Universidad Mariano Gálvez

Ingeniería en Sistemas de Información y Ciencias de la Computación

Noveno Ciclo

Proyecto de Graduación



Implementación de Modelos Estadísticos para el Análisis y Predicción A Través de la visualización de los Datos Criminales de la Subdirección General de Operaciones del Municipio de Guatemala Departamento de Guatemala mediante la Implementación del Lenguaje de Programación R

Marcos Mauricio Rivas López

Carne: 4490-14-2309

## Antecedentes

Existen varios estudios estadísticos realizados en Guatemala que tienen relación con datos criminales, aunque muchos de ellos solo se encargan de recolectar la información y colocarla en gráficas, ejemplos de este tipo de estudio pueden ser los reportes generados por el Instituto Nacional de Estadística, que se encargan de graficar toda la información y colocarle ciertos atributos a cada una de las gráficas como, por ejemplo, ascenso y descenso con respecto a otro periodo de tiempo, una descripción de los datos representados en dichas gráficas; ahora bien, todo esto es estadística, pero podríamos llamarle un estudio limitado ya que solo nos proporciona toda la información, que a pesar de que está ordenada no existe un ente encargado de darle sentido a todos estos datos, ya que esto permitiría saber, a parte del “qué” está pasando, el “porqué” está pasando y mejor aún el “cómo” está pasando. Existen algunos estudios guiados a resolver estas preguntas.

ci (Dudley, 2017) Escribió un artículo acerca del resultado de una larga investigación llevada a cabo por la organización periodística InSight Crime, la cual presenta varios datos acerca de la cantidad de información sobre crímenes en Guatemala para tratar de determinar si el narcotráfico contribuye a las tazas altas de violencia en ciertos departamentos del país, explica por ejemplo que un 49% de los homicidios ocurridos en Guatemala pertenece a la categoría “desconocido”, siendo ésta una categoría que no nos proporciona mayor información acerca de preguntas clave para saber que pasó. Ellos mencionan que todos estos datos son recolectados por personas que no están capacitadas para el análisis del crimen, lo que no permite obtener información confiable para relacionar una gran cantidad de asesinatos a una o varias organizaciones criminales; incluso menciona que los altos mandos de la policía nacional civil y el ministerio público reportan información inexacta y que un 75% de los homicidios ocurridos en Guatemala es por medio de un arma de fuego, a pesar de que el reporte de armas registradas con respecto a su estimado es mucho menor en Guatemala que en otros países de Centroamérica.

Causa

No se ha destinado presupuesto para el desarrollo de tecnologías que faciliten los procesos internos de la organización.

Problema

No existe una herramienta específica para el almacenamiento y visualización de los datos estadísticos de una manera sencilla y eficaz

Consecuencia

El ingreso y tratamiento de los datos es muy tardío y difícil de interpretar

¿Qué puedo hacer?

Desarrollar un software para facilitar la interpretación de los datos estadísticos

# Planteamiento del Problema

A nivel Latinoamérica ocurre que la implementación de tecnología en los procesos de las distintas instituciones es muy escaza, ya que la prioridad de muchos países en dicha zona es cubrir las necesidades básicas antes que la tecnología, por lo que, si nos orientamos más en herramientas de software para la recolección y tratamiento de datos estadísticos los resultados serán aún más bajos.

Así mismo en Guatemala específicamente en el departamento de Guatemala existen diferentes organizaciones gubernamentales y no gubernamentales que se dedican a realizar estudios estadísticos en diferentes partes del país y se observa que no cuentan con una herramienta especializada para facilitar la visualización de los resultados.

En la Subdirección General de Operaciones de la Policía Nacional Civil que se encuentra ubicada en el departamento de Guatemala también se ha observado que no existe una herramienta específica para el almacenamiento y visualización de los datos estadísticos de una manera sencilla y eficaz.

Entre las múltiples causas que pueden estar originando este problema la que hemos detectado es que no se ha destinado el presupuesto suficiente para el desarrollo de tecnologías que faciliten los procesos de la institución.

Las consecuencias que puede traer este problema son muchas, entre las que he detectado están que el ingreso y tratamiento de los datos es muy tardío y difícil de interpretar, además de que se utilizan herramientas por separado y se requiere mano humana para llevar acabo todo el proceso por el cuál pasan los datos.

Lo que podría hacer como profesional del área de tecnología es desarrollar e implementar un software para agilizar el proceso de ingreso e interpretación de los datos, así mismo aplicar ciencia de datos para descubrir nuevas formas de interpretación y aplicarlas para generar información útil.

## Formulación del Problema

Las consideraciones expuestas nos llevan a plantear la siguiente interrogante:

¿Qué relación existe entre la implementación de tecnologías de software específicas que aplican ciencia de datos, con la interpretación de los datos estadísticos de la Subdirección General de Operaciones de la Policía Nacional Civil?

¿Puede la implementación de tecnologías de software específicas que aplican ciencia de datos mejorar la interpretación de los datos estadísticos de la Subdirección General de Operaciones de la Policía Nacional Civil del departamento de Guatemala?

# Implementación de tecnologías de software específicas que aplican ciencia de datos para mejorar la interpretación de los datos estadísticos de la Subdirección General de Operaciones del departamento de Guatemala

## JUSTIFICACIÓN

La falta de información criminal concreta en Guatemala causa una angustia en las personas que son víctimas de los diferentes hechos criminales ocurridos en el país ya que saben que es muy probable que nunca se llegue a dar con el culpable, a la vez causa miedo en las personas que viven en zonas de alta criminalidad ya que los diferentes grupos criminales como pandillas o estructuras que pertenecen a la categoría denominada crimen organizado poseen mucho poder y ni siquiera la policía puede actuar para que las diferentes operaciones de estos grupos termine.

Es necesario darle una solución a este problema, existen varias posibilidades, pero para cada una de ellas se tienen que ir evaluando diferentes factores para prevenir fallar en el intento, uno de los principales factores es la falta de colaboración de los empleados de las organizaciones que recolectan datos criminales para que ejecuten su trabajo de la mejor manera, aunque puede existir una falla no en los empleados si no en los procesos, para verificar esto se necesitaría una evaluación a la organización lo que se sabe que también sería tardío y costoso.

Lo que necesitamos es un proceso para convertir todos estos datos estadísticos en respuestas hacia preguntas que son necesarias de responder, porque conforme más preguntas se vayan respondiendo, mejor conoceremos la situación criminal actual en el país, por lo que, en este caso la mejor solución es un sistema informático que basado en la recolección y procesamiento de los diferentes datos les aplique procesos estadísticos para generar respuestas e incluso predecir tendencias criminales, este tipo de solución sería de bajo costo en comparación al costo de contratar una serie de empleados que realicen estos procesos de una manera manual, y sin la posibilidad de reutilizar procesos para aplicar estadística a diferentes conjuntos de datos; fue por estos factores que se escogió este tipo de solución.

## SUPUESTOS Y EXPECTATIVAS DEL TEMA

### Supuestos

Los supuestos que se tienen para el desarrollo del presente proyecto de investigación son:

* Se contarán con múltiples series de datos en formato de Excel para la realización de pruebas.
* El presente trabajo de investigación será guiado por 2 profesionales del área de sistemas y el área de investigación.

## Expectativas

Se espera que este sistema ayuda a resolver preguntas claves para la aclaración de tendencias criminales en el país y que la mejora en estos procesos ayude a:

* Entender con claridad la situación criminal actual de país, sabiendo así las zonas rojas, y darles sentido a hechos de violencia ocurridos en las mismas.
* Predecir hechos delictivos en base a los datos que anteriormente se tienen.
* Agilizar el proceso de presentación de datos hacia los altos mandos de la Policía Nacional Civil y el Ministerio Público, para que así se logre un eficiencia en las oficinas de estas organizaciones.

Se espera que el sistema realizado ayude en varios procesos para la definición correcta de la información criminal, y pueda expandir su funcionalidad hacia el análisis de datos de otras áreas de interés.

## Planteamiento del Problema

A nivel mundial ocurre

## Alcances

1. Se utilizarán los datos recolectados por la Subdirección General de Operaciones que opera en el municipio de Guatemala como muestra para la aplicación de los modelos estadísticos.
2. La construcción del sistema se llevará a cabo en un tiempo tentativo de 4 a seis meses.
3. Se tiene la posibilidad de desarrollar el sistema en un lenguaje de programación de código abierto

## Límites

1. Muy poca cantidad de datos y piezas faltantes de información.
2. Información no ordenada por ningún patrón de datos.

## Hipótesis

La implementación de modelos estadísticos para el análisis y predicción de datos ayudan a tener un mejor panorama de algunos factores causantes de los hechos criminales ocurridos en áreas determinadas del país.

Copyright © 2015 por Cristian Rojas & “ESTA PAGINA ES OPCIONAL” Miguel Gutiérrez. Todos los derechos reservados.

**Dedicatoria**

ESTA PÁGINA ES OPCIONAL.

Dedicamos esta plantilla a los usuarios de nuestros sitios: normasapa.com, normasieee.com, normasicontec.org.**Agradecimientos**

Gracias por su preferencia de normasapa.com, no olviden recomendarnos con sus colegas y compañeros. ESTA PAGINA ES OPCIONAL **Abstract**

Este es un documento de Word de ejemplo que puede ser usado como plantilla para dar formato a su tesis o disertación. El abstract o resumen debe contar con **350 palabras o menos**.

**Prefacio**

Esta página es opcional.**Tabla de Contenidos**

[Capítulo 1 Introducción e información general 2](#_Toc410628920)

[Título 2 2](#_Toc410628921)

[Título 2 2](#_Toc410628922)

[Título 3. 2](#_Toc410628923)

[Título 3. 2](#_Toc410628924)

[Capítulo 2 Figuras y tablas 2](#_Toc410628925)

[Título 2 2](#_Toc410628926)

[Título 3. 2](#_Toc410628927)

[Título 3. 2](#_Toc410628928)

[Capítulo 4 Resultados y discussion. 2](#_Toc410628929)

[List of References 2](#_Toc410628930)

[Apéndice 2](#_Toc410628931)

[Vita 2](#_Toc410628932)

**Lista de tablas**

[Tabla 1. El título debe ser breve y descriptivo. 2](#_Toc410629016)

**Lista de figuras**

[Figura 1. Formas y descripción de las formas. 2](#_Toc410629185)

# Capítulo 1 Introducción e información general

## Título 2

Debe haber solo un salto de línea entre párrafo y párrafo ,este salto de línea se puede hacer presionando la tecla ENTER.

Para añadir un capitulo adicional se debe crear un salto de página entre los dos capítulos, esto se puede hacer tecleando CTRL + ENTER al final del párrafo previo al nuevo párrafo.

## Título 2

Usa los subtítulos consistentemente. Revisando constantemente el espaciado, mayúsculas y puntuación.

### Título 3.

El uso de estilos es de ayuda a la hora de generar una tabla de contenidos. Este documento de ejemplo usa los títulos, subtítulos y demás estilos para generar automáticamente la tabla de contenido, lista de tablas y lista de figuras. Este documento está configurado para seguir las normas APA.

### Título 3.

Acá puede ir otra idea del documento.

# Capítulo 2 Figuras y tablas

Las tablas y figuras junto con el texto deben ser puestos en la misma página donde son mencionados por primera vez en el texto. Las tablas y figuras grandes deben ser agregadas en una página separada. La tabla 1 es más grande que media página y por lo tanto fue agregada en una página para sí misma. La página antes de la figura debe ser una página llena de texto a menos que esta esté al final del capítulo. Esto aplica incluso si un párrafo debe ser dividido en varias páginas.

## Título 2

Tablas y figuras deben ser puestas en páginas diferentes independientemente de su tamaño. No se debe dejar espacios en blanco en las páginas de texto, pero es posible dejar espacio en blanco en páginas que solo contienen tablas y figuras.

### Título 3.

Tablas y figuras pueden ser puestas en un apéndice al final de la tesis o disertación. Si se hace esto se debe estar seguro de indicar que las tablas y figuras están ubicadas en el apéndice. Esto puede ser a través de paréntesis o con pies de página. Es posible poner todas o solo algunas de las tablas y figuras en el apéndice, si todas las tablas y figuras son puestas en el apéndice se debe indicar que “Todas las tablas y figuras están ubicadas en el apéndice” después de la primera mención de una tabla o figuras..

### Título 3.

Los títulos de las tablas deben ser puestos sobre las mimas. En el caso de las figuras deben ser puestos debajo. Todas las tablas deben contar con mínimo 2 columnas y una fila de títulos. Las tablas deben contar a menos con 3 líneas divisorias.

Tabla 1. El título debe ser breve y descriptivo.

|  |  |
| --- | --- |
| ***Column One*** | ***Column Two*** |
| Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data | Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data  Table data |

Estas líneas son la línea incluida en la parte superior de la tabla, la línea entre el la cabecera de la tabla y el contenido y la línea debajo de la tabla.

#### Título 4.

#### Las figuras pueden estar blanco y negro o a color. Si se usa color se debe asegurar que la figura tenga sentido si se imprime a blanco y negro.En la figura 1 se muestran algunas formas.

Figura 1. Formas y descripción de las formas.

# Capítulo 4 Resultados y discusión.

Más texto.

# Lista de referencias

Andrews, S. Fastqc, (2010). A quality control tool for high throughput sequence data.

Augen, J. (2004). Bioinformatics in the post-genomic era: Genome, transcriptome, proteome, and information-based medicine. Addison-Wesley Professional.

Blankenberg, D., Kuster, G. V., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., ... & Taylor, J. (2010). Galaxy: a web‐based genome analysis tool for experimentalists. Current protocols in molecular biology, 19-10.

Bolger, A., & Giorgi, F. Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data. URL http://www. usadellab. org/cms/index. php.

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., ... & Nekrutenko, A. (2005). Galaxy: a platform for interactive large-scale genome analysis. Genome research, 15(10), 1451-1455.

# Apéndice

Las tablas y figuras pueden ir en el apéndice como se mencionó anteriormente. También es posible usar el apéndice para incluir datos en bruto, instrumentos de investigación y material adicional.

# Vita

Acá se incluye una breve biografía del autor de la tesis.