

**Colegio Universitario de Cartago**

**Escuela de Tecnologías de la Información**

**Big Data**

**Curso:** Minería de Datos

**Título:**

**Análisis Dataset Crímenes**

**en Costa Rica**

**Autores:** Roberto Montoya Leiva

Fiorella Abarca Valverde

**Profesor:** Osvaldo González Chaves

Cartago, Costa Rica

2024

Contents

[**INTRODUCCION** 3](#_Toc164358445)

[**OBJETIVO GENERAL:** 4](#_Toc164358446)

[**OBJETIVOS ESPEFICOS:** 4](#_Toc164358447)

[**DESARROLLO:** 5](#_Toc164358448)

[**Definición de hipótesis para análisis** 5](#_Toc164358449)

[**Adquisición del conjunto de datos** 5](#_Toc164358450)

[**Exploración y limpieza de los datos.** 7](#_Toc164358451)

[**ANALISIS DE LOS DATOS.** 14](#_Toc164358452)

[**CONCLUSION:** 20](#_Toc164358453)

[**BIBLIOGRAFÍA:** 21](#_Toc164358454)

# **INTRODUCCION**

En la actualidad, en Costa Rica nos encontramos con una ola de criminalidad en aumento desde los últimos años. Se han reportado asaltos en paradas del servicio público e incluso dentro de los mismos autobuses, robos de vehículos, asaltos a centros comerciales, además de la posibilidad de tiroteos en la zona, lo que pone en riesgo la seguridad de las personas y sus familias.

Durante el 2023, el tema de la inseguridad y la delincuencia se posicionó como la principal preocupación de las y los costarricenses, en las tres Encuestas de Opinión Pública realizadas por el Centro de Investigación y Estudios Políticos (CIEP) de la Universidad de Costa Rica este año.

Los resultados de la última encuesta señalan que la mayoría de las y los costarricenses considera que la inseguridad y la delincuencia son el principal problema del país (41,3%), seguido por desempleo (12,5%) y el costo de vida (10,8%). Esto evidencia un aumento en la preocupación de los ticos respecto a la inseguridad y la delincuencia que atraviesa Costa Rica, con una subida de 13,5 puntos porcentuales respecto a la medición de setiembre, y 6,6 puntos por encima de lo percibido en abril del 2023.

Por ello, nos parece importante realizar un análisis de la criminalidad en Costa Rica en los últimos 10 años. Esto con el fin de tener un panorama del comportamiento de estos hechos, con una mejor estructura y visualización de los datos para su mejor entendimiento.

# **OBJETIVO GENERAL:**

Encontrar patrones criminalísticos en el país con el fin de determinar qué zonas son más propensas a delitos de tipo homicidios.

# **OBJETIVOS ESPEFICOS:**

* Definición de hipótesis para análisis
* Adquisición del conjunto de datos
* Análisis y limpieza de los datos
* Creación de consultas para responder a las hipótesis
* Visualización de los resultados
* Análisis basado en los resultados

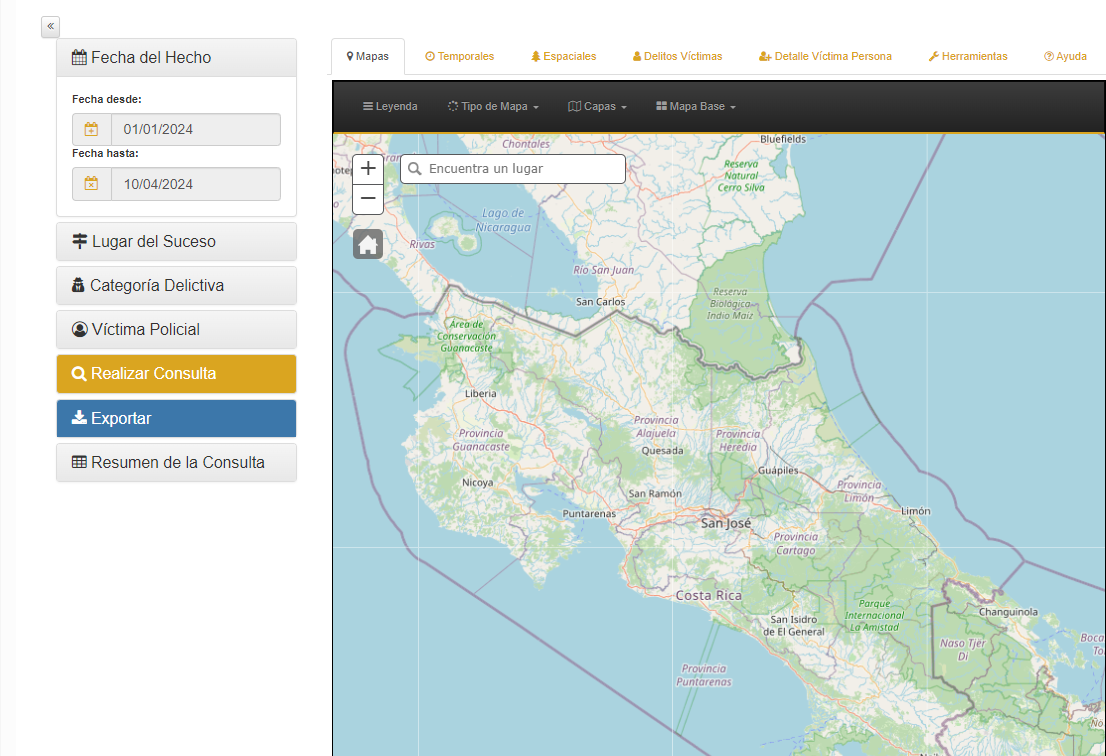
# **DESARROLLO:**

## **Definición de hipótesis para análisis**

1. *Las áreas específicas en Costa Rica donde son más recurrentes los homicidios se identifican claramente.*
2. *Las áreas específicas en Costa Rica donde son menos recurrentes los homicidios se identifican claramente.*
3. *Existe una variación significativa en la proporción de homicidios entre diferentes géneros.*
4. *Algunas nacionalidades presentan una tasa de homicidios significativamente mayor en comparación con otras.*
5. *Hay una diferencia significativa en la tasa de homicidios entre menores de edad y mayores de edad.*

## **Adquisición del conjunto de datos**

Para iniciar con el proyecto se necesita adquirir la data a la cual se le realizará el análisis, en este caso en particular “Análisis de la Criminalidad en CR” se adquiere por medio del siguiente enlace:  
  
<https://sitiooij.poder-judicial.go.cr/index.php/apertura/transparencia/estadisticas-policiales>



Nuestro filtro para la data es el siguiente:

* **Rango de fechas:** 1 Enero 2010 al 7 de Marzo 2024
* **Lugar del Suceso:** Todo el país (Provincias, Cantones, Distritos)
* **Categoría Delictiva:** Todas las categorías (Asalto, Hurto, Robo, Tacha, Robo de vehículo, Homicidio)
* **Victima Policial:** Todas las victimas (Edificación, Persona, Vehículo, Vivienda, otros)

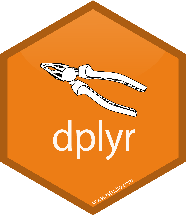
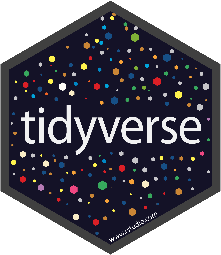
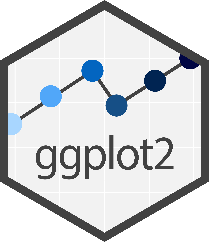
Una vez realizada la descarga obtenemos el archivo en formato csv, el cual dio como resultado **546594** resultados totales para nuestra base de datos.



## **Exploración y limpieza de los datos.**

Como primer paso, utilizaremos las librerías necesarias para el análisis en el programa RStudio.

En este caso, se emplearán las siguientes librerías:

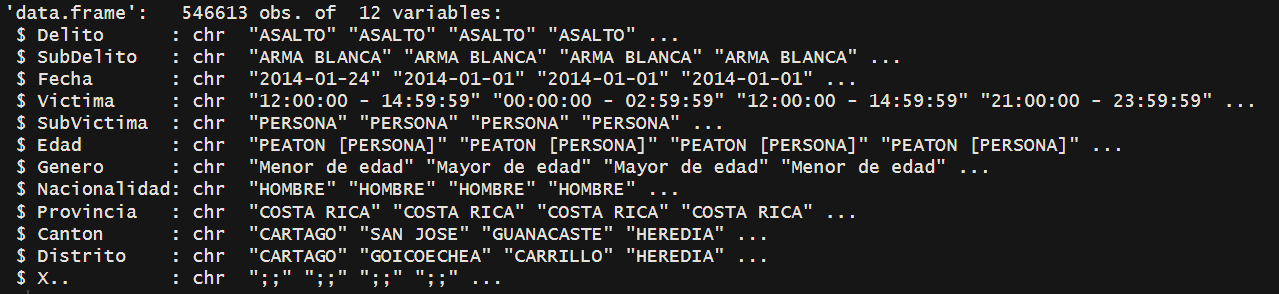
Una vez instaladas las librerías que necesitamos, se procede con la carga del Dataset por medio del siguiente código:



Se crea una variable que llamaremos Crimenes\_CR donde se guardará Dataset para facilitar el acceso a los datos. Posterior a esto se procede con la primera consulta que realizaremos y es obtener un resumen detallado de lo que contiene nuestro archivo csv:

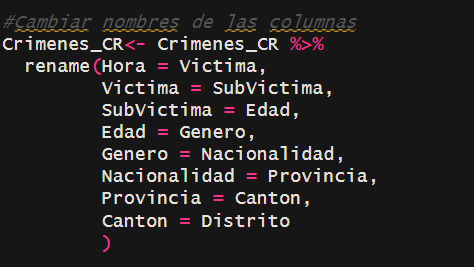


Y nos da como resultado:

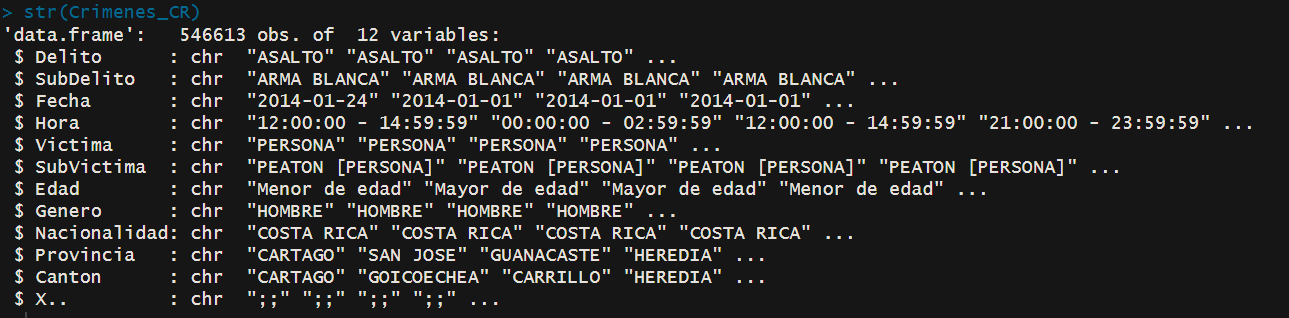


Se puede observar que el dataset **Crimenes\_CR** está conformado por 12 variables (columnas) y 546.613 mil observaciones (filas) con sus respectivos tipos de datos para cada uno de ellos.

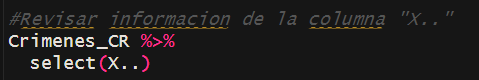
El primer problema identificado en la información es que la columna "Victima" contiene valores de tiempo. Esto se debe a que la columna "Fecha" en el archivo CSV está compuesta por dos valores: la fecha y la hora, separados por una coma (", "). Esto provoca que los valores se dividan y se desplacen a la siguiente columna. Por lo tanto, es necesario corregir este error en primer lugar por medio del siguiente código:



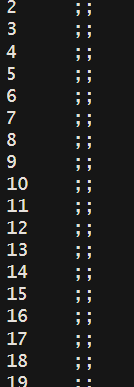
Donde se esta realizando el cambio de los nombres de las columnas luego confirmamos el cambio nuevamente con la función “str()”:



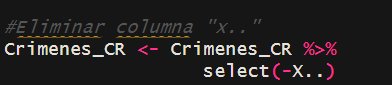
Se puede observar a la presencia de una columna llamada “X..” de tipo carácter que contiene datos que parecen ser irrelevantes, es necesario conocer los datos de dicha columna para observar si los valores son útiles para nuestro análisis. Para revisar los datos de la variable se realiza por medio de la siguiente línea de código

****

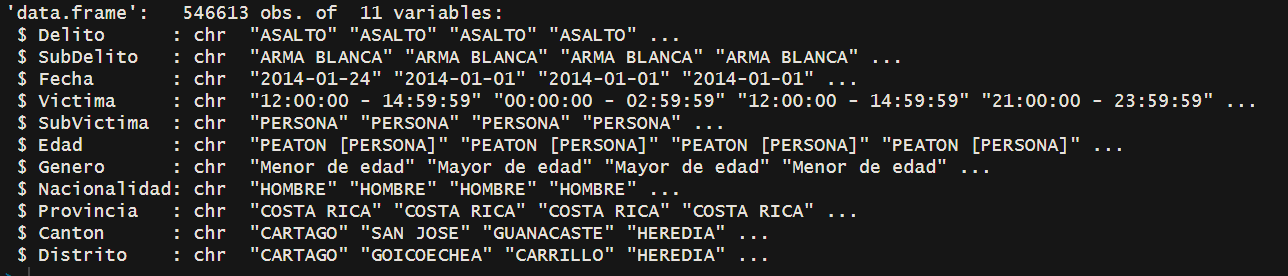
Se obtiene como resultado:

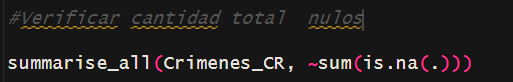


Dicha columna no es necesaria para nuestro enfoque, por lo que se dispone a eliminarla.

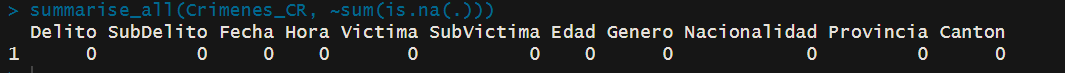


El código anterior, de igual manera utilizamos la variable **Crimenes\_CR** para guardar el resultado de la consulta donde seleccionamos el Dataset, donde la función **SELECT** selecciona la columna que deseamos, al agregarle el símbolo de **“ - “** al inicio elimina toda la columna. Se puede confirmar al aplicar de nuevo el comando de **“str “**al dataset:

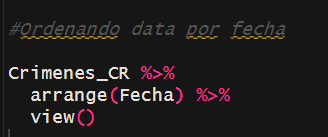


Es importante conocer si existen valores nulos cada columna:  


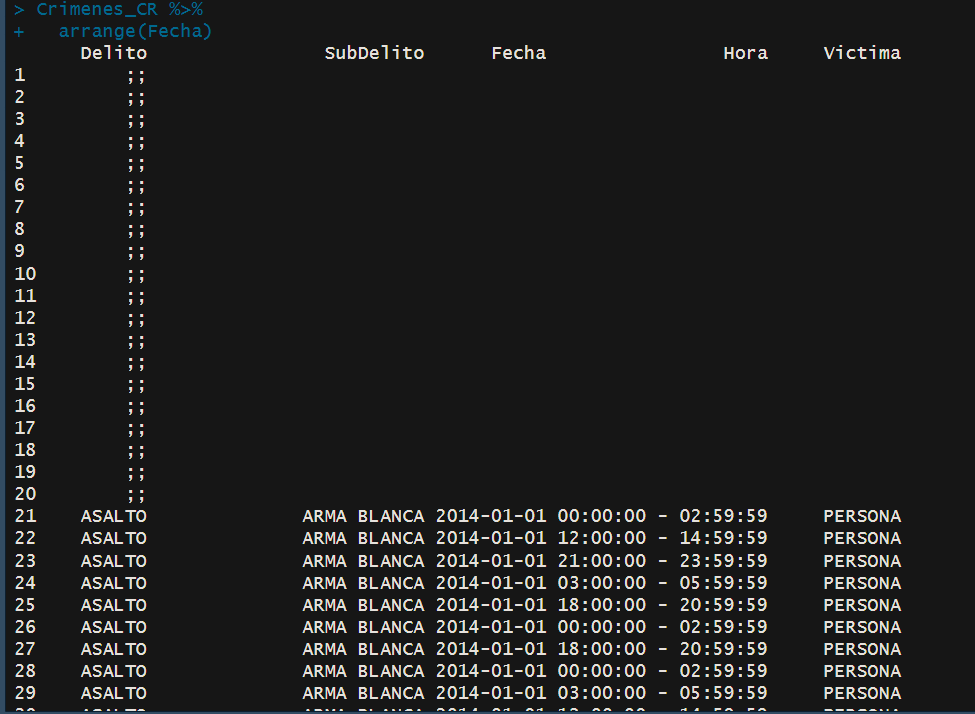
El código nos realiza una sumatoria total de cada columna de Crimenes\_CR donde cuyos valores que se encuentren sean nulos



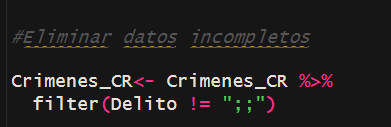
Se Confirma que no existen valores nulos en la información. Seguidamente procedemos a ordenar la data de forma ascendente por la variable de la fecha:

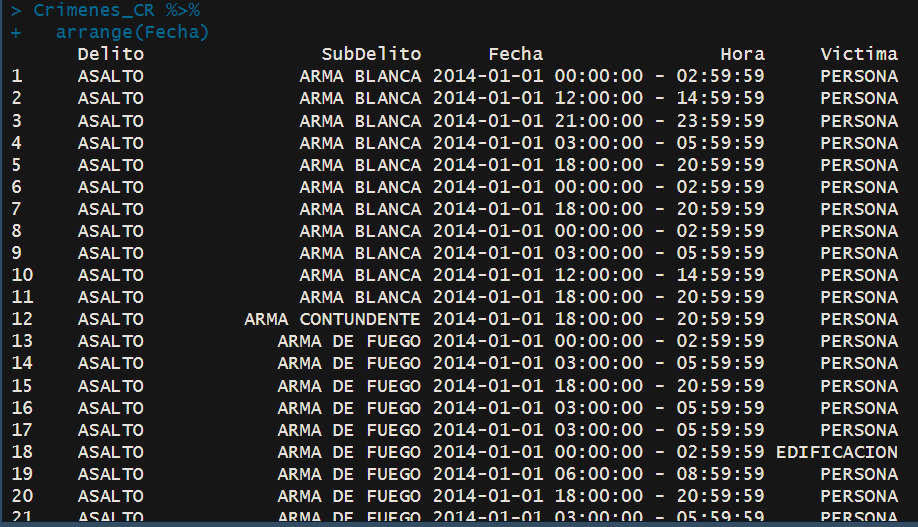


La consola nos muestra lo siguiente:

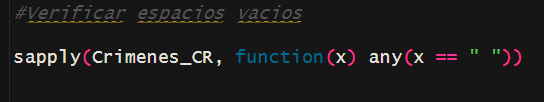


Se puede observar que en la columna “Delito” existen datos que no son útiles para nuestro análisis, por lo que se procede a eliminarlos:

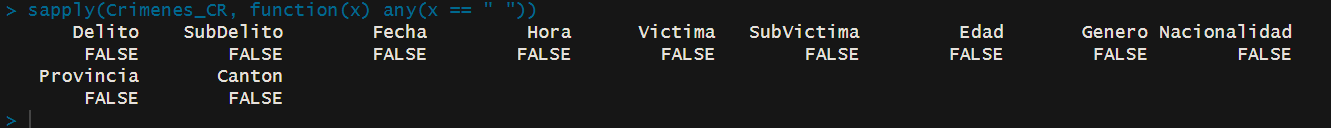


Realizamos a la confirmación nuevamente corriendo el código de ordenamiento por fecha:  


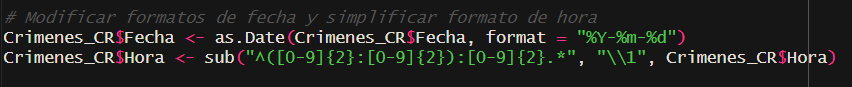
Se realiza un mapeo para verificar si existen valores en blanco:

****

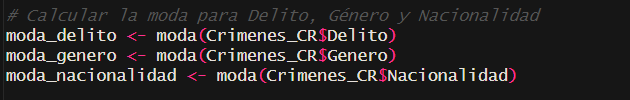
Se confirma que ninguna columna no tiene datos en blanco.

****

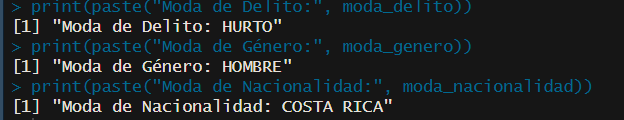
Se ambia el formato de la fecha y hora



Obtenemos la moda, media y mediana:

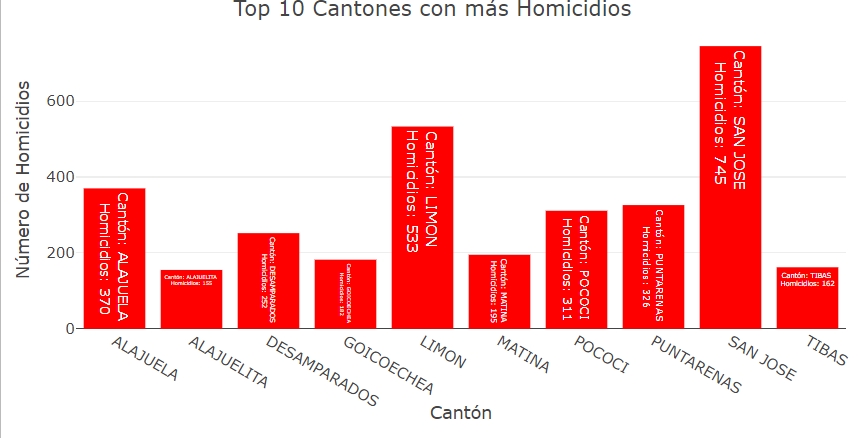


Lo que nos muestra la consola:



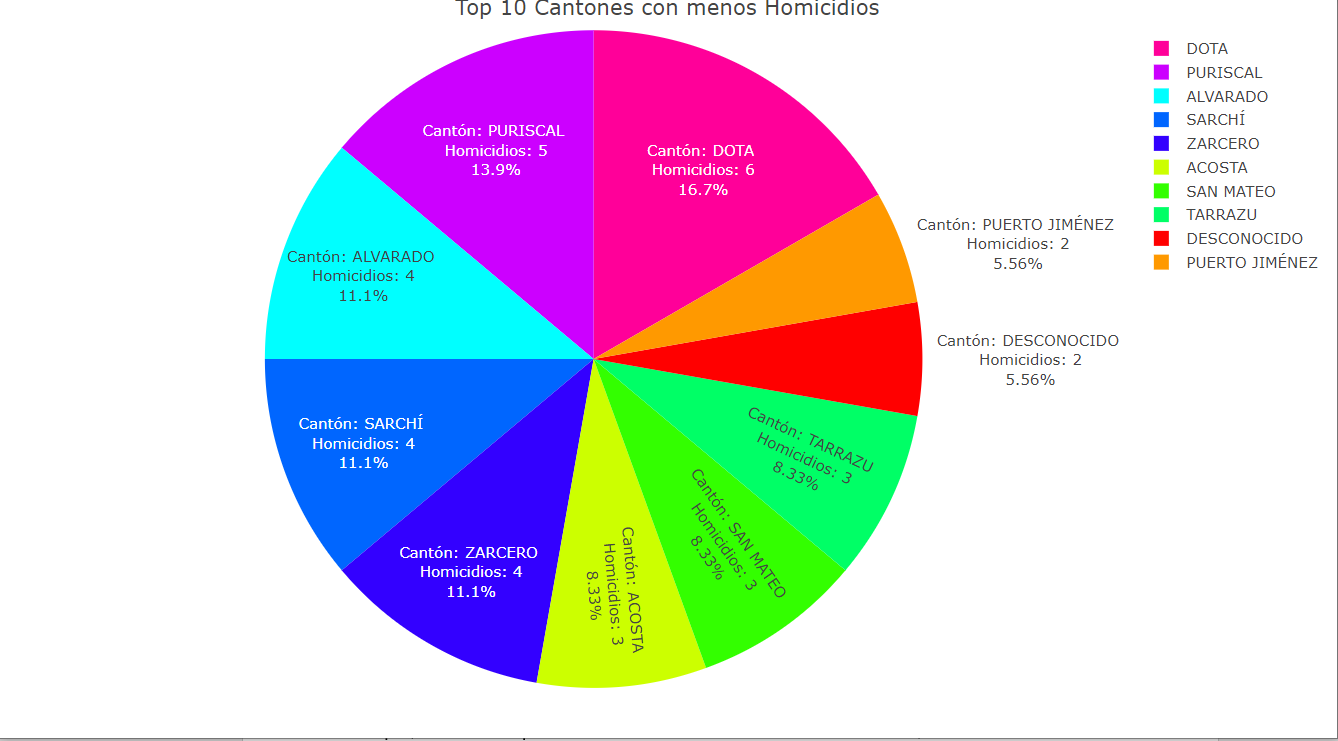
# **ANALISIS DE LOS DATOS.**

**Las áreas específicas en Costa Rica donde son más recurrentes los homicidios se identifican claramente.**



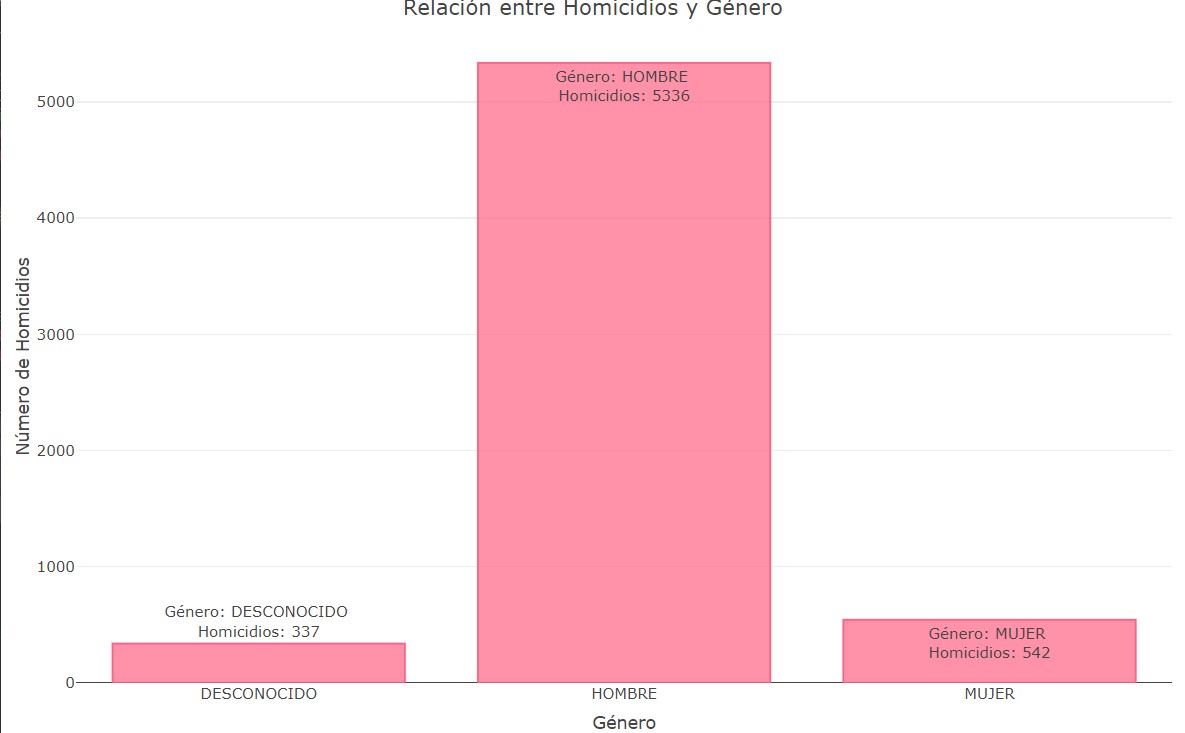
El grafico anterior nos muestra los 10 Cantones con mayor cantidad de homicidios en Costa Rica, podemos identificar al cantón de San José como uno de los que obtiene mayor cantidad de casos para un total de 600 homicidios en los últimos 10 años. Una de las mayores posibles causas de este numero tal alto se deba a la gran cantidad de personas que transitan día a día por la capital con el objetivo de trasladarse a sus trabajos u hogares o bien simplemente por recreación, lo que con lleva a la posibilidad de ser una posible víctima.

**Las áreas específicas en Costa Rica donde son menos recurrentes los homicidios se identifican claramente.**



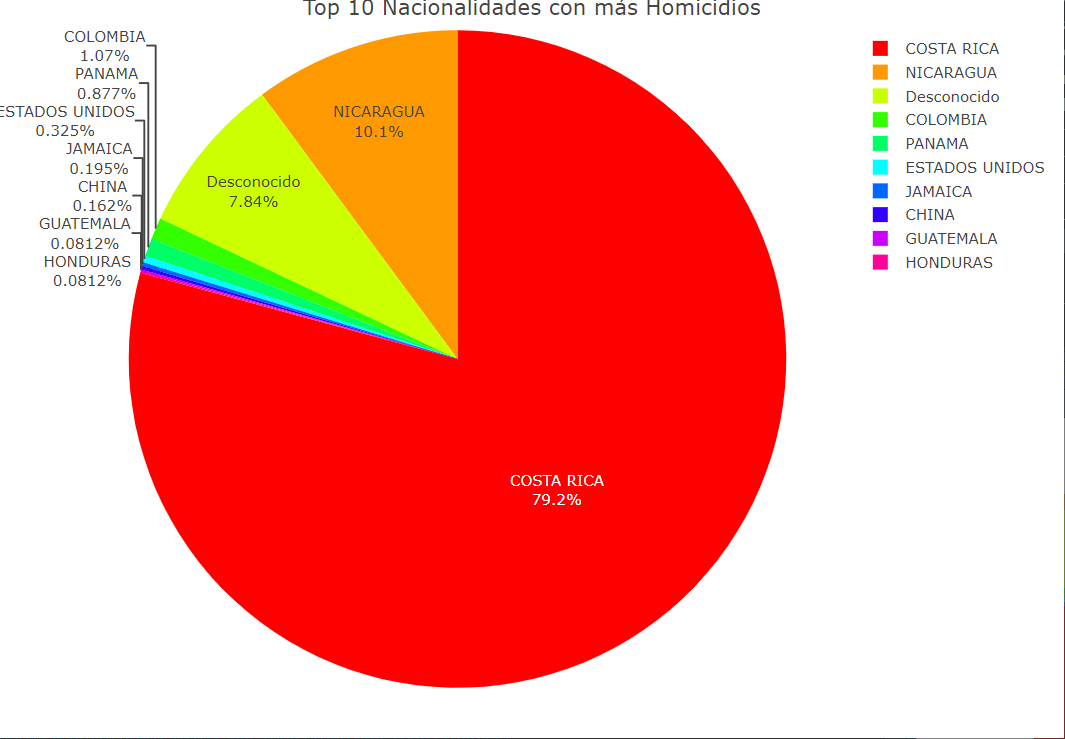
El grafico muestra el cantón de Puerto Jiménez como uno de los más seguros con tan solo 2 homicidios. Esto es un punto positivo para la provincia de Puntarenas ya que, al ser una opción con alto turismo nacional e internacional, beneficia mucho a los habitantes de esta provincia brindando mejores tasas de empleo.

**Existe una variación significativa en la proporción de homicidios entre diferentes géneros.**



Es importante conocer un poco de como se encuentran distribuidos los géneros de las personas que fueron victimas de homicidio, en este resultado se puede observar la gran diferencia entre hombres y mujeres, encabezando el resultado los hombres con un poco más de 5000 casos.

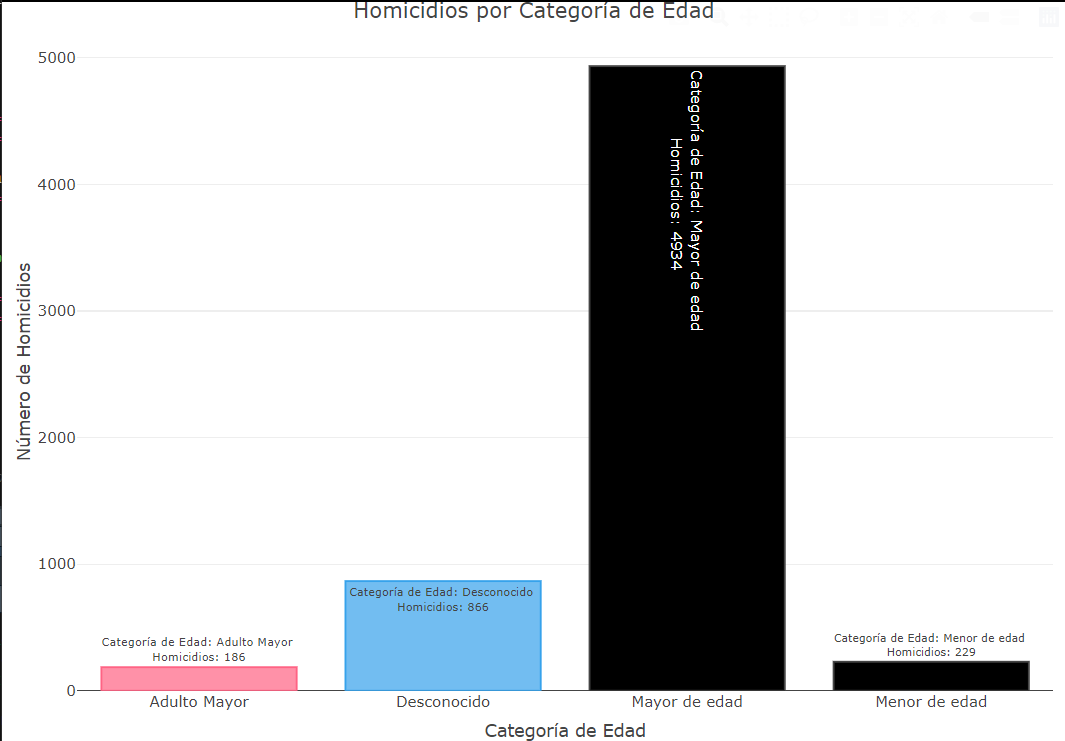
**Algunas nacionalidades presentan una tasa de homicidios significativamente mayor en comparación con otras.**



También es importante conocer cual es la nacionalidad de cada una de las victimas y por obvias razones podemos observar que Costa Rica es el país que encabeza la lista, seguido por Nicaragua país vecino por lo que es más recurrente la migración legal o ilegal de estas personas con el fin de obtener una mejor calidad de vida.

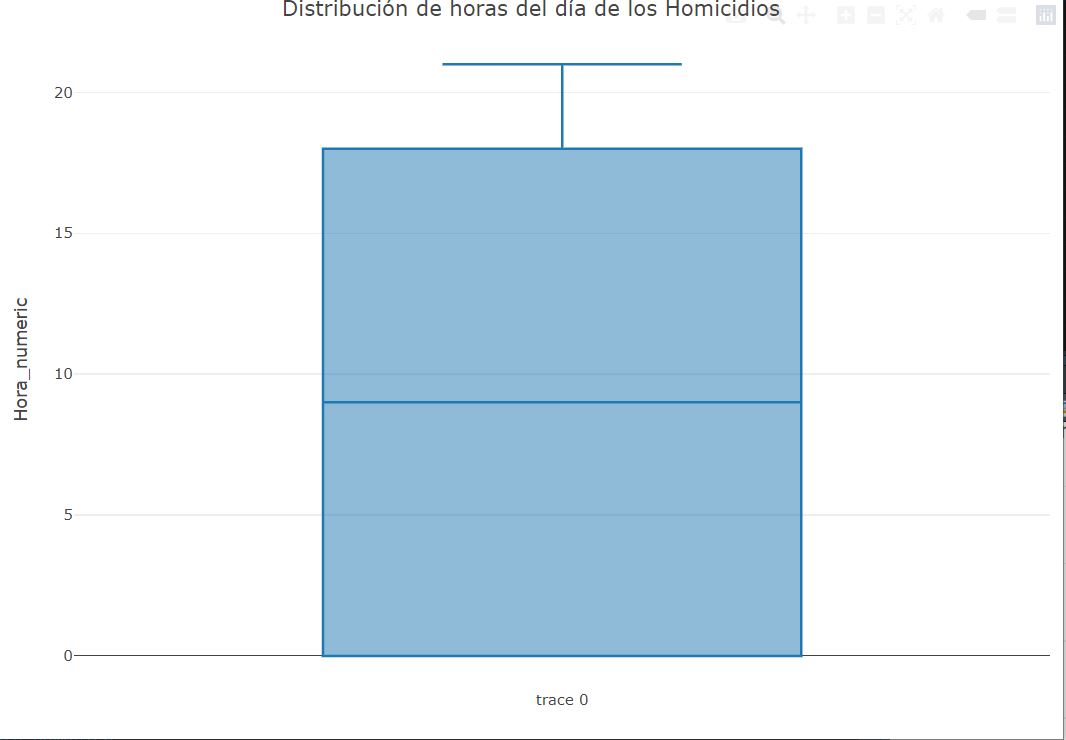
Es importante mencionar la nacionalidad con la tercera posición, la de origen **Desconocido**. Esto se debe a que las victimas posiblemente no se pudieron identificar luego pasado un tiempo para realizar las investigaciones para encontrar respuestas de la identificación de estas personas terminando como resultados negativos, se opta por catalogarlos dentro de este segmento.

**Hay una diferencia significativa en la tasa de homicidios entre menores de edad y mayores de edad.**



Se nota una gran diferencia de edades entre víctimas, mayores de edad encabezando con mucho la lista en comparación con la categoría de Menores de edad con una diferencia de un más de un 2000%.

Es posiblemente la mayoría de estos casos de menores de edad se deban a adolescentes pertenecientes a grupos delictivos que les abren los brazos ofreciéndoles una vida emocionante llena de dinero y lujos en un momento en la vida tarde o temprano, terminan siendo parte de las estadísticas.



En el grafico estilo Bloxpot se puede observar la distribución bastante equilibrada de los homicidios ocurridos en distintas horas alrededor del dia.

El análisis expresa que la mayor parte de los homicidios ocurren dentro de un rango horario relativamente amplio, desde las primeras horas de la mañana hasta las últimas horas de la noche.

# **CONCLUSION:**

En esta investigación acerca de la criminalidad en Costa Rica permitió comprender un poco más el nivel en el que nos encontramos actualmente como país a nivel seguridad.

A pesar de los desafíos para obtener, explorar, transformar y analizar la data, se fue capaz de superar los obstáculos y se logro obtener una data coherente y fácil de analizar. Estas transformaciones permitieron una visión profunda sobre la información que se esta investigando y encontrar patrones interesantes dentro de los filtros realizados al Dataset con el fin obtener respuestas a las hipótesis planteadas.

El análisis revela patrones que se encontraron como los serían los lugares con mayor tasa de criminalidad convirtiéndolos en zonas de alto riesgo para la ciudadanía. Este análisis podría permitir una mejora en parte de seguridad en dichos lugares con el fin de mitigar un poco los resultados y buscar disminuir las víctimas.

Esta investigación confirma la importancia de realizar un análisis de datos detallado ya que es una herramienta fundamental para obtener resultados más acordes a las hipótesis planteadas.

# **BIBLIOGRAFÍA:**

Poder Judicial de Costa Rica. (s.f.). Estadísticas policiales. Recuperado de <https://sitiooij.poder-judicial.go.cr/index.php/apertura/transparencia/estadisticas-policiales>