

**PROYECTO DE DATA SCIENCE**

**MINERIA DE DATOS PARA EL ANALISIS DE LOS DELITOS CIBERNETICOS**  
**EN COLOMBIA DURANTE EL AÑO 2015**

**PRESENTADO POR:**

JESUS DAVID SUAREZ PEÑA

**PRESENTADO AL DOCENTE:**

ALVARO AGUSTIN OÑATE BOWEN

**UNIVERSIDAD POPULAR DEL CESAR**

**FACULTAD DE INGENIERÍA Y TECNOLOGÍAS**

**INGENIERIA DE SISTEMAS**

**VALLEDUPAR - CESAR**

**2020**

## TABLA DE CONTENIDO

INTRODUCCIÓN.....	3
ESTADO DE LARTE.....	4
METODOLOGIA CRISP-DM.....	6
1. FASE 1: COMPRENSION DEL NEGOCIO .....	7
1.1. DELITOS CIBERNETICOS.....	7
1.2. AREA PROBLEMÁTICA.....	7
1.3. SITUACION ACTUAL.....	8
1.4. DEFINICION DE LOS OBJETIVOS COMERCIALES .....	9
1.5. VALORACION DE LA SITUACION.....	9
1.6 . RECURSOS UTILIZADOS.....	9
1.7. RECOPIACION DE DATOS INICIALES .....	10
1.8. TERMINOLOGIA.....	10
1.9. OBGETIVOS DE LA MINERIA DE DATOS .....	11
2: COMPRENSIÓN DE LOS DATOS .....	11
2. RECOPIACION DE DATOS INICIALES .....	11
2.1. DATOS EXISTENTES .....	11
2.2. ¿EXISTEN DATOS SUFICIENTES PARA OBTENER CONCLUSIONES GENERALES O REALIZAR PREDICCIONES PRECISAS? .....	12
2.3. ¿DISPONE DE ATRIBUTOS SUFICIENTES PARA SU MÉTODO DE MODELADO?.....	12
2.4. ¿HA CONSIDERADO CÓMO SE GESTIONAN LOS VALORES PERDIDOS EN CADA ORIGEN DE DATOS? .....	13
2.5. DESCRIPCION DE LOS DATOS.....	13
2.5.1. CANTIDAD DE DATOS Y CARACTERISTICAS DE LOS ATRIBUTOS .....	13
2.6. EXPLORACION DE DATOS .....	14

## INTRODUCCION

La variedad, amplitud y complejidad de los sistemas de información que adquieren, requieren o encuentran disponibles las organizaciones actuales, junto a la dinámica del permanente cambio observado en las tecnologías de la información y las comunicaciones, han impulsado de múltiples formas y, al mismo tiempo, condicionado las grandes transformaciones de las organizaciones, los mercados y el mundo de la modernidad y de la posmodernidad. Son cambios que, además de sus innegables ventajas, han traído simultáneamente para las personas y las organizaciones, amenazas, riesgos y espectros de incertidumbre en los escenarios de internet, intranet, desarrollo tecnológico, gestión de la información, la comunicación y los sistemas.

Con cada vez mayor frecuencia y mayor impacto, los dispositivos de almacenamiento y procesamiento de información -llámense servidores, estaciones de trabajo o simplemente PC- son vulnerados en sus elementos más sensibles, dejando expuestos no sólo múltiples y significativos datos de distinto valor (financiero, crediticio, estratégico, productivo...), sino los mismos patrimonios reales de personas y organizaciones y, aún más, su dignidad, su honra y su vida.

Con el avance de la tecnología informática y su influencia en casi todas las áreas de la vida social y empresarial, han surgido comportamientos ilícitos llamados de manera genérica *delitos informáticos*, que han abierto un amplio campo de riesgos y también de estudio e investigación, en disciplinas jurídicas y técnicas, pero especialmente en aquellas asociadas con auditoría de sistemas o auditoría informática.

## ESTADO DEL ARTE

La minería de datos se define como el proceso por el cual se descubren patrones en los datos que permitan una extracción de los mismos en un formato entendible y tratable. Es una de las técnicas más empleadas hoy en día en numerosos ámbitos, porque nos ayuda a conocer datos que nos conviene saber pero que no están a la vista, además de conocer eventos futuros, como, por ejemplo, conocer las modalidades, las causas y las circunstancias en que se llevan a cabo delitos cibernéticos, por lo que se han desarrollado numerosas aplicaciones y proyectos implementado la metodología CRISP-DM referido a lo antes mencionado

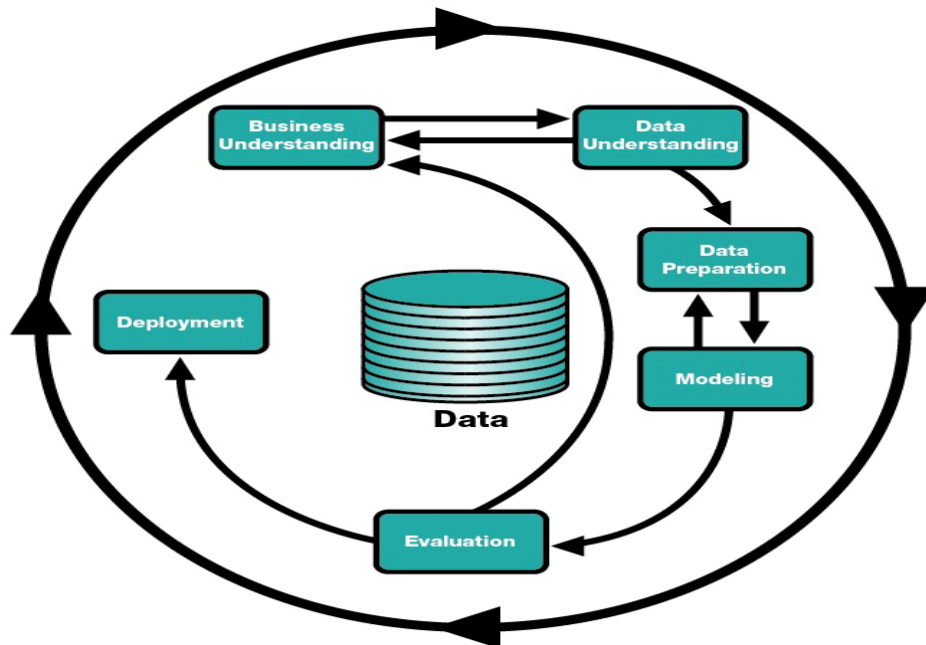
En primer lugar, Ignacio Perversi (2007) realizó un proyecto titulado “Aplicación de minería de datos para la exploración y detección de patrones delictivos en argentina” con el objetivo de aplicación de minería de datos para la exploración y detección de patrones delictivos en argentina, donde trabajan en un conjunto de datos para establecer la identificación y detección de patrones de homicidios dolosos cometidos en argentina durante 2005 en base a información suministrada por la dirección nacional de política criminal del ministerio de justicia y derechos humanos de la nación, organismo encargado de realizar las estadísticas oficiales de criminalidad en argentina.

Seguidamente se consideró el trabajo realizado por Andrea Licona Aguilar y María Camila Contreras B (2018). Llamado con el nombre de “Caracterización de los delitos en Cartagena mediante la aplicación de minería de datos” en lo cual se examinan específicamente 5 tipos de hurtos tales como automotores, motocicletas, residencias, establecimientos comerciales y entidades financieras en el transcurso de los años 2015 y 2016. Con el objetivo de realizar un análisis geoespacial que permita a los ciudadanos estar informados en lo correspondiente a los sucesos criminales que ocurren a su alrededor y de este modo mediante la divulgación del aplicativo web generado gracias al modelo matemático planteado

podrán tener una base sólida para tomar decisiones relacionadas con el lugar de vivienda, zonas para frecuentar, sectores peligrosos, entre otros aspectos.

Mediante la investigación de Carol Maribel Ñaupas Caraza (2016) llamando a su investigación con el nombre de “Minería de datos aplicada a la detección de fraude electrónico en entidades bancarias” donde propone a través de la aplicación de un proceso de descubrimiento de conocimientos en bases de datos, la generación de un modelo automático que permite clasificar las transacciones de la Banca por internet y de la Banca Móvil de personas naturales de una entidad financiera, como fraudulentas o íntegras, mediante la aplicación de técnicas de Minería Predictiva basada en arboles de Clasificación

## METODOLOGIA CRIPS-DM



La metodología CRISP-DM contempla el proceso de análisis de datos como un proyecto profesional, estableciendo así un contexto mucho más rico que influye en la elaboración de los modelos. Este contexto tiene en cuenta la existencia de un cliente que no es parte del equipo de desarrollo, así como el hecho de que el proyecto no sólo no acaba una vez se halla el modelo idóneo (ya que después se requiere un despliegue y un mantenimiento), sino que está relacionado con otros proyectos, y es preciso documentarlo de forma exhaustiva para que otros equipos de desarrollo utilicen el conocimiento adquirido y trabajen a partir de él.

## **FASE1: COMPRENSION DEL NEGOCIO**

### **1.1.DELITOS CIBERNETICOS**

Delito informático, delito cibernético o cibercrimen es toda aquella acción antijurídica que se realiza en el entorno digital, espacio digital o de Internet. Ante el extendido uso y utilización de las nuevas tecnologías en todas las esferas de la vida (economía, cultura, industria, ciencia, educación, información, comunicación, etc.) y el creciente número de usuarios, consecuencia de la globalización digital de la sociedad, la delincuencia también se ha expandido a esa dimensión. Gracias al anonimato y a la información personal que se guarda en el entorno digital, los delincuentes han ampliado su campo de acción y los delitos y amenazas a la seguridad se han incrementado exponencialmente.

Además de los ataques que tienen como objetivo destruir y dañar activos, sistemas de información u otros sistemas de computadoras, utilizando medios electrónicos y/o redes de Internet, se producen nuevos delitos contra la identidad, la propiedad y la seguridad de las personas, empresas e instituciones, muchos de ellos como consecuencia del valor que han adquirido los activos digitales para la big data empresarial y sus propietarios bien sean entes jurídicos o personas naturales. Existen también otras conductas criminales que, aunque no pueden considerarse como delito, se definen como ciberataques o abusos informáticos y forman parte de la criminalidad informática. La criminalidad informática consiste en la realización de un tipo de actividades que, reuniendo los requisitos que delimitan el concepto de delito, son llevados a cabo utilizando un elemento informático

### **1.2. AREA PROBLEMÁTICA**

Es común encontrar en internet, ya sea 'googleando' o en el buscador de alguna red social, que alguien, en cualquier lugar del mundo, ha sido víctima de fraude en la web. De hecho, a diario se puede ser víctima de este delito con correos

electrónicos que solicitan datos personales o páginas sospechosas pidiendo la actualización de su información bancaria.

Aunque no lo crea, cualquier tipo de intento de fraude, como el phishing o el robo de datos de páginas web inseguras, son los delitos informáticos que más se cometen en Colombia, seguido de la pornografía infantil que, en los últimos meses, se ha llevado toda la atención por el caso del 'Lobo Feroz'.

Según Wilson Pardo, subdirector de la Dijín, “fraudes como el phishing o modificación de portales web son los que más se investigan, e internacionalmente, son los que más se articulan. Pero hay que tener en cuenta que lo que les permite a los delincuentes actuar de esa forma es el comportamiento que tenemos en las redes”.

### **1.3. SITUACION ACTUAL**

Actualmente los delitos cibernéticos han aumentado debido al auge de las redes sociales y los dispositivos electrónicos, según cifras del Cuerpo Técnico de Investigación, CTI, mientras en 2015 entre enero y febrero los casos presentaron un aumento de 26,34%, este año, en el mismo periodo, el incremento fue de 60%.

Adicionalmente, el informe aseguró que nueve de 10 personas están expuestas a los delitos informáticos, desde el mismo momento en el que navegan en internet desde el celular o el computador, de ahí la necesidad de asumir comportamientos seguros.

En Colombia estos delitos se pagan con multas entre los 100 a 1.000 salarios mínimos legales mensuales vigentes, siempre que la conducta no constituya delito sancionado con pena más grave.

Pero si los delitos son graves pueden incluir penas de prisión, que van desde los 36 a los 96 meses (ocho años) dependiendo de la gravedad del delito informático



#### 1.4. DEFINICION DE LOS OBJETIVOS COMERCIALES

En el presente proyecto se aplicará minería de datos a través de la metodología CRISP-DM, con el cual encontraremos información que nos permitirá implementar estrategias que nos será útil para combatir los delitos cibernéticos.

- Identificar qué sitios son los más comunes en donde se presentan delitos cibernéticos.
- Determinar qué ciudad del país es la más afectada por delitos cibernéticos.
- Determinar los días de la semana en el que existen más riesgos de ser víctima de delitos cibernéticos.
- Identificar la hora del día donde se hace más frecuente los delitos cibernéticos.
- Identificar cuáles son los delitos cibernéticos más comunes que se presentan.

#### 1.5 VALORACION DE LA SITUACION

Para la implementación del proyecto en donde se podrá visualizar información obtenida, aplicar análisis estadísticos, entre otros se tiene a disposición un hardware con pocos requisitos.

- **DATOS:** Se dispone de un repositorio que posee todos los casos de delitos cibernéticos realizados en el año 2015. Los datos con lo que se llevará a cabo el desarrollo del proyecto se puede descargar desde el siguiente enlace: <https://www.datos.gov.co/Seguridad-y-Defensa/Delitos-Cibern-ticos/2tqs-zpz4>
- **PERSONAL:** Si dispone de una persona capacitada para llevar a cabo el desarrollo del proyecto.

- **FACTOR DE RIESGO:** Este proyecto se deberá de completarse para mediados del mes de junio, por lo que el tiempo, en el que se llevara a cabo el proyecto es el principal factor de riesgo.

## 1.6 RECURSOS UTILIZADOS

- Python
- Power Bi
- R studio

## 1.7 RECOPIACION DE DATOS INICIALES

Los datos con los que se realizara el proyecto se descargaron en la pagina <https://www.datos.gov.co/Seguridad-y-Defensa/Delitos-Cibern-ticos/2tqs-zpz4>

Esta página es utilizada como repositorio público donde se puede encontrar y descargar conjuntos de datos como son: datos del gobierno colombiano para investigar, desarrollar aplicaciones, crear visualizaciones e historias que pueden ser de utilidad para la realización de proyectos educativos o de cualquier tipo.

## 1.8 TERMINOLOGIA

- Delitos: Acción que va en contra de lo establecido por la ley y que es castigada por ella con una pena grave. Circunstancia de haber cometido una persona una acción contraria a la ley.
- La cibernética: es el estudio interdisciplinario de la estructura de los sistemas reguladores
- CTI: El Cuerpo Técnico de Investigación o CTI es una dirección de la Fiscalía General de la Nación del poder judicial de Colombia
- Fraude informático: es el acto de usar una computadora para tomar o alterar datos electrónicos, o para obtener un uso ilegal de una computadora o sistema.

## 1.9 OBJETIVOS DE LA MINERÍA DE DATOS

El proyecto desarrollará principalmente capacidades predictivas, y también descriptivas, se tienen los siguientes objetivos:

- Preparar un conjunto de datos apropiados para aplicar las diferentes técnicas de la minería de datos.
- Proponer un modelo descriptivo para la visualización de los datos.
- Aplicar técnicas para predecir patrones ocultos.

Criterios de la minería de datos:

- Lograr explorar los datos aplicando las técnicas de minería ya que esta va a permitir obtener unos resultados óptimos.
- Realizar cada una de las fases de la metodología CRISP-DM para tener más claridad de los datos y de la información que se puede obtener a través de estos.

## 2 FASE 2: COMPRESION DE LOS DATOS

### RECOPIACION DE DATOS INICIALES

#### 2.1. DATOS EXISTENTES

El conjunto de datos sobre delitos cibernéticos a nivel nacional del año 2015 está conformado por diferentes denuncias. El dataset está conformado por 14 atributos o columnas en total, los cuales son:

ATRIBUTO	DESCRIPCION
Fecha	Fecha en que fue realizado el delito
Departamento	Departamento donde ocurrió el delito
Municipio	Municipio donde ocurrió el delito

Dia	Dia de la semana en que ocurrió el delito
Hora	Hora en que ocurrió del delito
Barrio	Nombre del barrio donde ocurrió el delito
Zona	Zona donde ocurrió el delito URBANA, RURAL
Clasedecitio	Especificación de a que sitio u organización donde ocurrió el delito
Delito	Descripción del delito según el articulo
Armaempleada	Descripción del arma que usaron durante el delito
Movilagresor	Vehículo en el que se movilizo el agresor
Movilvictima	Vehículo en el que se movilizo la victima
Codigodane	Código identificador del dane
2015	Año

## 2.2 ¿EXISTEN DATOS SUFICIENTES PARA OBTENER CONCLUSIONES GENERALES O REALIZAR PREDICCIONES PRECISAS?

Este conjunto de dato cuenta con 7.531 registros de denuncias, los suficientes para el análisis y la realización del proyecto

## 2.3 ¿DISPONE DE ATRIBUTOS SUFICIENTES PARA SU MÉTODO DE MODELADO?

Para la realización de este proyecto de disponen de 11 atributos los cuales se estudiarán para determinar el método de modelo que se implementara, ya que estos contienen información significativa para sacar conclusiones acerca de los delitos cibernéticos en Colombia del año 2015.

## 2.4 ¿HA CONSIDERADO CÓMO SE GESTIONAN LOS VALORES PERDIDOS EN CADA ORIGEN DE DATOS?

Se ha considerado utilizar algún método estadístico como la media o la moda, percentil para completar o rellenar valores faltantes.

## 2.5 DESCRIPCION DE LOS DATOS

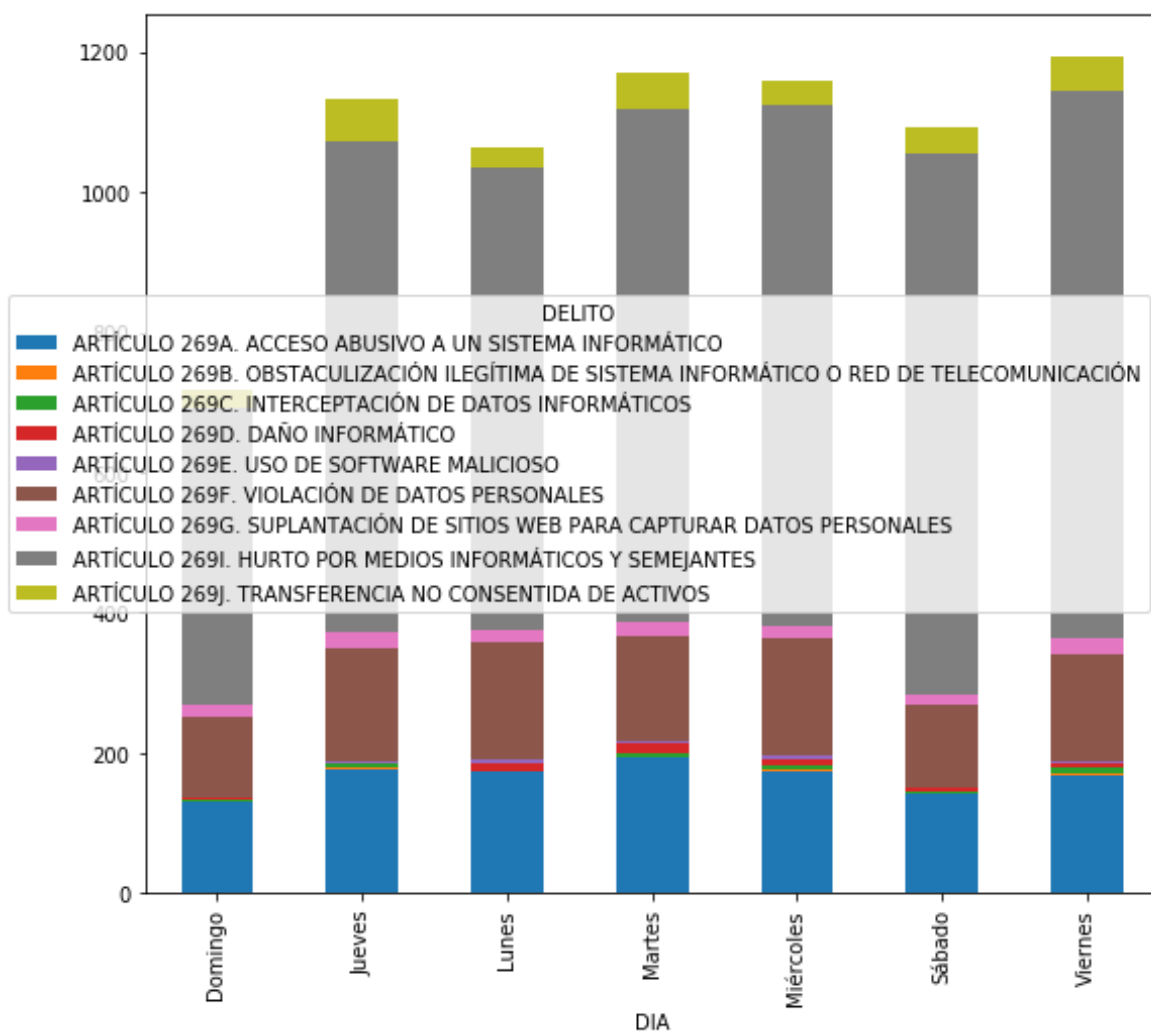
### 2.5.1 CANTIDAD DE DATOS Y CARACTERISTICAS DE LOS ATRIBUTOS

ATRIBUTO	TIPO
Fecha	Date
Departamento	Variable cualitativa nominal
Municipio	Variable cualitativa nominal
Dia	Variable cualitativa nominal
Hora	Variable cualitativa nominal
Barrio	Variable cualitativa nominal
Zona	Variable cualitativa nominal
Clasedecitio	Variable cualitativa nominal
Delito	Variable cualitativa nominal
Armaempleada	Variable cualitativa nominal
Movilagresor	Variable cualitativa nominal
Movilvictima	Variable cualitativa nominal
Codigodane	Numérico descrita
2015	Numérico descrita

## 2.6. EXPLORACION DE DATOS

Para explorar los datos de nuestro dataset con la que implementa el proyecto utiliza la herramienta Python y PowerBi, para generar graficas que nos ayudaran a entender los datos almacenados

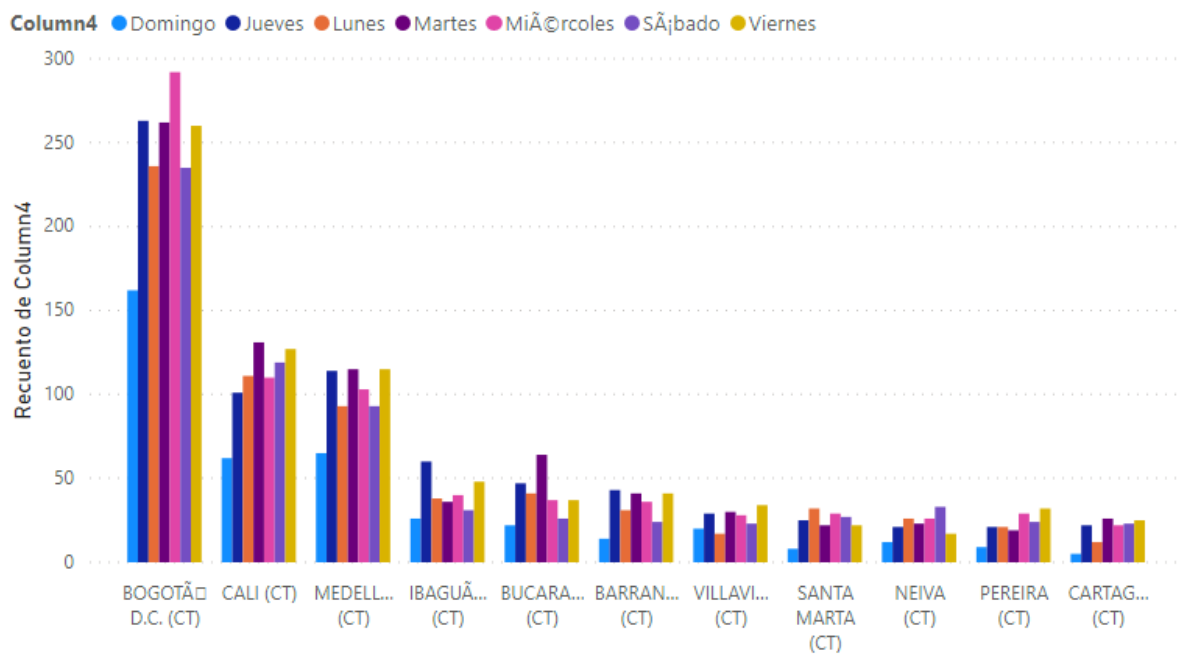
- ¿Según el día de la semana cuales son los delitos cibernéticos mas comunes?



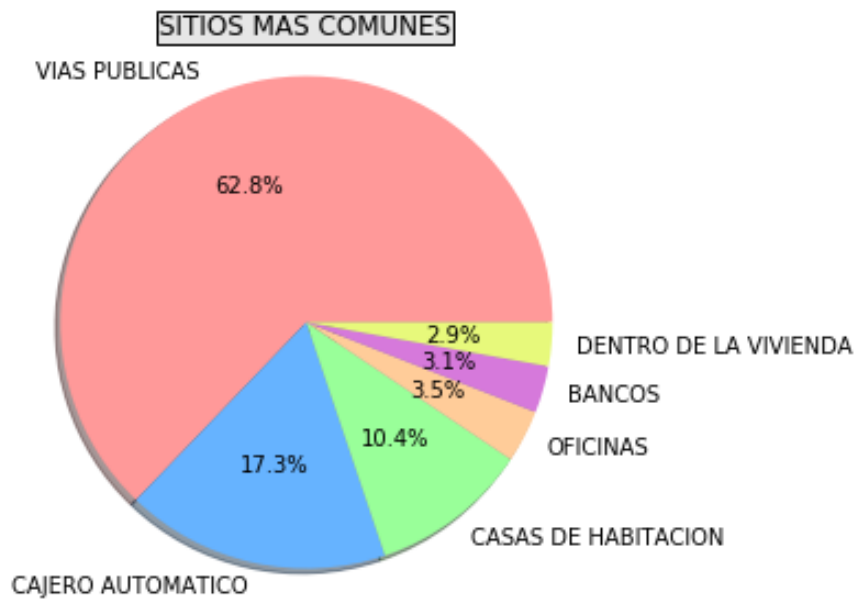
- ¿En qué parte de Colombia se presentan más casos de delitos cibernéticos?



- ¿Según las ciudades más afectadas en que días se presenta más delitos?



- Top 6 de los lugares mas comunes de los delitos cibernéticos



- ¿Según el día de la semana a que hora se comenten mas delitos cibereticos?

