y ESPECIFICOS**

**Objetivo principal:**

El objetivo principal que persigue la seguridad ciudadana es que las personas puedan desarrollarse y alcanzar la calidad de vida que deseen en un marco de libertad, sin temores a contingencias o peligros que afecte su vida cotidiana.

**Objetivos secundarios:**

- I. Reducir la inseguridad ciudadana a favor de la población.
  - i. Consolidar la participación para prevenir y afrontar la inseguridad ciudadana.
  - ii. Combatir la delincuencia Común por delitos y faltas contra el patrimonio.
- II. Reducir el crimen organizado en el territorio nacional.

## **2.5. FORMULACIÓN DE LAS HIPOTESIS**

Debido a los diversos problemas de la Seguridad Ciudadana en la que está expuesta la población. Se ha desarrollado una investigación en el cual vamos a determinar los problemas de esta situación, considerando los posibles aspectos que se involucrarían y serían los motivos a los problemas de la seguridad ciudadana. Es posible que la insuficiente presencia policial en las calles, la falta de recursos en sus equipos de comunicación y transporte y la decadente infraestructura de las comisarias sean factores de los problemas de la Seguridad Ciudadana.

## **3. FUNDAMENTO TEORICO:**

### **3.1. ESTADO DEL ARTE**

#### **ESTADO DEL ARTE (ESTUDIOS ANTERIORES DE ACUERDO A LA SEGURIDAD)**

La seguridad ciudadana en los últimos 20 años ha tenido avances gracias a la evolución de la tecnología. De esta forma, se ha pasado de contar con grabaciones en formato analógico, que no generaban valor digital, a tener soluciones digitales que se pueden explotar desde la analítica.

Pese a los avances, en la actualidad, la información se sigue almacenando en silos y no es explorada en su totalidad ya que se depende del hombre para procesar los datos, lo que genera que se reaccione de forma reactiva ante las situaciones que se presentan.

Hoy se cuenta con una gran variedad de hardware y software para procesar los datos en el menor tiempo posible. Las soluciones analíticas permiten encontrar los patrones que se encuentran ocultos en la información, lo que se traduce en beneficios y avances para que las ciudades superen los retos cotidianos, entre ellos el crimen.

Emplear datos de múltiples fuentes y diversos formatos, permite tener mejores insumos para la prevención del crimen. De esta manera, las ciudades pueden aprovechar datos como: localizaciones geográficas, el clima, la hora y eventos históricos para tomar decisiones en tiempo real, coordinar rápidamente las agencias y resolver los casos basándose en información confiable y no solamente en el conocimiento de los operadores de los centros de control.

Por Ejemplo, la policía de Chicago usa modelos predictivos para combatir el crimen, así pueden prever en qué zonas de la ciudad ocurrirán actos como tiroteos; de esta manera, los agentes pueden actuar incluso antes de que sucedan. El uso de los datos para sus estrategias, los ha llevado a reducir hasta en un 39% los tiroteos en la ciudad.

En definitiva, unos de los pasos indispensables para que las ciudades sean más inteligentes y avancen tecnológicamente es apalancarse en el análisis y estudio de la información (BI) que recolectan diariamente.

### **3.2.ENFOQUES**

#### **3.2.1. APLICACION DE MINERIA DE DATOS PARA LA EXPLORACION Y DETECCION DE PATRONES DELICTIVOS EN ARGENTINA.**

La Minería de Datos es un proceso iterativo de búsqueda de información no trivial en grandes volúmenes de datos [Kantardzic, 2002] que busca generar información similar a la que podría generar un experto humano: patrones, asociaciones, cambios, anomalías y estructuras significativas [Britos et al, 2005].

Actualmente la Dirección Nacional de Política Criminal (DNPC) del Ministerio de Justicia y Derechos Humanos de la Nación (MJDHN) de Argentina analiza la información proveniente de sus sistemas (SNIC, SAT y SNEEP) mediante un análisis estadístico básico, sin hacer un aprovechamiento exhaustivo de la información.

El presente trabajo aplica técnicas y herramientas de minería de datos sobre la información relevada por la DNPC mediante un software gratuito que permita a los analistas complementar el análisis actual con conclusiones de mayor valor agregado. Como resultado se obtuvo que la minería de datos ha demostrado no sólo que es factible aplicar minería de datos a la información criminal en Argentina, sino también su alto valor agregado para el análisis y la generación de nuevo conocimiento.

#### **3.2.2. Un Novedoso Sistema de Bajo Tiempo de Procesamiento para la Detección de Actividades Delictivas Aplicado a Centros de Comando y Control de Seguridad Ciudadana**

Este trabajo muestra un Novedoso Sistema de Bajo Tiempo de Procesamiento enfocado a la detección de actividades delictivas basado en análisis de video en tiempo real aplicado a Centros de

Comando y Control de Seguridad Ciudadana. Este sistema se aplicó a la detección y clasificación de hechos delictivos en un subsistema de videovigilancia en tiempo real en el Centro de Comando y Control de Seguridad Ciudadana de la Policía Nacional de Colombia. Fue desarrollado utilizando una aplicación novedosa de Deep Learning, específicamente una Red Convolutiva Basada en Regiones (R-CNN) más rápida para la detección de actividades delictivas tratadas como "objetos" para ser detectados en video en tiempo real. Con el fin de maximizar la eficiencia del sistema y reducir el tiempo de procesamiento de cada cuadro de video, se utilizó el modelo preentrenado CNN (Red Neural Convolutiva) AlexNet y el entrenamiento fino se llevó a cabo con un conjunto de datos construido para este proyecto, formado por objetos de uso común en actividades delictivas como armas de fuego cortas y armas blancas. Además, el sistema fue entrenado para la detección de robos en las calles. El sistema puede generar alarmas al detectar hurto en la calle, armas de fuego cortas y armas blancas, mejorando el conocimiento situacional y facilitando la toma de decisiones estratégicas en el Centro de Comando y Control de Seguridad Ciudadana de la Policía Nacional de Colombia.

### **3.2.3. Modelo de redes neuronales para predecir la tendencia de víctimas de secuestro en Colombia**

El modelo utiliza un conjunto de datos tomado del portal de datos abiertos del gobierno y publicado por la fiscalía nacional de Colombia, el cual, cuenta con un total de 39.863 registros de secuestros realizados en Colombia entre los años 1982 y 2020, donde se especifica el total de víctimas en cada caso.

Del conjunto de datos original se descartaron datos nulos, incompletos, y que no tuvieran referencia a Colombia, ya que hay registros de secuestros en otros países. Para entrenar los modelos se agrupó el total de víctimas por año en que ocurrió el secuestro y se procedió a generar un dataset de 6 columnas, donde las primeras 5 columnas corresponden a una ventana de tiempo de 5 años y la última columna corresponde al año siguiente.

La Red Neuronal se implementó usando Keras de Tensorflow y MLPRegressor de sklearn, así mismo, se configuró con 6 capas ocultas y 505 nodos y fue entrenada con una iteración máxima de 150 épocas y el algoritmo de optimización Adam. Los datos de entrada de la red corresponden en un porcentaje del 85% a datos de entrenamiento, y un porcentaje del 15% a datos de prueba para evaluar el modelo; la división de datos de entrenamiento y prueba se realizó sin modificar el orden temporal de los datos.

Considerando que los pesos en la Red Neuronal se establecen de

forma aleatoria al iniciar el entrenamiento, las implementaciones se entrenaron y almacenaron en varias oportunidades hasta hallar los resultados con menor error cuadrático medio. Este documento presentó la evaluación de dos modelos: Máquinas de Soporte Vectorial para regresión (SVR) y Redes Neuronales donde se pudo estimar la cantidad de víctimas de secuestros en los próximos diez años en todo el territorio colombiano, y se encontró que la implementación de la Red Neuronal tuvo mejor comportamiento que la Máquina de Soporte Vectorial.

### **3.3. ANTECEDENTES**

En Perú, como en toda Latinoamérica, el alto grado de inseguridad generado por la violencia y la delincuencia obstaculizan el crecimiento económico y la reducción de la pobreza. Sin embargo, la falta de datos precisos impide formular en forma adecuada el problema. En el caso de Perú este asunto es, al parecer, particularmente grave. Aunque no existen estudios concluyentes al respecto, se estima que sólo el 25% de los actos delictivos son denunciados. Por otro lado, los estudios realizados hasta el momento para el caso peruano han incidido en una u otra manifestación violenta o criminal para sugerir así los niveles de inseguridad imperantes. Asimismo, gran parte de las inquietudes se han focalizado en el ámbito de Lima Metropolitana

El INEI en una investigación para el fortalecimiento estratégico de los procesos de auditoría interna y externa para el mejoramiento continuo de la calidad de datos obtuvo que la tasa de victimización en el 2015 en el Perú fue de 30,8% y que los crímenes violentos en el país vienen creciendo en el tiempo. En los últimos cinco años, del 2011 al 2015, la tasa de homicidios creció 32,0%, los delitos con arma de fuego crecieron 45,0%, aumentaron las denuncias por violencia sexual y por lesiones, aparecieron nuevas formas delictivas como la extorsión, los asaltos, el sicariato; se observa también que los delitos tienen una alta concentración geográfica: 80,0% de estos ocurren en cerca de 100 de los 1.874 distritos del país (2015), pero solo el 15,0% de las víctimas denuncian estos hechos

Estos enfoques si bien son valiosas contribuciones no dejan de ser parciales. Este documento busca suplir estos vacíos. En ese sentido ha reunido y ordenado la información oficial disponible tratando de darle coherencia estadística mediante el levantamiento de índices. También ha interrelacionado las diversas manifestaciones de violencia que se producen en todo el país, según tipo y frecuencia.

En conclusión, se recomienda impulsar un proceso de implementación de una plataforma de interoperabilidad entre las instituciones para compartir datos e información sobre la criminalidad y seguridad ciudadana, de tal manera que el flujo de información sea cada vez más rápido y permita realizar consultas en línea.

### 3.4. ARGUMENTOS

La seguridad pública es actualmente el tema de mayor preocupación en Perú. La solución al problema de la delincuencia requiere la acción integrada por parte del Estado, pero el rol particular de las fuerzas policiales resulta crucial. Según estudio realizado por Miguel Ángel Carpio (2014) el efecto de la presencia policial sobre diferentes tipos de delitos en Perú: un incremento de 1.00% en el número de policías por 100,000 habitantes genera una disminución en la probabilidad de experimentar un delito de 0.52%. Asimismo, el estudio detecta que este incremento genera impactos diferentes según la naturaleza del delito: -0.65% en el caso de robo de dinero, cartera o celular, -3.17% en el caso de estafa y -2.23% en el caso de vehículos o autopartes. Al mismo tiempo, no tiene ningún impacto sobre robo de vivienda ni sobre amenazas e intimidaciones, lo cual probablemente se explique porque se trata de delitos que no ocurren en la vía pública.

La Policía Nacional del Perú (PNP), entidad estatal encargada de la seguridad interna, no cuenta con un sistema único de información del delito tecnologizado y completo, lo cual afecta negativamente el empleo de una buena georreferenciación. La información recopilada sirve como insumo para el diagnóstico que realice alguna política pública. Asimismo, la novedad del tema da un acercamiento a la importancia de los medios tecnológicos como instrumentos de la gestión de conocimiento para el diseño de las políticas públicas (Mora, 2015)

La lucha contra la delincuencia común y organizada requiere de herramientas que permitan a los departamentos de investigación criminal - DEPINCRIS cumplir a cabalidad sus funciones. En el caso de las unidades móviles asignadas, como son las camionetas, autos y motos, sorprende el número de vehículos inoperativos. Esto reduce la capacidad de la policía para atender a la población y a las víctimas, con lo adecuado para acudir a las escenas del delito. Urge atención en dos niveles: reparación de vehículos y repotenciación del parque automotor. Defensoría del Pueblo (2019)

## 4. CONCEPTOS

### 4.1. Data Mart:

Un Data Mart es un almacén de datos orientado a un área específica, como por ejemplo, Ventas, Recursos Humanos u otros sectores en una organización. Por ello, también se le conoce como una base de información departamental. Este almacén permite que una empresa pueda acceder a datos claves de un área de forma sencilla, además de realizar diversas funciones, tales como:

- La organización de información para su posterior análisis.
- La elaboración de indicadores clave de rendimiento (KPI).
- La creación de informes para un aprendizaje automático.
- La evaluación de datos sobre el cumplimiento de objetivos de un sector.



Su implementación es fundamental en un negocio, pues de esta manera la empresa puede integrar y combinar la información en un Data Mart para descubrir oportunidades y amenazas indetectables en las bases de datos operacionales. En ellas, solo se almacena la información de las transacciones diarias de un negocio; no obstante, al integrarlas con las bases de la organización, es posible conseguir un análisis de valor. Una compañía que implementa un Data Mart obtiene diversos beneficios, tales como:

- La generación de datos dirigidos a perfiles de negocio. Gracias al almacenamiento sencillo de información, la entrega de datos es óptima y rápida, además de útil para la toma de decisiones importantes en la organización.
- Los colaboradores encargados de gestionar la información pueden utilizarla para potenciar las herramientas analíticas con las que laboran.
- Es un mecanismo de respuesta rápido para consultar información antes de realizar un cambio en un proceso empresarial.
- A ligera la carga de trabajo. Los ejecutivos pueden realizar consultas rápidas en las fuentes de datos, en cualquier momento y lugar.

#### **4.2. MySQL:**

MySQL es un sistema de base de datos relacional o RDBMS (*Relational Database Management System*) que emplea un modelo cliente-servidor.

Es multihilo y multiusuario, y además cuenta con millones de instalaciones en todo el mundo. MySQL es el segundo motor de bases de datos relacional más utilizado en todo el mundo, por detrás de Oracle.

Es desarrollado por MySQL AB desde 1995 (actualmente subsidiaria de Oracle Corporation) como software libre. El objetivo era cumplir el estándar SQL(*Structured Query Language*) sin sacrificar velocidad, fiabilidad o usabilidad.

MySQL es el tipo de Bases de Datos más popular. Combinada con PHP, es una herramienta que ofrece a desarrolladores y diseñadores web numerosas posibilidades.

#### **4.3. Power BI:**

Power BI es una herramienta que se utiliza principalmente para crear cuadros de mando que faciliten la toma de decisiones. La información se puede actualizar de manera automatizada o manual y permite la compartición de los informes mediante la propia herramienta.

Esto permite pasar de un sistema con varias herramientas de gestión (un ERP, un CRM, varias tablas de control en Excel) y un controller que cruza información entre ellas habitualmente en Excel para entregar a Dirección informes en Excel o en papel a una situación en la que el controller prepara los informes Power BI que necesita el gerente recopilando la información más importante en un cuadro de mandos que se actualiza automáticamente.

El cuadro de mando equivale en su función al salpicadero de un coche, mediante el cual el gerente puede recibir información adecuada y suficiente para conducir su empresa de una manera bastante más autónoma y libre de los errores humanos que pueden producirse en el cruce de información y actualización cuando no se utiliza un sistema de Business Intelligence como es Power BI.

¿Cuáles son las capacidades principales de Power BI?

- Permitir importar datos y moldearlos a voluntad.
- Facilita cruzar datos, hacer cálculos y gráficas sofisticadas de manera relativamente sencilla.
- Permite que los cálculos se reutilicen tantas veces como sea necesario con la información actual recogida en las bases de datos de origen, por lo que se puede contar con información actualizada y libre de errores humanos.

## 5. METODOLOGIA

### 5.1. DESARROLLO DEL DATA MART

- **Objetivo estratégico.**

El objetivo estratégico del presente trabajo reducir la inseguridad ciudadana a favor de la población, ya que actualmente el índice de delincuencia va incrementando año tras año y esto causa cada vez mas inseguridad en los ciudadanos del país, ya sea porque no se proporciona la cantidad necesaria de policías en las calles o porque no se brinda los implementos necesarios para que los policías puedan realizar su trabajo.

- **Fuente de datos.**

Los datos para este trabajo fueron extraídos de la pagina web del INEI, de la encuesta CENSO NACIONAL DE COMISARIAS del año 2017 por departamento anual, de las 9 encuestas que se encuentran trabajamos las siguientes encuestas con códigos módulos: 1226,1227,1231,1232.

De los datos que se encuentran en el código modulo 1226 nos quedamos con las siguientes columnas:

- IDN\_N, UBIGEO, CCDI, NOMBREDI, INF109, INF109B3, INF110\_TOT, INF112\_MOT, INF113\_PIE, INF114\_INV\_DF, INF115\_INV\_AT, INF116\_INV\_VF, GPSLONGITUD\_INF, GPSLATITUD\_INF.

De los datos que se encuentran en el código modulo 1227 nos quedamos con las siguientes columnas:

- ID\_N, UBIGEO, CCDI, NOMBREDI, INF207, INF212A\_1, INF263.

De los datos que se encuentran en el código modulo 1231 nos quedamos con las siguientes columnas:

- ID\_N, UBIGEO, CCDI, NOMBREDI, INF301, INF301\_TOTAL, INF334, INF337, INF342\_1, INF342\_2, INF342\_3, INF342,4, INF342\_5.

De los datos que se encuentran en el código modulo 1232 nos quedamos con las siguientes columnas:

- ID\_N, UBIGEO, CCDI, NOMBREDI, INF638, INF650\_1\_2016, INF650\_1\_2017, INF650\_2\_2016, INF650\_2\_2017, INF650\_3\_2016, INF650\_3\_2017, INF650\_4\_2016, INF650\_4\_2017, INF650\_5\_2016, INF650\_5\_2017, INF650\_6\_2016, INF650\_6\_2017, INF650\_7\_2016, INF650\_7\_2017, INF650\_8\_2016, INF650\_8\_2017, INF650\_9\_2016, INF650\_9\_2017, INF650\_10\_2016, INF650\_11\_2016, INF650\_12\_2016, INF650\_TOT\_2016, INF650\_TOT\_2017, INF652\_1\_2016, INF652\_1\_2017, INF652\_2\_2016, INF652\_2\_2017, INF652\_3\_2016, INF652\_3\_2017, INF652\_4\_2016, INF652\_4\_2017, INF652\_5\_2016, INF652\_5\_2017, INF652\_6\_2016, INF652\_6\_2017, INF652\_7\_2016, INF652\_7\_2017, INF652\_8\_2016, INF652\_8\_2017, INF652\_9\_2016, INF652\_9\_2017, INF652\_10\_2016, INF652\_11\_2016, INF652\_12\_2016, INF652\_TOT\_2016, INF652\_TOT\_2017, INF653\_1\_2016, INF653\_1\_2017, INF653\_2\_2016, INF653\_2\_2017, INF653\_3\_2016, INF653\_3\_2017, INF653\_4\_2016, INF653\_4\_2017, INF653\_5\_2016, INF653\_5\_2017, INF653\_6\_2016, INF653\_6\_2017, INF653\_7\_2016, INF653\_7\_2017, INF653\_8\_2016, INF653\_8\_2017, INF653\_9\_2016, INF653\_9\_2017, INF653\_10\_2016, INF653\_11\_2016, INF653\_12\_2016, INF653\_TOT\_2016, INF653\_TOT\_2017.

Luego, creamos 6 entidades:

✓ **CIUDAD.**

Con los siguientes atributos:

- Id\_ciudad
- Ubigeo
- Ccdi
- Nombredi
- gpslongitud\_inf
- gpslatitud\_inf

✓ **POLICIAS.**

Con los siguientes atributos:

- id\_policia
- inf110\_tot
- inf112\_mot
- inf113\_pie
- inf114\_inv\_df
- inf115\_inv\_at
- inf116\_inv\_vf

✓ **COMISARIA.**

Con los siguientes atributos:

- id\_nnm
- inf109
- inf109b3
- inf207
- inf212a\_1
- inf263

✓ **EQUIPOS.**

Con los siguientes atributos:

- id\_equipo
- inf301
- inf301\_total
- inf334
- inf337
- inf638

✓ **VEHICULOS.**

Con los siguientes atributos:

- id\_vehiculo
- inf342\_1
- inf342\_2
- inf342\_3
- inf342\_4
- inf342\_5

✓ **DETENCIONES.**

Con los siguientes atributos:

- |                   |                   |
|-------------------|-------------------|
| - id_detenciones  |                   |
| - inf650_1_2016   | - inf650_1_2017   |
| - inf650_2_2016   | - inf650_2_2017   |
| - inf650_3_2016   | - inf650_3_2017   |
| - inf650_4_2016   | - inf650_4_2017   |
| - inf650_5_2016   | - inf650_5_2017   |
| - inf650_6_2016   | - inf650_6_2017   |
| - inf650_7_2016   | - inf650_7_2017   |
| - inf650_8_2016   | - inf650_8_2017   |
| - inf650_9_2016   | - inf650_9_2017   |
| - inf650_10_2016  | - inf650_11_2016  |
| - inf650_12_2016  |                   |
| - inf650_tot_2016 | - inf650_tot_2017 |
| - inf652_1_2016   | - inf652_1_2017   |

- inf652\_2\_2016
- inf652\_3\_2016
- inf652\_4\_2016
- inf652\_5\_2016
- inf652\_6\_2016
- inf652\_7\_2016
- inf652\_8\_2016
- inf652\_9\_2016
- inf652\_10\_2016
- inf652\_12\_2016
- inf652\_tot\_2016
- inf653\_1\_2016
- inf653\_2\_2016
- inf653\_3\_2016
- inf653\_4\_2016
- inf653\_5\_2016
- inf653\_6\_2016
- inf653\_7\_2016
- inf653\_8\_2016
- inf653\_9\_2016
- inf653\_10\_2016
- inf653\_12\_2016
- inf653\_tot\_2016
- inf652\_2\_2017
- inf652\_3\_2017
- inf652\_4\_2017
- inf652\_5\_2017
- inf652\_6\_2017
- inf652\_7\_2017
- inf652\_8\_2017
- inf652\_9\_2017
- inf652\_11\_2016
- inf652\_tot\_2017
- inf653\_1\_2017
- inf653\_2\_2017
- inf653\_3\_2017
- inf653\_4\_2017
- inf653\_5\_2017
- inf653\_6\_2017
- inf653\_7\_2017
- inf653\_8\_2017
- inf653\_9\_2017
- inf653\_11\_2016
- inf653\_tot\_2017

- **Modelo Entidad-Relación.**

Se mostrará el modelo creado de las 6 entidades.

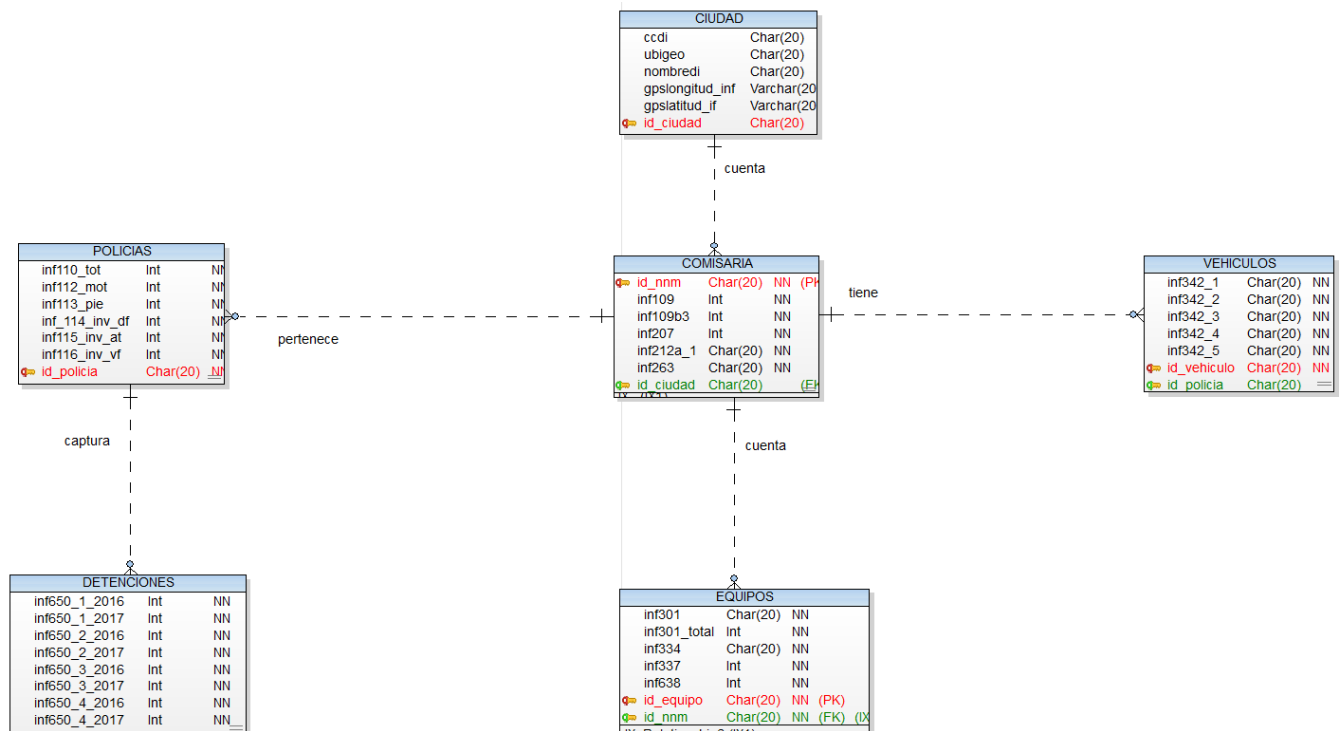


Figura 1: modelo Entidad-Relación obtenida del Toad Data Modeler.

- **Descripción del Proceso de ETL.**

El proceso de ETL es importante ya que con este proceso se va a obtener el datamart para realizar algunas técnicas que nos van a permitir tomar decisiones.

El proceso ETL (Extract, Transform, Load) cuenta con tres etapas:

**Fase de extracción:**

Para esta etapa de extracción las entidades con sus respectivos atributos se encuentran en un archivo de extensión xlsx.

A continuación, mostraremos los datos en las respectivas variables.

✓ **CIUDAD**

Se muestran los primeros registros.

	A	B	C	D	E	F
1	CCDI	UBIGEO	NOMBREDI	GPSLONGITUD_INF	GPSLATITUD_IF	ID_CIUADAD
2	01	100601	RUPA-RUPA	-76.00260836	-9.29806953	0001
3	06	100606	MARIANO DAMASO BERAUN	-75.95306846	-9.49461342	0002
4	04	100604	JOSE CRESPO Y CASTILLO	-76.11188986	-8.93140901	0003
5	01	221001	TOCACHE	-76.51227108	-8.1884396	0004
6	02	221002	NUEVO PROGRESO	-76.3241724	-8.45341229	0005
7	05	221005	UCHIZA	-76.46217742	-8.45046975	0006
8	01	250301	PADRE ABAD	-75.50689252	-9.03794253	0007
9	05	250305	ALEXANDER VON HUMBOLDT	-75.051607	-8.82538	0008
10	02	250302	IRAZOLA	-75.21614922	-8.82951452	0009
11	01	200101	PIURA	-80.63408522	-5.19025882	0010
12	15	200115	VEINTISEIS DE OCTUBRE	-80.668668	-5.185295	0011
13	04	200104	CASTILLA	-80.59871083	-5.18713728	0012
14	05	200105	CATACAOS	-80.6637766	-5.24248333	0017
15	07	200107	CURA MORI	-80.66488714	-5.32463504	0019
16	09	200109	LA ARENA	-80.71298264	-5.34800031	0020
17	10	200110	LA UNION	-80.74340563	-5.40232231	0021
18	11	200111	LAS LOMAS	-80.24390865	-4.65659873	0022
19	14	200114	TAMBO GRANDE	-80.34184538	-4.93225789	0024
20	01	200201	AYABACA	-79.71481699	-4.63912067	0026

Figura 2: primeros registros de la entidad CIUDAD.

✓ **POLICIAS**

Se muestra las primeras filas de la tabla entidad.

*Figura 3: primeros registros de la entidad POLICIAS.*

	A	B	C	D	E	F	G	H
1	INF110_TOT	INF112_MOT	INF113_PIE	INF114_INV_DF	INF115_INV_	INF116_INV_VF	ID_POLICIA	ID_NNM
2	58	13	0	9	6	0	000001	0001
3	18	4	0	3	3	1	000002	0002
4	27	9	0	2	2	1	000003	0003
5	55	20	7	2	3	2	000004	0004
6	28	12	0	4	2	2	000005	0005
7	35	10	2	4	2	2	000006	0006
8	29	8	2	8	2	2	000007	0007
9	22	4	0	4	4	4	000008	0008
10	21	4	2	4	2	2	000009	0009
11	162	47	71	6	4	2	000010	0010
12	110	31	39	11	2	3	000011	0011
13	55	24	0	6	2	3	000012	0012
14	68	27	11	4	1	4	000013	0013
15	79	42	0	7	2	3	000014	0014
16	13	0	4	1	1	1	000015	0015
17	61	18	15	6	2	2	000016	0016
18	23	8	4	1	1	1	000017	0017
19	119	45	40	5	2	2	000018	0018
20	16	0	5	1	1	2	000019	0019
21	24	6	3	2	1	2	000020	0020
22	46	19	9	2	1	2	000021	0021
23	34	9	9	2	2	2	000022	0022
24	10	1	2	1	1	1	000023	0023
25	47	19	10	2	3	2	000024	0024
26	17	4	0	2	2	2	000025	0025
27	31	4	11	2	2	1	000026	0026

✓ **COMISARIA**

Se muestra los primeros registros de la entidad comisaria.

	A	B	C	D	E	F	G
1	ID_NNM	INF109	INF109B3	INF207	INF212A_1	INF263	id_ciudad
2	0001	5	2	22	2	1	0001
3	0002	2	1	18	2	1	0002
4	0003	4	2	27	2	1	0003
5	0004	2	0	20	2	1	0004
6	0005	2	1	19	2	1	0005
7	0006	1	1	20	2	1	0006
8	0007	4	1	17	3	1	0007
9	0008	3	0	24	2	1	0008
10	0009	3	1	7	2	1	0009
11	0010	6	1	47	2	1	0010
12	0011	5	0	48	2	1	0011
13	0012	5	1	20	2	1	0012
14	0013	5	1	47	3	1	0012
15	0014	5	1	25	3	1	0012
16	0015	3	0	30	2	1	0012
17	0016	5	1	12	2	1	0010
18	0017	3	0	27	3	1	0013
19	0018	5	1	2	2	1	0013
20	0019	2	1	56	2	1	0014
21	0020	4	1	11	2	1	0015
22	0021	5	3	32	2	1	0016
23	0022	3	1	52	3	1	0017
24	0023	2	0	8	2	1	0017
25	0024	5	1	15	2	1	0018
26	0025	3	0	23	2	1	0018
27	0026	5	1	40	3	1	0019

*Figura 4: primeros registros de la entidad COMISARIA.*



✓ **EQUIPOS**

Se muestra los primeros registros de la entidad equipo.

	A	B	C	D	E	F	G
1	INF301	INF301_TOTAL	INF334	INF337	INF638	ID_EQUIPO	ID_NNM
2	1	3	1	1	0	0001	0001
3	2	0	2	1	0	0002	0002
4	2	0	2	1	0	0003	0003
5	1	1	2	0	0	0004	0004
6	2	0	1	1	NA	0005	0005
7	2	0	2	0	6	0006	0006
8	1	2	2	2	7	0007	0007
9	1	1	1	4	3	0008	0008
10	1	1	2	3	15	0009	0009
11	1	1	1	2	0	0010	0010
12	1	4	1	5	0	0011	0011
13	2	0	2	1	0	0012	0012
14	1	1	1	9	0	0013	0013
15	1	2	1	3	0	0014	0014
16	1	1	1	1	0	0015	0015
17	1	2	1	1	NA	0016	0016
18	1	1	1	12	0	0017	0017
19	1	1	1	3	0	0018	0018
20	2	0	1	2	NA	0019	0019
21	1	1	1	1	NA	0020	0020
22	1	3	1	2	NA	0021	0021
23	1	1	2	1	0	0022	0022
24	1	1	2	2	NA	0023	0023
25	1	1	2	2	NA	0024	0024
26	1	1	2	2	0	0025	0025
27	2	0	2	3	NA	0026	0026

Figura 5: primeros registros de la entidad EQUIPOS.

✓ **VEHICULOS**

Se muestran los primeros registros de la entidad vehículos.

	B	C	D	E	F	G	H
1	INF342_2	INF342_3	INF342_4	INF342_5	ID_VEHICULO	ID_POLICIA	ID_NNM
2	2	2	1	1	0001	000001	0001
3	2	1	2	2	0002	000002	0002
4	2	1	1	1	0003	000003	0003
5	2	1	2	1	0004	000004	0004
6	2	1	2	1	0005	000005	0005
7	2	1	2	1	0006	000006	0006
8	2	1	2	1	0007	000007	0007
9	2	2	1	2	0008	000008	0008
10	2	2	2	1	0009	000009	0009
11	2	2	1	1	0010	000010	0010
12	2	1	2	1	0011	000011	0011
13	2	1	1	1	0012	000012	0012
14	2	2	1	1	0013	000013	0013
15	2	2	1	2	0014	000014	0014
16	2	2	2	2	0015	000015	0015
17	2	1	1	1	0016	000016	0016
18	2	1	2	1	0017	000017	0017
19	1	2	1	1	0018	000018	0018
20	2	2	1	2	0019	000019	0019
21	2	2	1	1	0020	000020	0020
22	2	2	1	1	0021	000021	0021
23	2	1	2	1	0022	000022	0022
24	2	2	2	1	0023	000023	0023
25	2	1	1	1	0024	000024	0024
26	2	2	2	1	0025	000025	0025
27	2	1	2	2	0026	000026	0026

Figura 6: primeros registros de la entidad VEHICULOS.

## ✓ DETENCIONES

Se muestran los primeros registros de la entidad detenciones con todos sus atributos.

	A	B	C	D	E	F	G	H	I	J
1	INF650_1_2016	INF650_1_2017	INF650_2_2016	INF650_2_2017	INF650_3_2016	INF650_3_2017	INF650_4_2016	INF650_4_2017	INF650_5_2016	INF650_5_2017
2	0	1	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	2	0	0
4	0	0	0	0	0	0	0	0	0	0

	J	K	L	M	N	O	P	Q	R	S
1	INF650_5_2017	INF650_6_2016	INF650_6_2017	INF650_7_2016	INF650_7_2017	INF650_8_2016	INF650_8_2017	INF650_9_2016	INF650_9_2017	INF650_10_2016
2	0	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	2	0	1	0
4	0	0	0	0	0	0	0	0	0	0

	T	U	V	W	X	Y	Z	AA	AB	AC
1	INF650_11_2016	INF650_12_2016	INF650_TOT_2016	INF650_TOT_2017	INF652_1_2016	INF652_1_2017	INF652_2_2016	INF652_2_2017	INF652_3_2016	INF652_3_2017
2	0	0	0	1	0	0	0	0	0	0
3	0	0	0	6	9	5	14	3	7	6
4	0	0	0	0	0	0	0	0	0	0

	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
1	INF652_4_2016	INF652_4_2017	INF652_5_2016	INF652_5_2017	INF652_6_2016	INF652_6_2017	INF652_7_2016	INF652_7_2017	INF652_8_2016	INF652_8_2017
2	0	26	0	34	0	10	0	19	0	17
3	14	8	13	4	9	6	12	7	4	10
4	0	0	0	0	0	0	0	0	0	0

	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
1	INF652_9_2016	INF652_9_2017	INF652_10_2016	INF652_11_2016	INF652_12_2016	INF652_TOT_2016	INF652_TOT_2017	INF653_1_2016	INF653_1_2017	INF653_2_2016
2	2	19	15	5	10	32	125	0	0	0
3	10	13	7	12	5	116	62	9	5	14
4	0	0	0	0	0	0	0	0	0	0

	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG
1	INF653_2_2017	INF653_3_2016	INF653_3_2017	INF653_4_2016	INF653_4_2017	INF653_5_2016	INF653_5_2017	INF653_6_2016	INF653_6_2017	INF653_7_2016
2	0	0	0	0	26	0	34	0	10	0
3	3	7	6	14	8	13	4	9	6	12
4	0	0	0	0	0	0	0	0	0	0

	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS
1	INF653_7_2017	INF653_8_2016	INF653_8_2017	INF653_9_2016	INF653_9_2017	INF653_10_2016	INF653_11_2016	INF653_12_2016	INF653_TOT_2016	INF653_TOT_2017	ID_DETENCIONES	ID_POLICIA
2	19	0	17	2	19	15	5	10	32	125	000000001	000001
3	7	4	10	10	13	7	12	5	116	62	000000002	000002
4	0	0	0	0	0	0	0	0	0	0	000000003	000003

Figura 7: primeros registros de la entidad DETENCIONES.

## Fase de transformación

Esta fase consiste en procesar los datos de forma que sean coherentes, para la entidad ciudad, se eliminaron las instancias duplicadas respecto al atributo nombredi para saber si habían nombres escritos de forma errónea con la función quitar duplicados del mismo Excel , pero se encontró que si estaban escritos de forma correcta; luego para los NA que presentan algunas instancias en las entidades se decidió cambiarlas por ceros ya que indicaba que no había cantidad para esto utilizamos la función de buscar y reemplazar en el Excel para cada entidad.

## Fase de carga

Cargamos el script generado en el toad para crear las entidades vacías, para posteriormente ir llenando las entidades mediante Import Wizard.

A continuación, se mostrará la carga completa de los datos de cada entidad al mysql.

### ✓ CIUDAD

The screenshot shows the 'Import Wizard' window at Step 6 of 8. The 'Source Table' is 'Hoja1' and the 'Target Table' is 'ciudad'. A table lists the field mappings:

Target Field	Source Field
ccdi	CCDI
ubigeo	UBIGEO
nombredi	NOMBREDI
gpslongitud_inf	GPSLONGITUD_INF
gpslatitud_if	GPSLATITUD_IF
id_ciudad	ID_CIUDAD

At the bottom, there are navigation buttons: '<<', '< Back', 'Next >', '>>', and 'Cancel'. The 'Next >' button is highlighted.

The screenshot shows the 'Import Wizard' window at Step 8 of 8. It displays the following statistics:

Tables:	1
Processed:	996
Errors:	0
Added:	996
Updated:	0
Deleted:	0
Time:	0.630s

Below the statistics, a message box shows the import process details:

```
[Msg] [Imp] Import start
[Msg] [Imp] Import type - Excel2007 file
[Msg] [Imp] Import from - C:\Users\Wayra\Downloads\ciudad (1).xlsx
[Msg] [Imp] Import table [ciudad]
[Msg] [Imp] Processed:996, Added:996, Updated:0, Deleted:0, Errors:0
[Msg] [Imp] Finished - Successfully
```

At the bottom, there are buttons: 'Save', 'Log', 'Start', and 'Close'. The 'Start' button is highlighted.

Figura 8: Carga de los datos (importación de Excel al MySQL) a la entidad CIUDAD.

### ✓ POLICIAS

The screenshot shows the 'Import Wizard' window at Step 6 of 8. The 'Source Table' is 'Hoja1' and the 'Target Table' is 'policias'. A table lists the field mappings:

Target Field	Source Field
inf110_tot	INF110_TOT
inf112_mot	INF112_MOT
inf113_pie	INF113_PIE
inf114_inv_df	INF114_INV_DF
inf115_inv_at	INF115_INV_AT
inf116_inv_vf	INF116_INV_VF
id_policia	ID_POLICIA
id_nnm	ID_NNIM

At the bottom, there are navigation buttons: '<<', '< Back', 'Next >', '>>', and 'Cancel'. The 'Next >' button is highlighted.

The screenshot shows the 'Import Wizard' window at Step 8 of 8. It displays the following statistics:

Tables:	1
Processed:	1495
Errors:	0
Added:	1495
Updated:	0
Deleted:	0
Time:	1.051s

Below the statistics, a message box shows the import process details:

```
[Msg] [Imp] Import start
[Msg] [Imp] Import type - Excel2007 file
[Msg] [Imp] Import from - C:\Users\Wayra\Downloads\POLICIA (2).xlsx
[Msg] [Imp] Import table [policias]
[Msg] [Imp] Processed:1495, Added:1495, Updated:0, Deleted:0, Errors:0
[Msg] [Imp] Finished - Successfully
```

At the bottom, there are buttons: 'Save', 'Log', 'Start', and 'Close'. The 'Start' button is highlighted.

Figura 9: Carga de los datos (importación de Excel al MySQL) a la entidad POLICIAS.

## ✓ COMISARIA

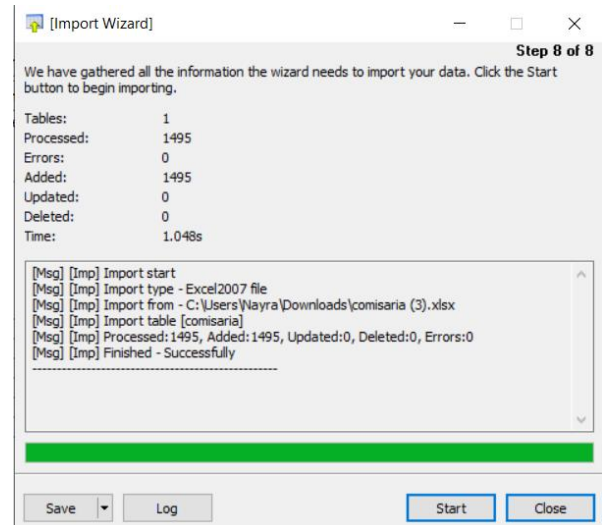
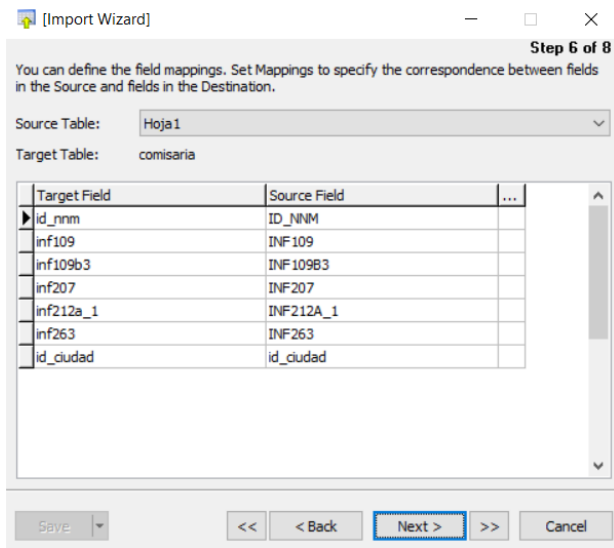


Figura 10: Carga de los datos (importación de Excel al MySQL) a la entidad COMISARIA.

## ✓ EQUIPOS

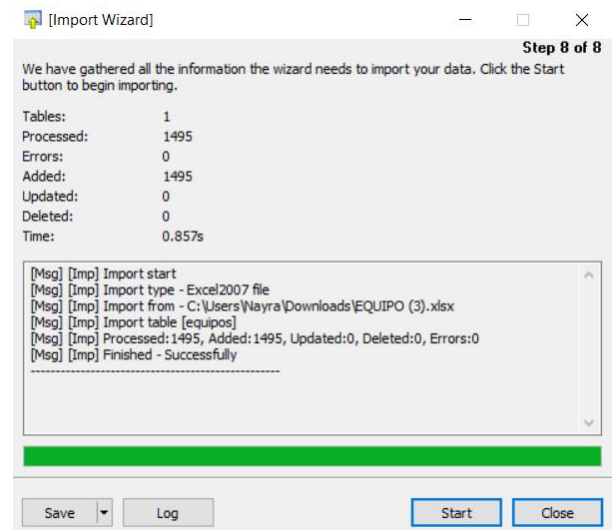
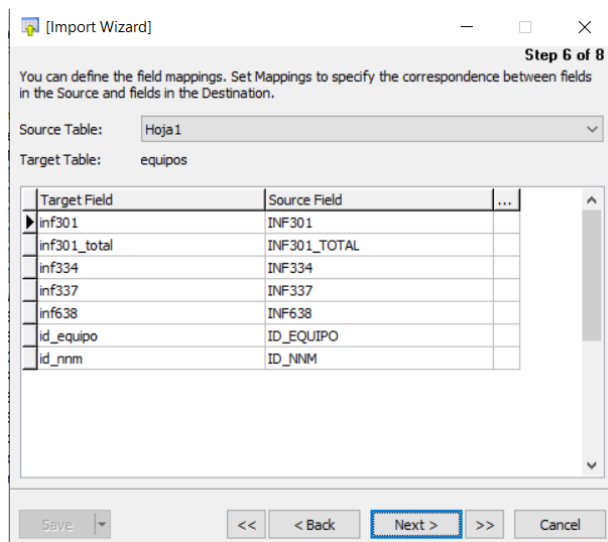


Figura 11: Carga de los datos (importación de Excel al MySQL) a la entidad EQUIPOS.

## ✓ VEHICULOS

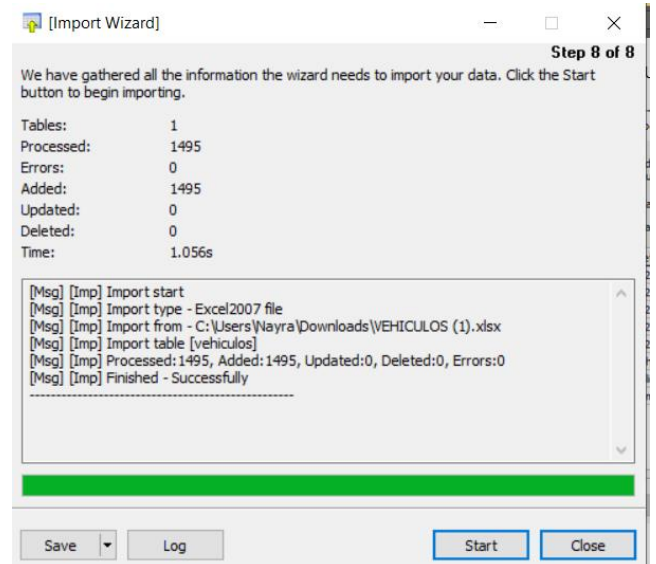
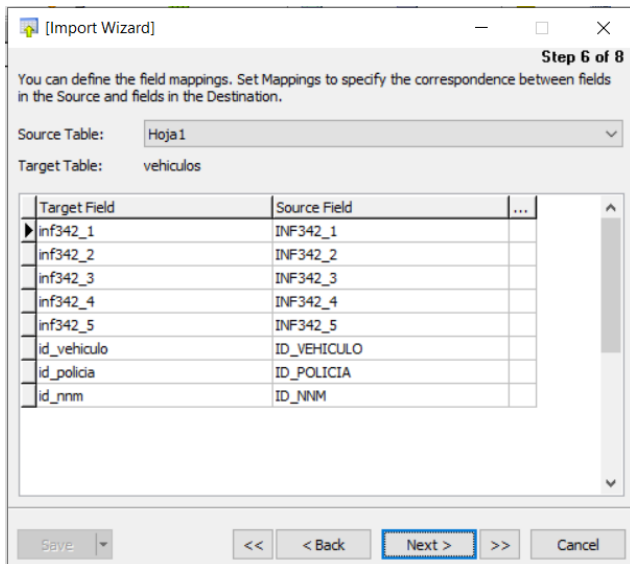


Figura 12: Carga de los datos (importación de Excel al MySQL) a la entidad VEHICULOS.

## ✓ DETENCIONES

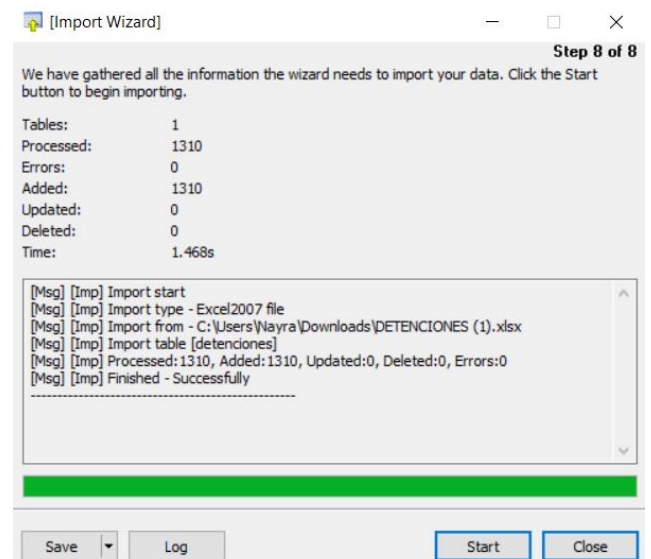
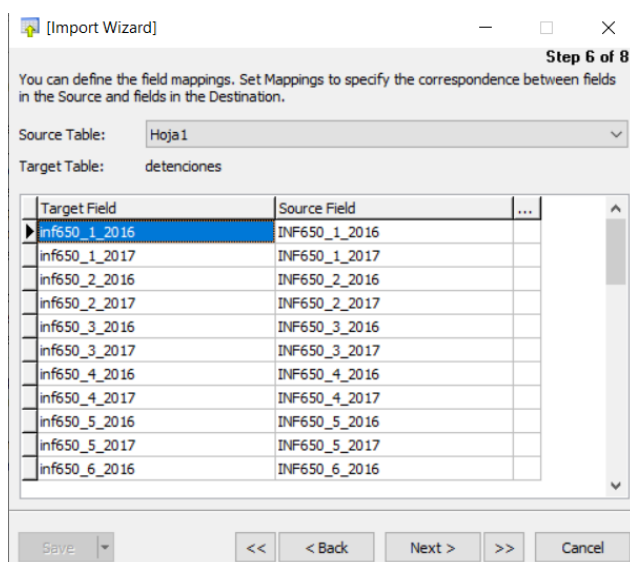
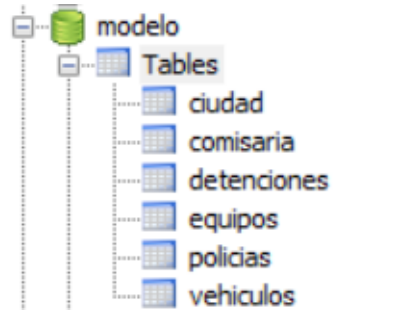


Figura 13: Carga de los datos (importación de Excel al MySQL) a la entidad DETENCIONES.

Luego de haber realizado las fases del proceso ETL, se logró tener todas las entidades dentro de un solo sistema como es el MySQL, en la base de datos modelo.



*Figura 14: Los datos integrados a cada entidad dentro de la base de datos modelo.*

### **Creación del Data Mart**

Se desarrolla el modelo dimensional o modelo estrella, con la finalidad de facilitar la comprensión de la base de datos.

Dicho modelo cuenta con lo siguiente:

Dimensiones:

- CIUDAD
- VEHICULOS
- COMISARIA\_EQUIPO

Tabla de hechos:

- FAC\_TABLE

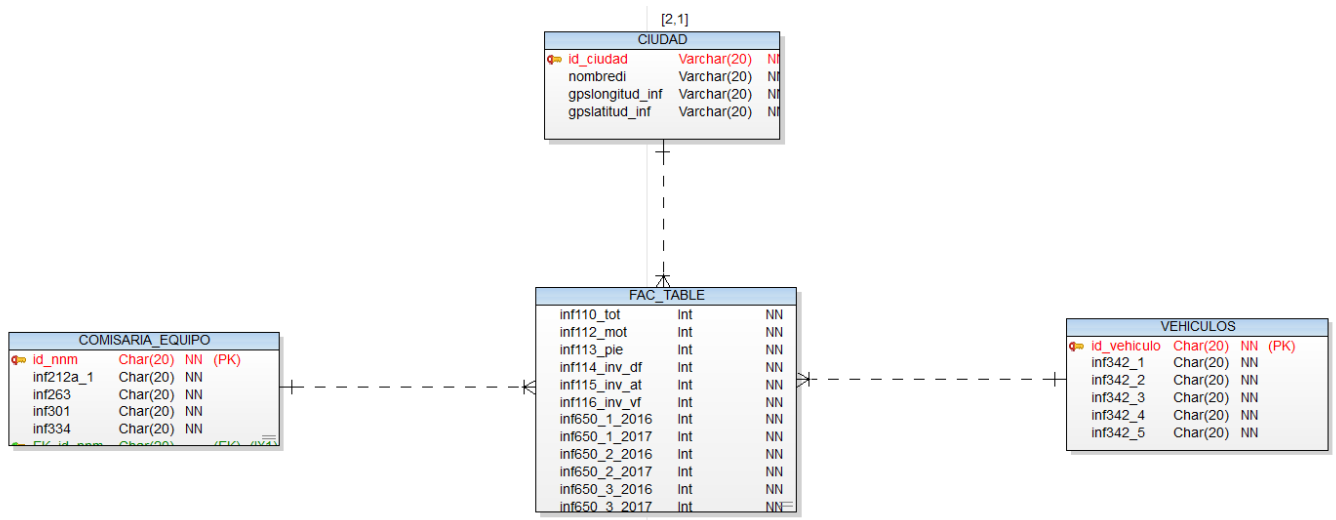


Figura 15: modelo entidad-relación obtenida del Toad data Modeler.

### Creación de las dimensiones en MySQL

Se pasó a crear las dimensiones mencionadas en MySQL usando el EXCEL, pero antes de esto se copio el script generado en el toad para importar los datos del Excel al navicat mediante Import Wizard.

A continuación, se mostrará las dimensiones creadas, las cuales están almacenadas en la base de datos llamada modelo\_dimensional.

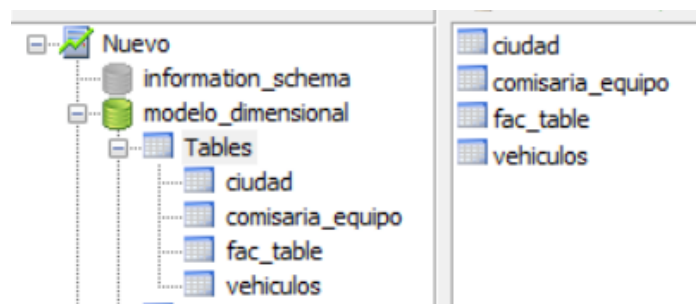


Figura 16: dimensiones creadas dentro de la base de datos modelo\_dimensional.



Ahora se irá importando para cada dimensión desde el Excel:

### ✓ CIUDAD

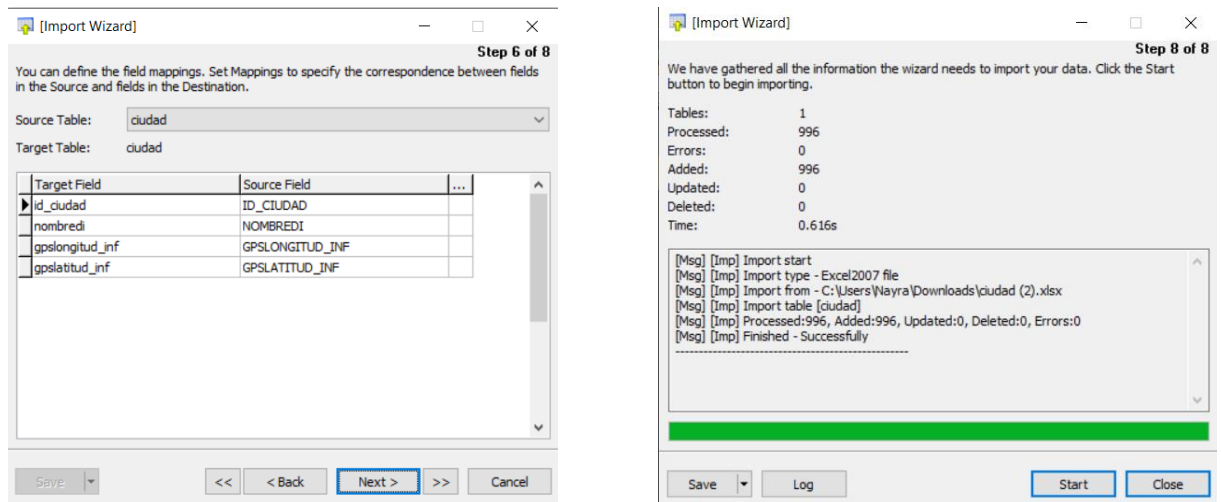


Figura 17: Carga de los datos (importación de Excel al MySQL) a la dimensión CIUDAD.

### ✓ VEHICULOS

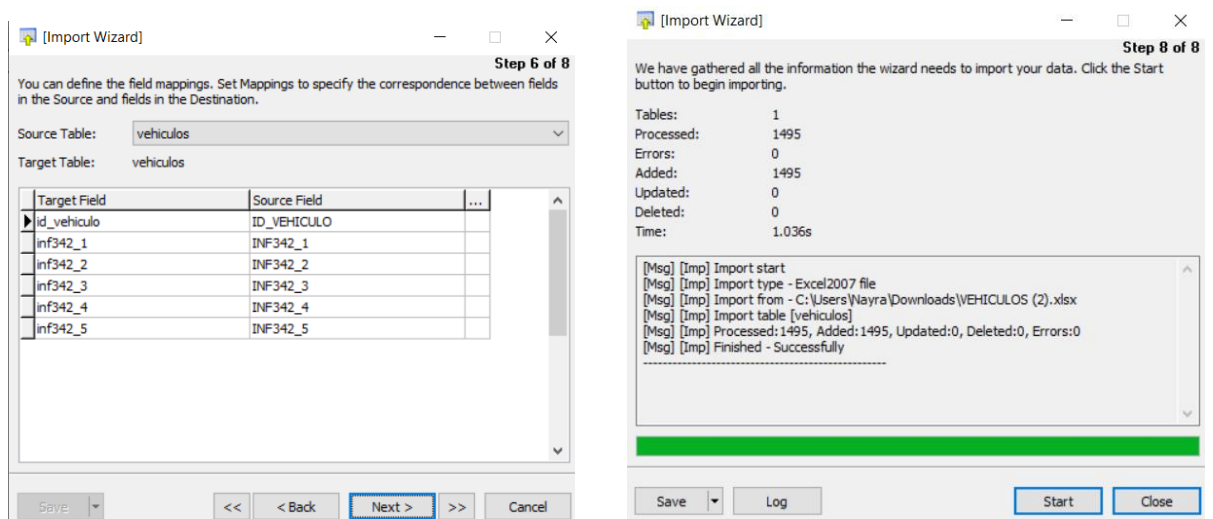


Figura 18: Carga de los datos (importación de Excel al MySQL) a la dimensión VEHICULOS.

## ✓ COMISARIA\_EQUIPO

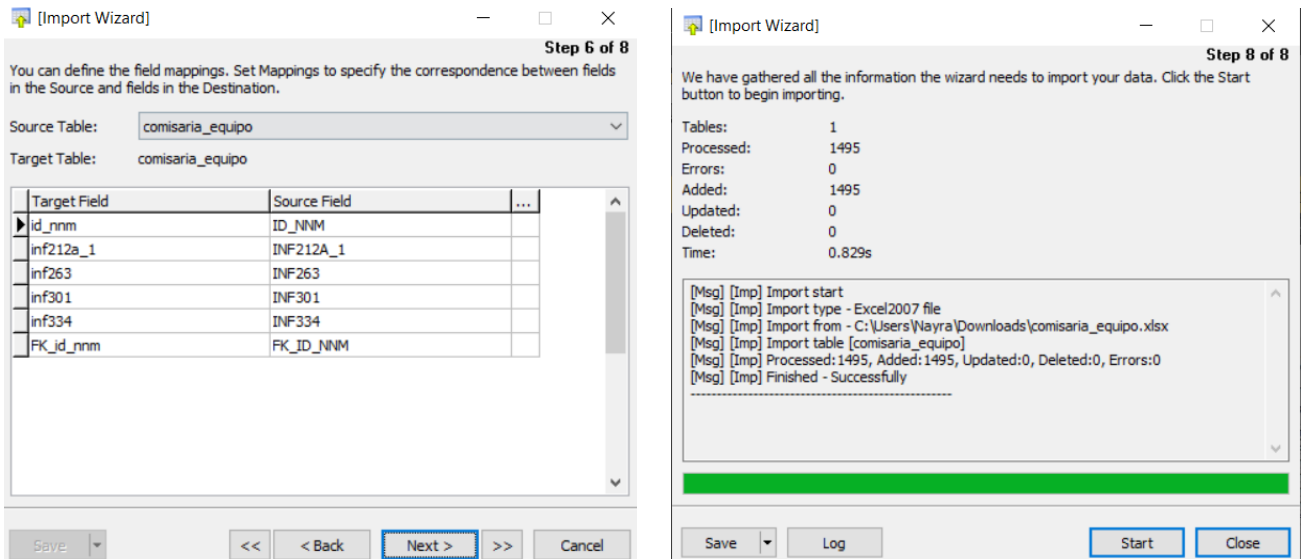


Figura 19: Carga de los datos (importación de Excel al MySQL) a la dimensión COMISARIA\_EQUIPO.

## ✓ FAC\_TABLE

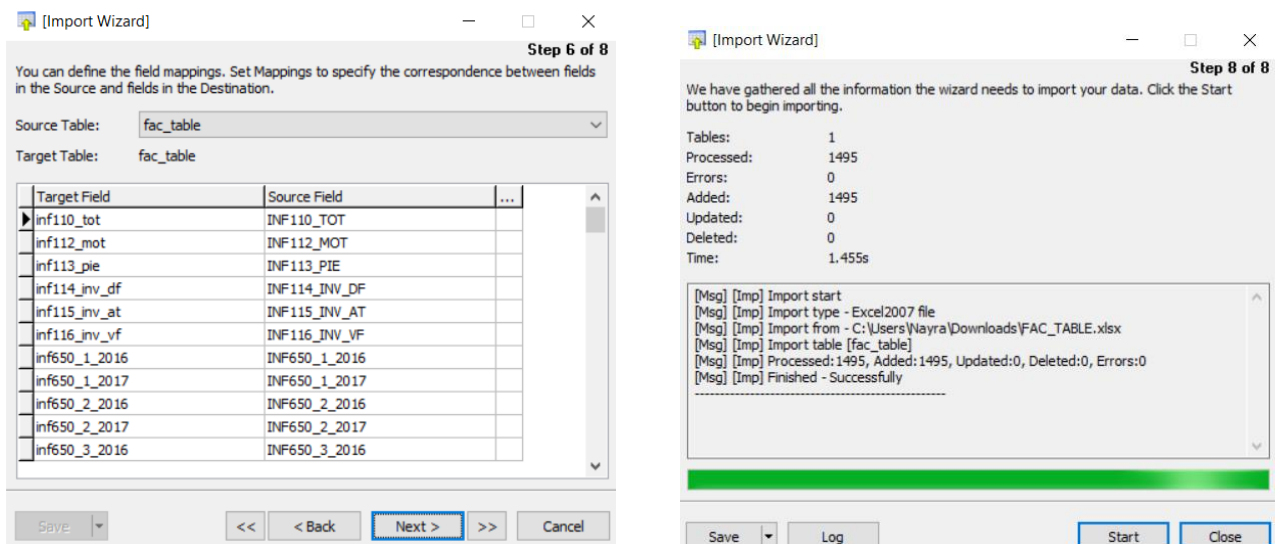


Figura 20: Carga de los datos (importación de Excel al MySQL) a la métrica FAC\_TABLE.

Como todas las dimensiones y el fac table ya están creadas en MySQL, entonces ahora se puede aplicar técnicas estadísticas y creación de Dashboard en Power BI.

## 5.2. EXPLOTACIÓN DE DATOS

### Técnica de Clasificación: Cluster Jerárquico

Para realizar este método de clasificación usaremos el número de personas detenidas y el número de bandas delictivas por comisaria en los años 2016 y 2017 para poder así clasificar las comisarias y a su vez zonas de mayor actividad delictiva.

- ✓ Lectura de datos y preprocesamiento

```
data <- read.xlsx("Datos_Seguridad.xlsx", sheetIndex = 4)
names(data)
```

```
## [1] "NOMBRE_COMISARIA" "ID_COMISARIA" "B_1_2016"
## [4] "B_1_2017" "B_2_2016" "B_2_2017"
## [7] "B_3_2016" "B_3_2017" "B_4_2016"
## [10] "B_4_2017" "B_5_2016" "B_5_2017"
## [13] "B_6_2016" "B_6_2017" "B_7_2016"
## [16] "B_7_2017" "B_8_2016" "B_8_2017"
## [19] "B_9_2016" "B_9_2017" "B_10_2016"
## [22] "B_11_2016" "B_12_2016" "BANDAS_TOT_2016"
## [25] "BANDAS_TOT_2017" "PI_1_2016" "PI_1_2017"
## [28] "PI_2_2016" "PI_2_2017" "PI_3_2016"
## [31] "PI_3_2017" "PI_4_2016" "PI_4_2017"
## [34] "PI_5_2016" "PI_5_2017" "PI_6_2016"
## [37] "PI_6_2017" "PI_7_2016" "PI_7_2017"
## [40] "PI_8_2016" "PI_8_2017" "PI_9_2016"
## [43] "PI_9_2017" "PI_10_2016" "PI_11_2016"
## [46] "PI_12_2016" "P_INTERV_TOT_2016" "P_INTERV_TOT_2017"
## [49] "PD_1_2016" "PD_1_2017" "PD_2_2016"
## [52] "PD_2_2017" "PD_3_2016" "PD_3_2017"
## [55] "PD_4_2016" "PD_4_2017" "PD_5_2016"
## [58] "PD_5_2017" "PD_6_2016" "PD_6_2017"
## [61] "PD_7_2016" "PD_7_2017" "PD_8_2016"
## [64] "PD_8_2017" "PD_9_2016" "PD_9_2017"
## [67] "PD_10_2016" "PD_11_2016" "PD_12_2016"
## [70] "DETENIDOS_TOT_2016" "DETENIDOS_TOT_2017" "CODIGO_NOMBRE"
```

```
#Reduciendo las variables a considerar para perfilar mejor los clusters
```

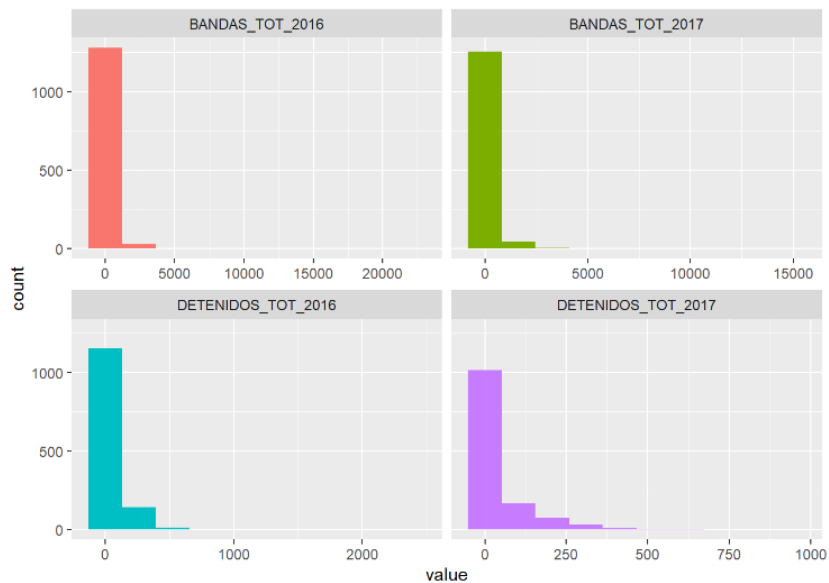
```
datac <- select(data, BANDAS_TOT_2016, BANDAS_TOT_2017, DETENIDOS_TOT_2016, DETENIDOS_TOT_2017)
str(datac)
```

```
## 'data.frame': 1310 obs. of 4 variables:
## $ BANDAS_TOT_2016 : num 0 0 0 100 0 0 800 150 300 0 ...
## $ BANDAS_TOT_2017 : num 50 300 0 900 0 0 1000 200 750 0 ...
## $ DETENIDOS_TOT_2016: num 32 116 0 97 42 47 177 212 129 23 ...
## $ DETENIDOS_TOT_2017: num 125 62 0 283 35 60 327 195 106 10 ...
## - attr(*, "na.action")= 'omit' Named int [1:189] 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 ...
## .. attr(*, "names")= chr [1:189] "1311" "1312" "1313" "1314" ...
```

```
# Gráfico de variables numéricas
```

```
plot_num(datac)
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.
```



## ✓ Elaboración del modelo de clasificación

```
#-----  
  
# Cluster Jerárquico Aglomerativo: AGNES  
  
# Usando la distancia euclidiana  
d <- dist(datac, method = "euclidean")  
  
# Cluster Jerárquico usando el método de enlace average  
res.hc <- hclust(d, method = "average" ) # calcula el cluster jerárquico  
  
res.hc
```

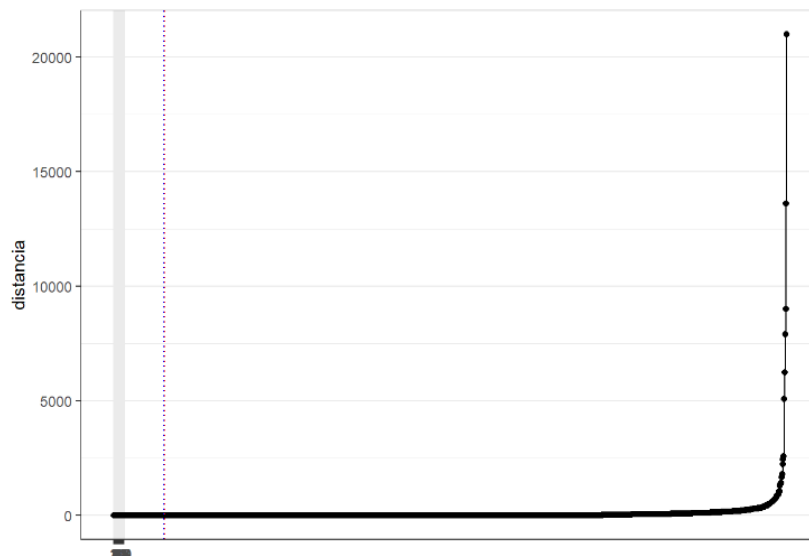
```
##  
## Call:  
## hclust(d = d, method = "average")  
##  
## Cluster method : average  
## Distance : euclidean  
## Number of objects: 1310
```

```
## List of 7
## $ merge      : int [1:1309, 1:2] -3 -20 -22 -40 -41 -55 -65 -66 -67 -68 ...
## $ height     : num [1:1309] 0 0 0 0 0 0 0 0 0 ...
## $ order      : int [1:1310] 665 447 801 80 673 1059 379 329 1072 357 ...
## $ labels     : chr [1:1310] "1" "2" "3" "4" ...
## $ method     : chr "average"
## $ call       : language hclust(d = d, method = "average")
## $ dist.method: chr "euclidean"
## - attr(*, "class")= chr "hclust"
```

```
# Proceso de agrupamiento indicando los individuos
res.hc$merge # hace un ranking de los individuos mas parecidos
```

## ✓ Calculando el criterio de elección del número de Clusters

```
#Grafica
library(ggplot2)
ggplot(alturas) + aes(x=etapa,y=distancia) + geom_point() + geom_line()+
  scale_x_continuous(breaks=seq(1,20)) +
  geom_vline(xintercept = 99,col="red",lty=3) +
  geom_vline(xintercept = 98,col="blue",lty=3) +
  theme_bw()
```



```
#luego se cuenta los puntos desde la vertical de referencia, hay 2 cluster

# Dividir en 2 clusters
grp <- cutree(res.hc, k = 2) # funcion que forma cluster y te las da en una columna
grp
```

## ✓ Gráfico final de los Clusters

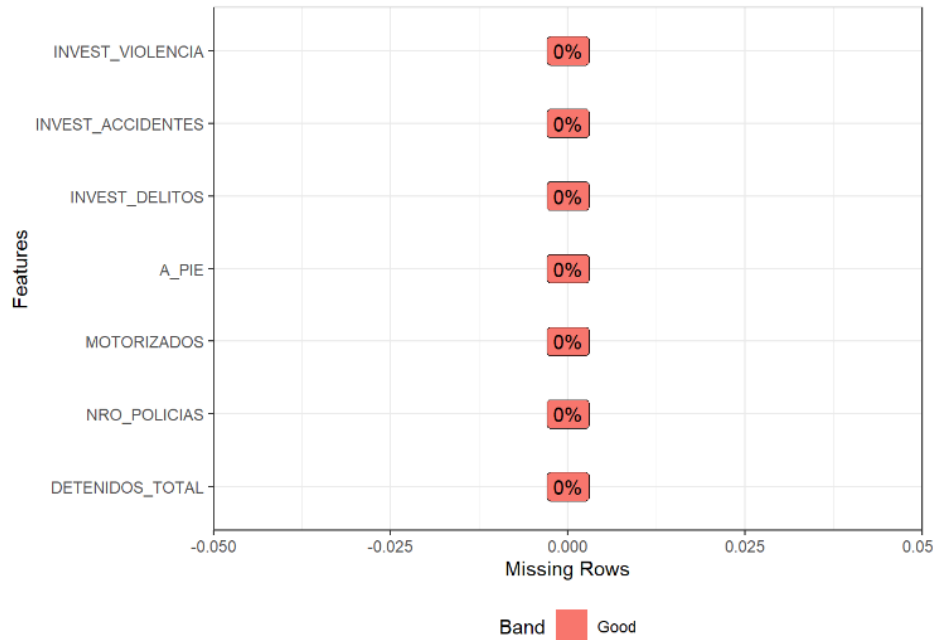
```
# Ploteo de Cluster en poligonos
row.names(datac) <- data$CODIGO_NOMBRE

fviz_cluster(list(data = datac, cluster = grp),
  palette = c("#2E9FDF", "#E7B800"),
  ellipse.type = "convex", # Concentration ellipse
  repel = F, # Avoid label overplotting (slow)
  show.clust.cent = FALSE, ggtheme = theme_minimal())
```



```
# Detectando y graficando los % de datos perdidos
```

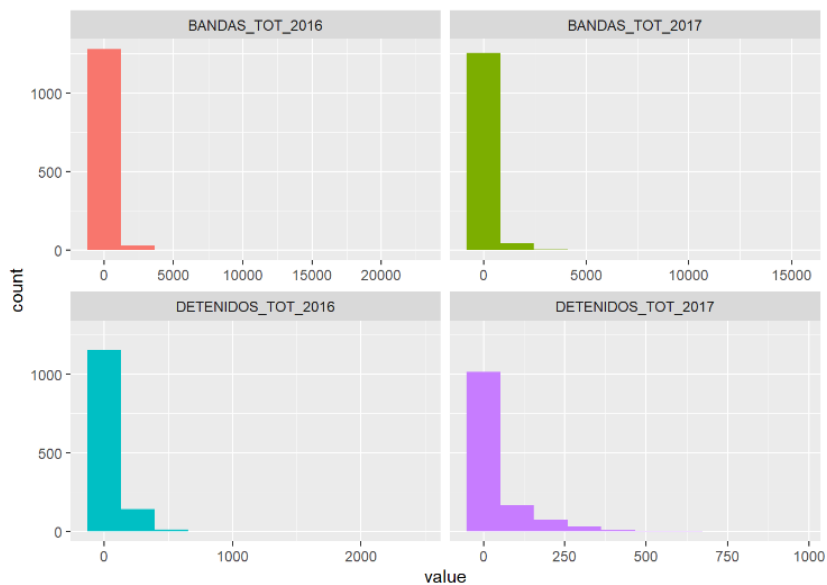
```
plot_missing(data, ggtheme=theme_bw()) +  
  labs(title = "Datos perdidos por variable", y = "Datos perdidos",  
        x = "Variables")
```



```
# Gráfico de variables numéricas
```

```
plot_num(datac)
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.
```



## ✓ Elaboración del modelo predictivo

```
# Elaboracion del modelo predictivo
```

```
modelo0 <- lm(DETENIDOS_TOTAL~.,data=data)
summary(modelo0)
```

```
##
## Call:
## lm(formula = DETENIDOS_TOTAL ~ ., data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -170.15  -93.35  -79.02   -8.41  3085.25
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    103.3114     8.8994   11.609  <2e-16 ***
## NRO_POLICIAS     -0.4829     0.9416   -0.513    0.608
## MOTORIZADOS      0.6632     1.2585    0.527    0.598
## A_PIE            1.7370     1.3075    1.328    0.184
## INVEST_DELITOS    0.8085     3.0348    0.266    0.790
## INVEST_ACCIDENTES -4.0222     5.8715   -0.685    0.493
## INVEST_VIOLENCIA -2.3323     3.1136   -0.749    0.454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 205.9 on 1303 degrees of freedom
## Multiple R-squared:  0.004159,    Adjusted R-squared:  -0.0004267
## F-statistic: 0.907 on 6 and 1303 DF,  p-value: 0.4889
```

## ✓ Análisis de resultados

Podemos observar que con el modelo realizado obtenemos un R cuadrado igual a -0.0004267, lo cual significa que el modelo no tiene poder predictivo alguno y que las variables no tienen capacidad explicativa sobre la variable dependiente escogida. Esto nos haría optar por escoger otra variable dependiente e independientes, pero en el conjunto de datos la dependiente sería la única.

Al revisar los datos se puede observar que hay muchas comisarias que registran cero detenidos en un lapso de dos años, algo que no tiene ningún sentido lógico, por lo tanto, es muy probable que la columna de datos de detenidos esta corrupta.

### Dashboard en Power BI

- **Objetivo estratégico:**

- **Indicador:**

- ✓ Policías que patrullan entre el total de policías.
- ✓ Comisarias con sistema de denuncia en línea entre total de comisarias



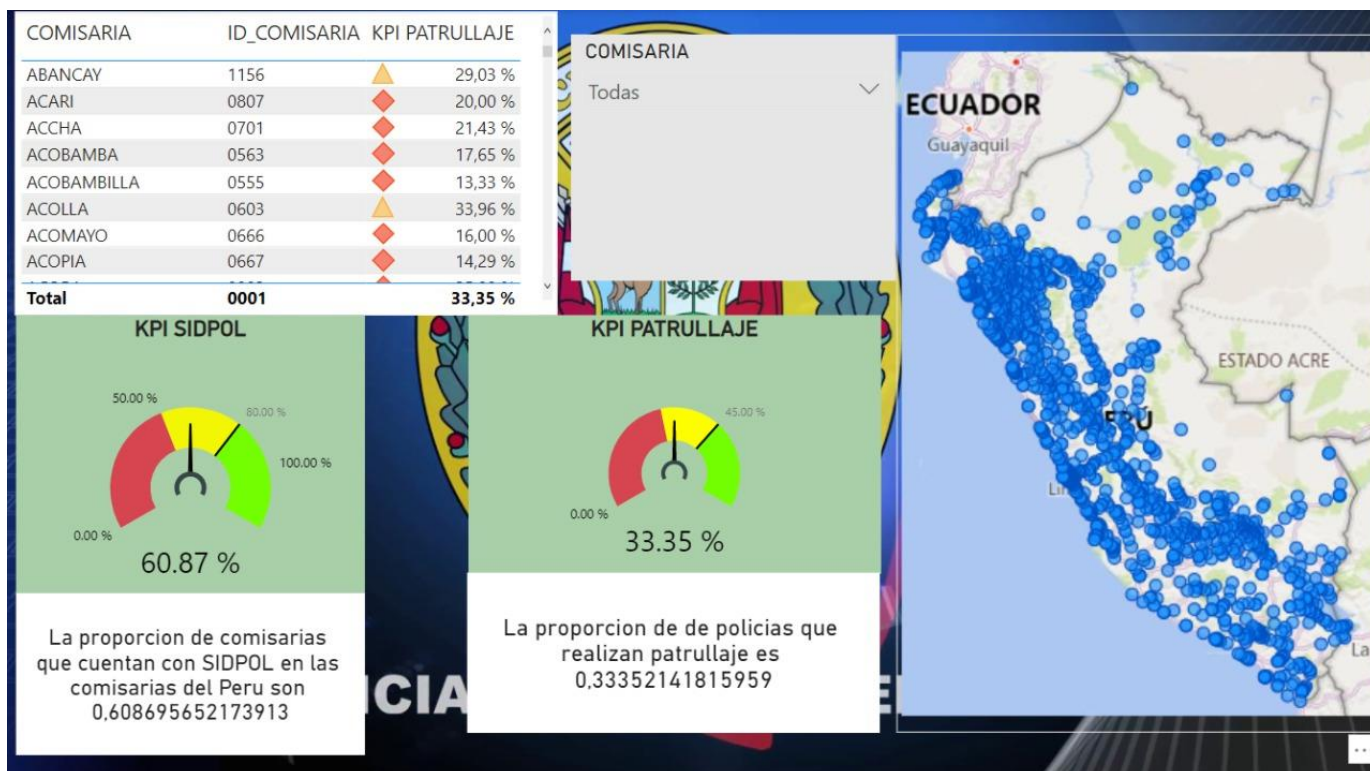


Figura 21: Dashboard para el objetivo estratégico desarrollado – Power BI

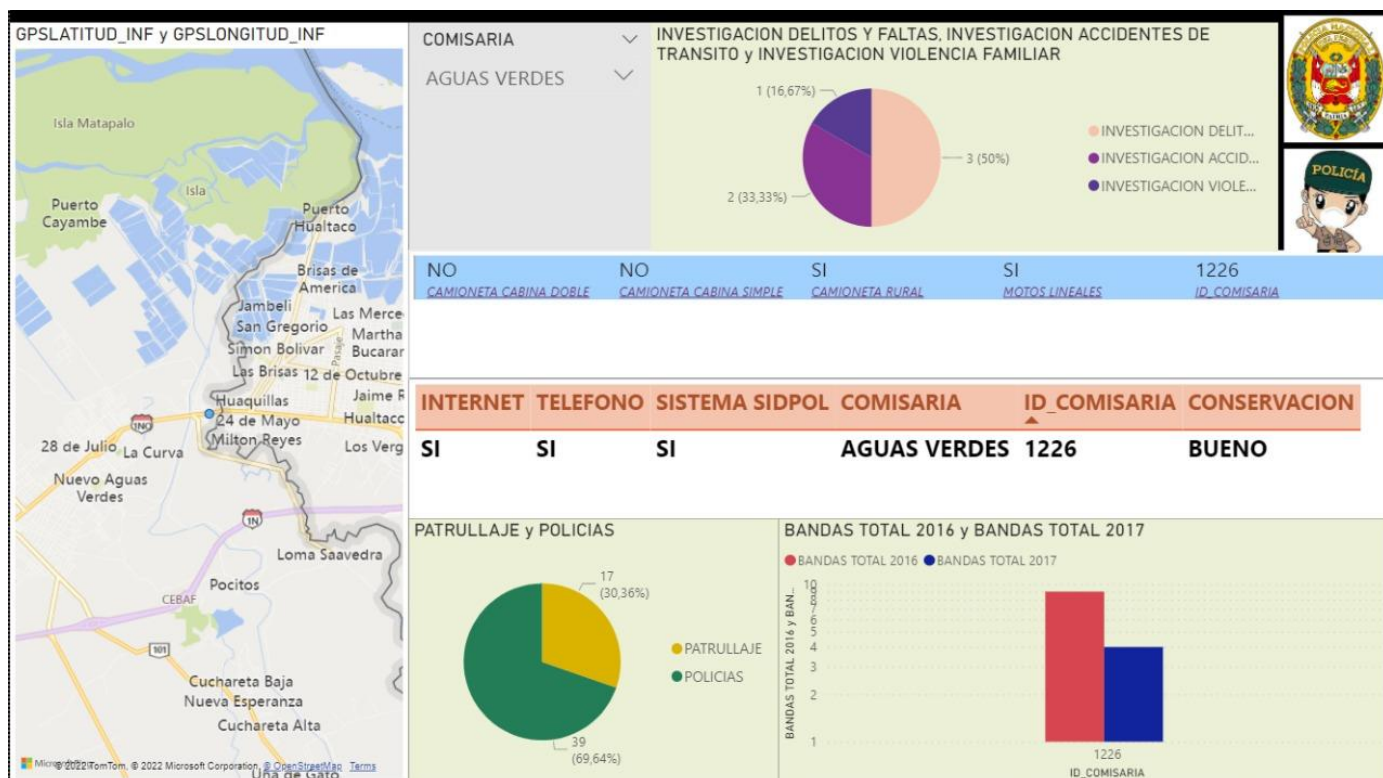


Figura 22: Dashboard para el objetivo estratégico desarrollado – Power BI

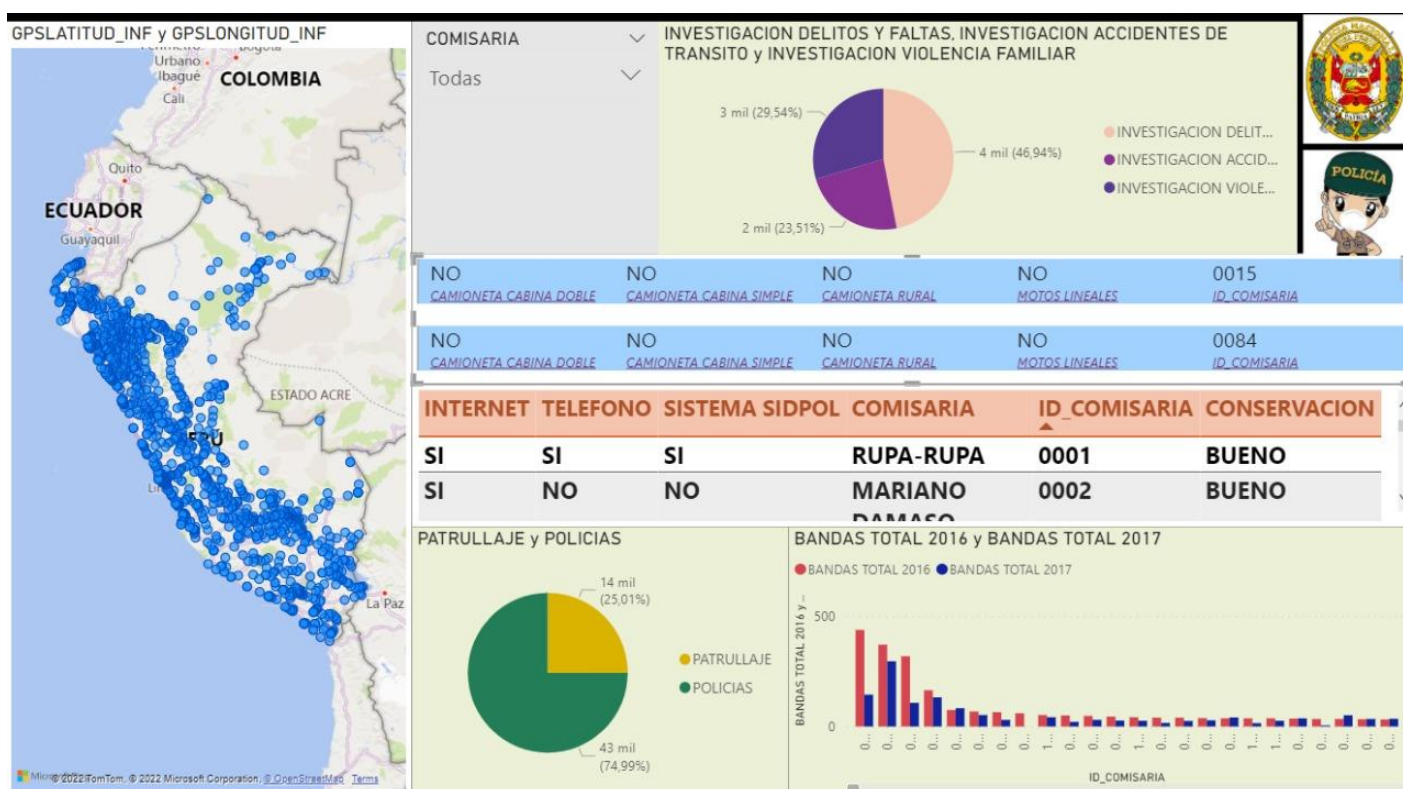


Figura 22: Dashboard para el objetivo estratégico desarrollado – Power BI

## 6. RESULTADOS

- En el dashboard y acerca de los KPIs planteados se concluye que en promedio el 33.35% de policías que se encuentran en una comisaría realizan labor de patrullaje, esto basándonos en que la meta de llegar al siguiente año 2018 es del 48%, aproximadamente la meta para el presente año 2017 sería de un 45%. No se ha llegado a la meta, pero tampoco se estuvo, por poco, en la zona “roja” el cual es menos del 30%
- También se concluye que el porcentaje de comisarías que cuentan con el Sistema Informático de Registro de Denuncias Policiales (SIDPOL) es del 60.87% la meta es del 80% aproximadamente este número alto se puede justificar con el constante desarrollo de la tecnología ya cada vez es más necesario este sistema en las comisarías. No se ha alcanzado a la meta del presente año, pero este KPI cada vez está en mayor crecimiento, al igual que el anterior KPI tampoco está en la zona “roja” el cual es menos del 50 % así que se encuentra en la “ámbar”.

## 7. CONCLUSIONES

- Este trabajo, contribuye a tener un control de las distintas medidas sobre las comisarías en el Perú, sobre todo del porcentaje de policías que realizan patrullaje y

del porcentaje de comisarias que tienen implementado el sistema de denuncias en línea.

- Se trabajó con las tablas generales que nos proporcionó el INEI y en base a ello pudimos integrar bien las distintas variables. Cabe mencionar que no es el único datamart que se podría sacar de las tablas ya que dependen de cuáles son tus objetivos estratégicos.
- Se pudo realizar la elaboración de un dashboard iterativo donde se muestran las comisarias con ubicación específica en el mapa además de gráficos que ayudan a obtener información de aquellas comisarias, además se presentó los dos KPIs con su respectiva representación del “semáforo” para cada comisaria y en total. Así con estos KPIs podemos tener respuesta a nuestros objetivos estratégicos.

## 8. RECOMENDACIONES

- Se recomienda, continuar la construcción del datamart con nuevos indicadores del área.
- incluir otras áreas, a fin de poder tener un lineamiento en función a la información de la institución policial.

## 9. BIBLIOGRAFIA

- Perversi, I. (2007). *APLICACIÓN DE MINERÍA DE DATOS PARA LA EXPLORACIÓN Y DETECCIÓN DE PATRONES DELICTIVOS EN ARGENTINA* [TESIS DE GRADO EN INGENIERÍA INDUSTRIAL, Instituto Tecnológico De Buenos Aires]. Biblioteca DR.Amadeo J.Galli  
<https://ri.itba.edu.ar/handle/123456789/3085>
- Suárez-Paez, J.; Salcedo-González, M.; CLIMENTE, A.; Esteve, M.; Gómez, JA; Palaos, CE; Pérez-Llopis, I. Un novedoso sistema de bajo tiempo de procesamiento para la detección de actividades delictivas aplicado a los Centros de Mando y Control de Seguridad Ciudadana. *Información* 2019, 10 , 365. <https://doi.org/10.3390/info10120365>
- Carpio, M., & Guerrero, M. (2014). *El efecto de la presencia policial sobre el delito en Perú. CIES: Documentos de trabajo. Lima, Perú.*
- Mora Luque, P. N. (2015). *Uso de tecnologías para sistematización de la información sobre el crimen* (usos, problemas de georreferencia y demás).
- Uriel, E. y Aldas, J. (2017). Análisis multivariante aplicado con R. Ediciones Paraninfo. Pág. 445 – Pág. 490
- James, G. (2013). *An Introduction to Statistical Learning with Applications in R*. Los Angeles, USA. Springer. 203-210 p.

## 10. ANEXOS

### ✓ CLUSTER

```
### Modelos de la Data_Seguridad ###

library(xlsx)

# Paquetes
library(pacman)
p_load(cluster,aplpack,fpc,foreign,TeachingDemos,
        factoextra,NbClust,ape,corrplot,DataExplorer,
        funModeling,compareGroups,tidyverse,dendextend,
        igraph,FeatureImpCluster,flexclust,LICORS)

setwd(dirname(rstudioapi::getActiveDocumentContext())$path)
getwd()

data <- read.xlsx("Datos_Seguridad.xlsx", sheetIndex = 4)
head(data)

#Elimina todas las filas que contengan algun valor NA

data <- na.omit(data)

data$BANDAS_TOT_2016 <- data$BANDAS_TOT_2016*50
data$BANDAS_TOT_2017 <- data$BANDAS_TOT_2017*50

# Estructura de los datos

str(data)

# Detectando y graficando los % de datos perdidos

plot_missing(data, ggtheme=theme_bw()) +
  labs(title = "Datos perdidos por variable", y = "Datos perdidos",
        x = "Variables")

#Reduciendo las variables a considerar para perfilar mejor los clusters
```

```

datac <-
select(data,BANDAS_TOT_2016,BANDAS_TOT_2017,DETENIDOS_TOT_2016,DETENIDOS_TOT_2017)
str(datac)

# Gráfico de variables numéricas

plot_num(datac)
#-----

# Cluster Jerárquico Aglomerativo: AGNES

# Usando la distancia euclidiana
d <- dist(datac, method = "euclidean")

# Cluster Jerárquico usando el método de enlace average
res.hc <- hclust(d, method = "average" ) # calcula el cluster jerárquico

res.hc

str(res.hc) # se ve la estructura

# Proceso de agrupamiento indicando los individuos
res.hc$merge # hace un ranking de los individuos mas parecidos

# Criterio de distancias con alturas
alturas <- data.frame(etapa=1:1309,distancia=res.hc$height)
alturas
#Grafica
library(ggplot2)
ggplot(alturas) + aes(x=etapa,y=distancia) + geom_point() + geom_line()+
  scale_x_continuous(breaks=seq(1,20)) +
  geom_vline(xintercept = 99,col="red",lty=3) +
  geom_vline(xintercept = 98,col="blue",lty=3) +
  theme_bw()
#luego se cuenta los puntos desde la vertical de referencia, hay 3 cluster

# Dividir en 3 clusters
grp <- cutree(res.hc, k = 3) # funcion que forma cluster y te las da en una
columna
grp

# Ploteo de Cluster en poligonos
row.names(datac) <- datac$CODIGO_NOMBRE

fviz_cluster(list(data = datac, cluster = grp),
  palette = c("#2E9FDF", "#E7B800", "#FC4E07"),
  ellipse.type = "convex", # Concentration ellipse

```

```
repel = F, # Avoid label overplotting (slow)
show.clust.cent = FALSE, ggtheme = theme_minimal())
```

## ✓ PREDICCIÓN

```
### Modelo de Prediccion ###
```

```
library(xlsx)
```

```
setwd(dirname(rstudioapi::getActiveDocumentContext())$path)
getwd()
```

```
# Lectura de datos
```

```
data <- read.xlsx("Datos_Seguridad.xlsx", sheetIndex = 2)
names(data)
```

```
data <- data[c(1:1310),]
```

```
# Detectando y graficando los % de datos perdidos
```

```
plot_missing(data, ggtheme=theme_bw()) +
  labs(title = "Datos perdidos por variable", y = "Datos perdidos",
        x = "Variables")
```

```
# Gráfico de variables numéricas
```

```
plot_num(data)
```

```
# Elaboracion del modelo predictivo
```

```
modelo0 <- lm(DETENIDOS_TOTAL~., data=data)
summary(modelo0)
```

