

**PROYECTO DE GRADO**

**APLICACIÓN DE LAS HERRAMIENTAS BIG DATA FRENTE A LA INSEGURIDAD EN LA CIUDAD DE BOGOTÁ**

**JENNIFER RODRIGUEZ VERA - 827822**

**KEVIN NICOLAS FLOREZ PINZON - 116382**

**YULI ANDREA PINEDA CASTILLO - 847090**

**CORPORACIÓN UNIVERSITARIA MINUTO DE DIOS**

**ESPECIALIZACIÓN TECNOLÓGICA EN HERRAMIENTAS PARA LA GESTIÓN DE DATOS – BIG DATA**

**NOVIEMBRE 2022**

**TABLA DE CONTENIDO**

1. **INTRODUCCIÓN**
2. **JUSTIFICACIÓN**
3. **OBJETIVOS**
   1. **Objetivo General**
   2. **Objetivos Específicos**
4. **PLANTEAMIENTO DEL PROBLEMA**
   1. **Antecedentes del Problema**
   2. **Formulación del Problema**
   3. **Alcances y Delimitaciones**
5. **MARCO DE REFERENCIA**
   1. **Marco Conceptual**
   2. **Marco Teórico**
   3. **Estado del arte**
6. **METODOLOGIA**
7. **ANALISIS Y RESULTADOS**
8. **CONCLUSIONES**
9. **BIBLIOGRAFIA**

**AGRADECIMIENTOS**

En la elaboración de este proyecto agradecemos a los docentes que nos acompañaron y apoyaron durante el proceso de formación en la Especializacion en Herramientas Tecnologicas para la gestión de datos – Big Data, permitiéndonos adquirir grandes conocimientos para un mejor desempeño en nuestras vidas profesionales y personales.

1. **INTRODUCCIÓN**
2. **JUSTIFICACIÓN**

La inseguridad en la ciudad de Bogotá es un problema que no ha disminuido a través de los años siendo una problemática que afecta la calidad de vida de la población, cada día llegan miles de denuncias a las autoridades competentes, las cuales en su gran mayoría se encuentran sin resolver. Es por esto que se desarrollan nuevas estrategias para mitigar y encontrar soluciones a través de los decretos establecidos por el Gobierno Nacional, entre los más importantes se destacan el Decreto 2268 expedido en diciembre de 2017, con el cual se restringió el porte de armas de fuego. En octubre de 2018, el Distrito expidió el Decreto 599 que restringe la compra y venta de armas blancas, algo que ya consagraban el Código de Policía de Bogotá de 2003 y el Acuerdo 517 de 2012, con el fin de dar orden ciudadano. A partir de esta necesidad se pretende aplicar la evolución tecnológica con las herramientas de análisis de datos – Big Data por medio de la información recolectada en las bases de datos de las entidades estatales donde destacan las denuncias, capturas y cantidad de delitos ejecutados en un periodo de tiempo.

(https://www.personeriabogota.gov.co/sala-de-prensa/notas-de-prensa/item/470-alarmante-situacion-de-inseguridad-en-bogota)

Para minimizar la problemática de inseguridad en la ciudad de Bogotá se busca emplear herramientas de análisis de datos para generar valor a la información suministrada por la Secretaria de Seguridad Convivencia y Justicia, con el fin de obtener estrategias optimas que permitan reducir los índices de delitos y mejorar la perspectiva ciudadana.

1. **OBJETIVOS**
   1. **Objetivo General**

Implementar las herramientas del análisis de datos Big Data con el fin de determinar el impacto de la inseguridad en la ciudad de Bogotá antes, durante y después de la pandemia, a fin de generar valor de la información en la toma de decisiones para las autoridades competentes.

* 1. **Objetivos Específicos**
* Analizar las diferentes fuentes de datos que permita realizar procesos de ETL (Extracción, Transformación y Carga) para el uso correcto de la información.
* Desarrollar un análisis estadístico utilizando algoritmos de predicción y correlación con el fin focalizar los puntos críticos de inseguridad a futuro en la ciudad de Bogotá.
* Utilizar el modelo de regresión lineal que permita trabajar con las líneas de tiempo, para determinar la frecuencia de delitos en las diferentes localidades de la ciudad de Bogotá.
* Integrar la metodología CRISP DM durante las diferentes fases del desarrollo de la investigación.

1. **PLANTEAMIENTO DEL PROBLEMA**

Actualmente muchos colombianos se preguntan ¿Qué hacemos con la inseguridad que inunda el país? esto ante las denuncias de robos y atracos que cada día ocurren en las ciudades principales, donde quedan los videos de cámaras de seguridad que muestran las acciones de la delincuencia, pero sobre todo en la ciudad de Bogotá donde los niveles de violencia a los que han llegado algunos delincuentes han impactado gravemente la ciudadanía, primero atacan, golpean, hieren o incluso matan a las víctimas y posteriormente las roban.

De acuerdo con las estrategias que utilizan las entidades que brindan seguridad a la ciudadanía se encuentra la Policía Metropolitana de Bogotá que ha ofrecido montos de hasta 20 millones de pesos por información que permita la identificación y consecutivamente judicialización de estos delincuentes, sin embargo, se ha evidenciado que las familias más afectadas reclaman una intervención más rigurosa por parte de las autoridades para que estos hechos no sigan ocurriendo. Las cifras son alarmantes, ya que solamente en los primeros tres meses del año se registraron 302 hurtos confirmados por día. Lo que serían 3.915 casos más denunciados de los que se habían presentado durante enero, febrero y marzo del año 2021. (Revista Semana, 2022)

¿Por qué pareciera estar aumentando la inseguridad en la ciudad de Bogotá?

Hoy la ciudadanía por medio de entrevistas informales ha manifestado que su percepción de la inseguridad ha pasado a ser un hecho real, no hay ninguna ciudad en la que el homicidio no haya aumentado, en promedio el homicidio creció el 37%, el hurto a celulares el 17% y el robo de bicicletas el 3% en el año 2021, de acuerdo con estudios sobre esta problemática hay varios factores principales que están contribuyendo a ese crecimiento de inseguridad, entre ellos la pandemia, un año donde las bandas criminales también estuvieron encerradas porque la actividad comercial cerro sus puertas, las personas empezaron a resguardarse en sus hogares y victimas para atacar no se veían en las calles de la ciudad de Bogotá, esto provoco un reajuste de estructuras criminales y adicional fenómenos como el microtráfico, hurto callejero, extorsión al comercio, no solo se reactivó la economía sino también la inseguridad y con mayor potencia. (Ernesto Cortes, Editor General de El Tiempo, 2021)

Es importante aclarar que la Policía Nacional realiza constantes capturas todos los días en el país, sin embargo, aparentemente esto no es suficiente, la realidad deja entender que los criminales saben que los atrapan, duran capturados aproximadamente 36 horas y quedan en libertad, cabe resaltar que la responsabilidad no solo recae sobre la Policía Nacional, esto es una cadena, si la ciudadanía exige a estos resultados, capturas, retenes y detenciones, metiendo presos a los delincuentes, es importante ver que esto requiere una infraestructura mínima, un aparato de justicia que funcione, que no deje los delincuentes en la calle de nuevo, que realice un reconocimiento efectivo de las bandas criminales, pero para esto adicional se necesitan los Comandos de Atención Inmediata (CAI), estaciones de policía, ampliar la infraestructura de las cárceles para controlar el hacinamiento, solamente de población extranjera se dice que en la ciudad de Bogotá en el año 2021 han sido capturados 2500 personas cometiendo actos delictivos, casos donde la justicia tipifica que es un delito menor, porque la persona no está bien identificada, porque no se llevó a cabo el procedimiento correcto permitiendo que vuelvan a quedar libres, lo que lleva a la reincidencia que actualmente no se está castigando, finalmente mientras no haya un aparato de justicia eficaz la Policía Nacional seguirá sintiendo un desgaste operativo al capturar siempre a las mismas personas por los mismos delitos.

¿Qué está fallando en la estrategia de seguridad desde la alcaldía y la Policía Metropolitana de Bogotá? Se evidencian varios factores, comenzando por el nivel de violencia que se esta viendo en las calles, parece un desprecio por la vida donde lo importante es robar.

¿Por qué ese nivel de violencia? Esto se debe a la falta de confianza entre la Policía y la ciudadanía que ha venido creciendo abismalmente, esto es algo que no puede pasar en la comunidad ya que las personas dejan de denunciar los casos, es necesario adaptar la inteligencia social y tecnológica a la nueva realidad frente a esta problemática. Actualmente hay bandas sofisticadas que roban desde las propias cárceles, hay bandas armadas, bandas especializadas en cometer delitos, bandas con armas de fuego, bandas con armas readaptadas y bandas con armas de fogueo que no se pueden controlar, las armas de fuego han incrementado el número de homicidios, este fenómeno del sicariato creció el 9% en Bogotá durante el año 2021, también se ha evidenciado el aumento de las riñas por intolerancia social y las riñas por ajuste de cuentas entre bandas, queriendo apoderarse de territorio de otras bandas donde finalmente se desatan guerras en la ciudad. Es indispensable generar estrategias que permitan proteger a la comunidad en su propio sitio, muchas veces la policía se concentra en grandes puntos donde hay focos de inseguridad, pero no contemplan que hay inseguridad en el barrio, en la puerta de la casa, cuando comparte con la familia, fundamental conocer cómo se puede mitigar esta realidad empleando un trabajo de inteligencia tecnológica.

En Medellín se ha comprobado que ha funcionado la implementación tecnológica por que han permitido evitar los hurtos y el robo de las motos, ya que en el caso que se llegue a presentar esta situación el afectado se comunica a la línea 123 y de inmediato por medio de un circuito de cámaras de seguridad evitan que el hurto se materialice. (Santiago Olivares, Periodista Judicial del Diario Q’ Hubo, 2021)

Finalmente ¿Que se podría hacer con estas bandas criminales que parecen estar mas organizadas y mejor armadas?

En el año 2022 la alcaldía de Bogotá propuso incrementar el número de Policías para el año 2023, lo que permitiría que la percepción de la ciudadanía “mejore” entre la gente vea más policías en la calle aparentemente va a sentir algo más de tranquilidad, sin embargo, esto no combate de raíz el problema, desde que se permitió que los motociclistas dejaran de usar chalecos y posteriormente cascos con las placas de la moto se ha dificultado el reconocimiento por cámaras de seguridad de estos delincuentes que en su momento estaban cometiendo algún acto ilícito. Hay que revisar cual es el problema de fondo y generar oportunidades ya que un delincuente prefiere pasar dos años de su vida en una cárcel por un robo que le genero más ingresos que emplearse en el mundo laboral que posiblemente solo le genere un S.M.L.V., adicional hay que ser mas rigurosos con la corrupción que se ve en los entes que deberían prestar la seguridad a la comunidad ya que últimamente se han visto casos donde se involucran a estas bandas delincuenciales.

¿Qué podría hacer un ciudadano desde su posición? Inicialmente es importante que se establezcan las denuncias en la Policía Nacional para poder llevar datos que permitan su análisis y por medio de estos mejorar la toma de decisiones al momento de aplicar estrategias, segundo es claro que por razones económicas algunos ciudadanos compran cosas robadas como celulares, repuestos de vehículos, bicicletas, entre otros, en lugares donde ofrecen mejores precios pero que su procedencia es de origen ilícito sin tener en cuenta que este objeto pudo arrebatarle la vida a una persona, lo que finalmente contribuye a mantener esa cadena criminal que debería romperse al dejar de comprar estos artículos robados.

* 1. **Antecedentes del Problema**
  2. **Formulación del Problema**

¿Cómo las herramientas de análisis de datos pueden apoyar la toma de decisiones para generar estrategias que minimicen los delitos en la ciudad de Bogotá?

* 1. **Alcances y Delimitaciones**

1. **MARCO DE REFERENCIA**
   1. **Marco Conceptual**
   2. **Marco Teórico**
   3. **Estado del arte**
2. **METODOLOGIA**

Conforme al objeto de estudio, se requiere realizar un análisis de la data obtenida por medio de la fuente de Datos Abiertos dispuestos por el Gobierno Nacional de Colombia, con el fin de generar estudios estadísticos por medio de diferentes variables que aporten un valor agregado a la recolección de datos.

Para lograr esta observación se utiliza la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), la cual facilita una representación del ciclo de vida de minería de datos.

La anterior metodología contiene las siguientes fases: comprensión del negocio, estudio y comprensión de los datos, análisis de datos y selección de las características, modelado, evaluación y despliegue, usando la herramienta tecnológica Anaconda Navigator, el framework notebook de Jupyter y el lenguaje de programación Python en la versión 2.7.18.

**FASES METODOLOGIA CRISP-DM**

1. **Fase 1: Comprensión del negocio**

El cometido general de esta fase es entender los objetivos y requerimientos del proyecto desde una perspectiva de negocio y convertir este conocimiento en la definición de un problema de minería de datos y plan preliminar para alcanzar los objetivos.

Como efecto se establece el objeto de estudio para caracterizar los focos de inseguridad que se presentan en las diferentes localidades de la ciudad de Bogotá, en un rango de tiempo de los últimos 5 años (prepandemia, pandemia y postpandemia), con el fin de identificar el aumento de la inseguridad y definir los patrones de comportamiento en un tiempo presente y futuro, de acuerdo con la agrupación de variables como se determinan en las siguientes tablas:

|  |
| --- |
| Características |
| 1. Datos 2018 |
| 2. Datos 2019 |
| 3. Datos 2020 |
| 4. Datos 2021 |
| 5. Datos 2022 |
| 6. Coordenadas Localidad |
| 7. Recuperación Vehículos Policía Nacional |
| 8. Reporte Capturas Policía Nacional |

Las características correspondientes a la tabla anterior se dividen en diferentes tipos de variables de la siguiente manera:

|  |  |
| --- | --- |
| Característica 1. | Variables |
|  | Año |
| Nro del Mes |
| Mes |
| Localidad |
| UPZ |
| Delito |
| Sexo |
| Nombre Dia |
| Rango del Dia |
| Arma Empleada |
| Numero Hechos |

|  |  |
| --- | --- |
| Característica 2. | Variables |
|  | Año |
| Nro del Mes |
| Mes |
| Localidad |
| UPZ |
| Delito |
| Sexo |
| Nombre Dia |
| Rango del Dia |
| Arma Empleada |
| Numero Hechos |

|  |  |
| --- | --- |
| Característica 3. | Variables |
|  | Año |
| Nro del Mes |
| Mes |
| Localidad |
| UPZ |
| Delito |
| Sexo |
| Nombre Dia |
| Rango del Dia |
| Arma Empleada |
| Numero Hechos |

|  |  |
| --- | --- |
| Característica 4. | Variables |
|  | Año |
| Nro del Mes |
| Mes |
| Localidad |
| UPZ |
| Delito |
| Sexo |
| Nombre Dia |
| Rango del Dia |
| Arma Empleada |
| Numero Hechos |

|  |  |
| --- | --- |
| Característica 5. | Variables |
|  | Año |
| Nro del Mes |
| Mes |
| Localidad |
| UPZ |
| Delito |
| Sexo |
| Nombre Dia |
| Rango del Dia |
| Arma Empleada |
| Numero Hechos |

|  |  |
| --- | --- |
| Característica 6. | Variables |
|  | Localidad |
| Longitud |
| Latitud |
| Código |
| Gp |

|  |  |
| --- | --- |
| Característica 7. | Variables |
|  | Departamento |
| Municipio |
| Código DANE |
| Clase bien |
| Fecha Hecho |
| Cantidad |

|  |  |
| --- | --- |
| Característica 8. | Variables |
|  | Departamento |
| Municipio |
| Código DANE |
| Fecha Hecho |
| Genero |
| Descripción Conducta Captura |
| Grupo Etario |
| Cantidad |

1. **Fase 2: Estudio y comprensión de los datos**

Se limpian los datos (eliminación de datos nulos y valores constantes e integración de los datos), se generan atributos adicionales a partir de los existentes por ganancia de información, se realizan transformaciones o cambios de formato a los valores de los atributos que se consideran necesarios, se eliminan los atributos reemplazados, así como los registros.

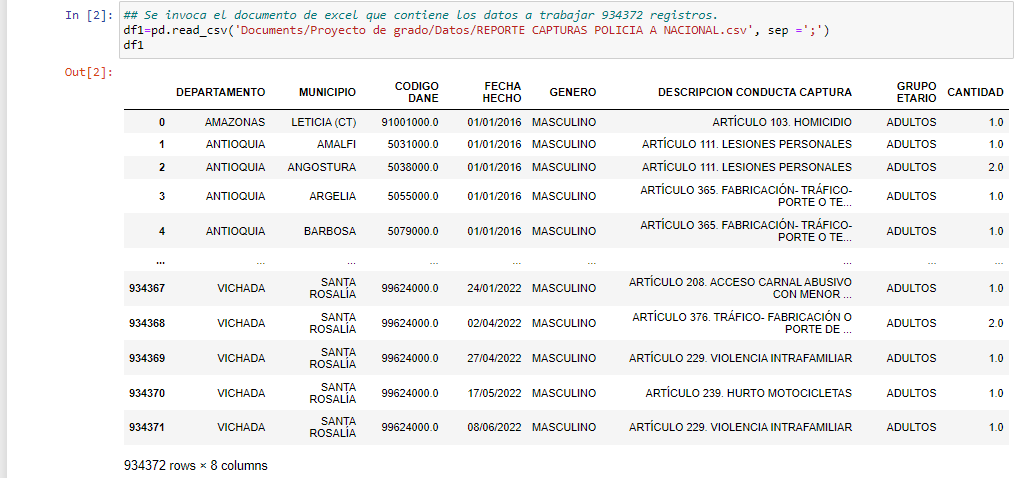
Se toma como referencia las fuentes de información de la página Datos Abiertos, con el fin de poder obtener y analizar la data que se tendrá en cuenta en la problemática a resolver de este proyecto. Como primera instancia se utiliza la herramienta de Python para iniciar la limpieza de datos por medio de un notebook.

**REPORTE CAPTURAS POLICIA A NACIONAL**

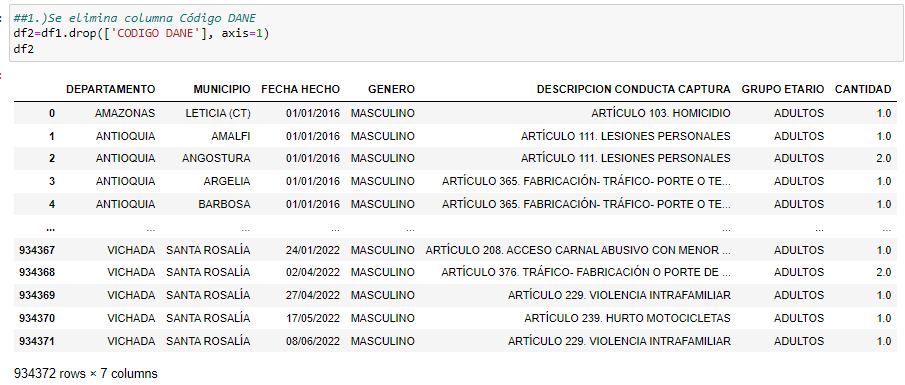
Se realiza la importación de las librerías Pandas y Numpy las cuales se tendrán presentes en el desarrollo de esta fase.



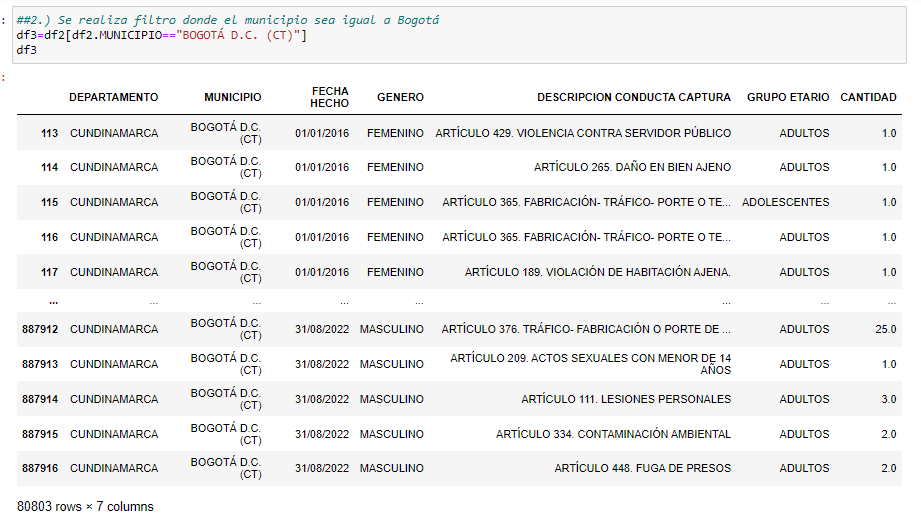
Se procede con la importación del archivo “REPORTE CAPTURAS POLICIA A NACIONAL.csv” para identificar el numero de registros y columnas de datos a trabajar.



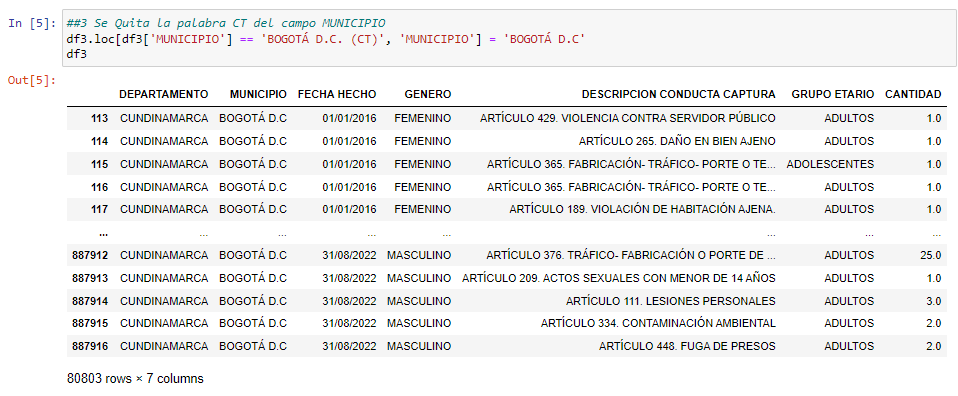
Respecto a la informacion anteriormente importada se inicia con la validacion del contenido de cada columna y se evidencia que para esta data existe la columna “CODIGO DANE”, se determina que dicha columna no sera tenida en cuenta para el analisis que requiere este proyecto, por lo cual se procede a eliminar.



Sabiendo que el alcance de este proyecto se enfoca únicamente en los delitos registrados que se han generado en la ciudad de Bogotá, se realiza un filtro por medio del campo municipio, donde su valor es igual a “BOGOTÁ D.C. (CT)”.



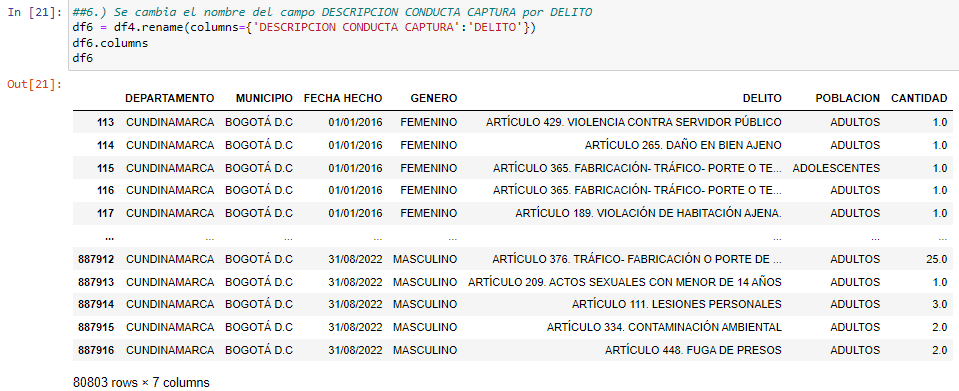
Para continuar con la limpieza de la información se inicia una homologación respecto al campo “MUNICIPIO” teniendo en cuenta que el valor actual es “BOGOTÁ D.C. (CT) “el anterior valor se reemplaza por “BOGOTÁ D.C” con el fin de obtener una data mas clara.



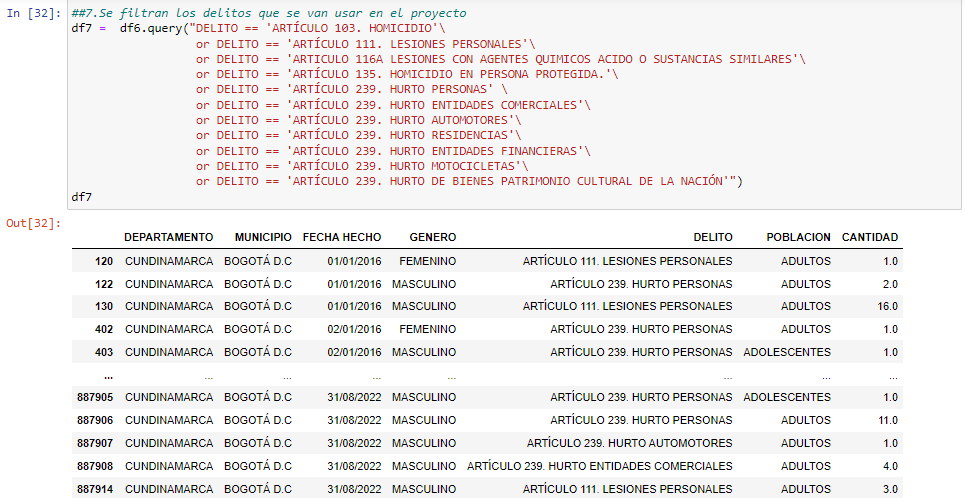
Continuando con la revisión de columnas se identifica que existe una columna cuyo nombre es “GRUPO ETARIO” la cual es reemplazada por el nombre de “POBLACION” de acuerdo a la información registrada.



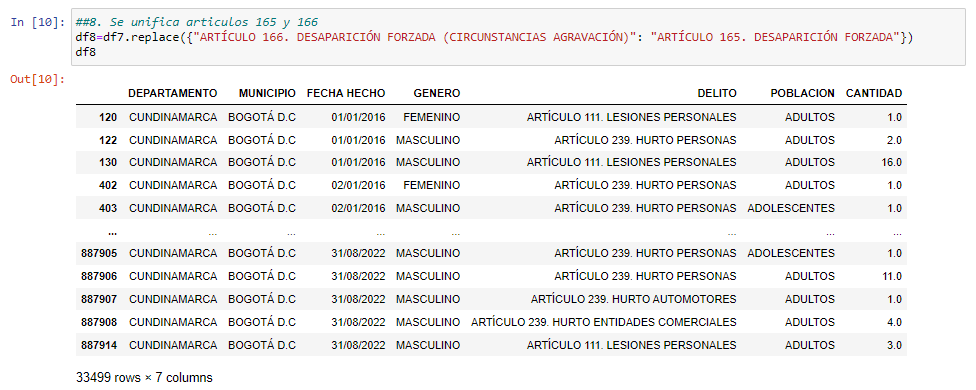
Validando las columnas del archivo se identifica una columna que se llama “DESCRIPCION CONDUCTA CAPTURA” la cual contiene un nombre extenso y se realiza el nombramiento con la palabra “DELITO” esto con el fin de obtener variables concisas y realizar manipulación de la data de una forma más factible.



Una vez realizada la limpieza anterior se filtra por los artículos que serán tenidos en cuenta para poder realizar el análisis y asi mismo hallar los casos que serán manejados en cuanto a los delitos en la ciudad de Bogotá.

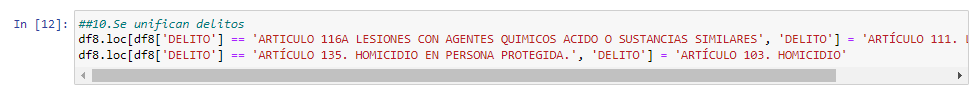


Se toma el “ARTÍCULO 166. DESAPARICIÓN FORZADA (CIRCUNSTANCIAS AGRAVACIÓN)” el cual es homologado por el articulo “ARTÍCULO 165. DESAPARICIÓN FORZADA” con el fin de manejar una sola variable.

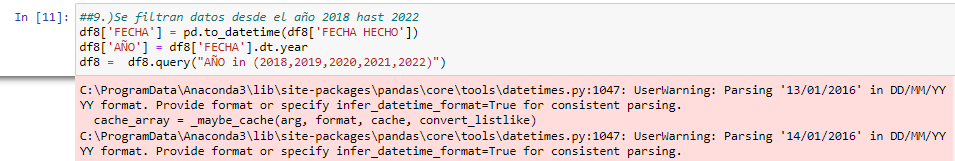


Se homologa “ARTICULO 116A LESIONES CON AGENTES QUIMICOS ACIDO O SUSTANCIAS SIMILARES” por el “ARTÍCULO 111. LESIONES PERSONALES”

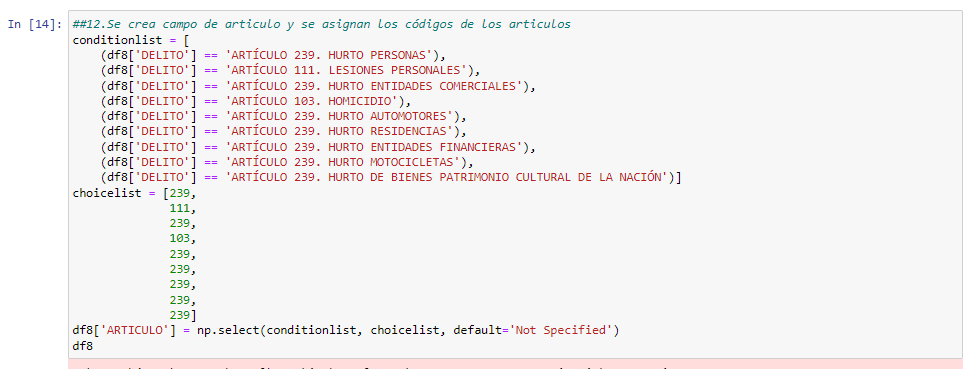
Se homologa “ARTÍCULO 135. HOMICIDIO EN PERSONA PROTEGIDA” por el “ARTÍCULO 103. HOMICIDIO”



De acuerdo a las líneas de tiempo que se tomaran para realizar el análisis de delitos y capturas se toma como base un rango de años desde el 2018 hasta el 2022.

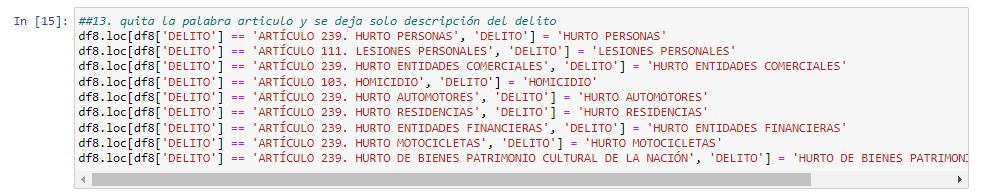


Se evidencia que en el campo del “DELITO” se encuentra la descripción y el código del articulo de forma concatenada por lo cual se inicia la extracción del código del articulo.

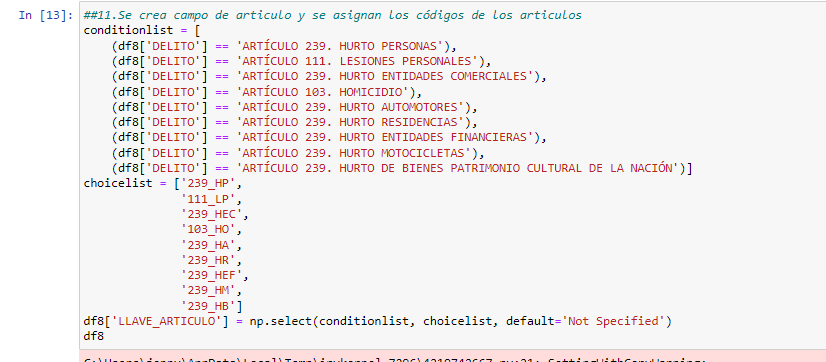




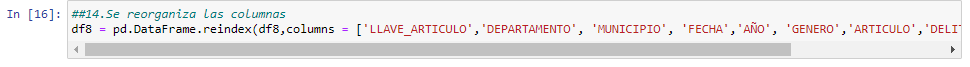
Se quita la palabra “ARTICULO” del campo “DELITO” y se deja únicamente la descripción con el fin de minimizar y generar una información más detallada.



Se crea una nueva columna cuyo nombre será “LLAVE\_ARTICULO” el cual sirve para la identificación de cada registro de esta base y podrá utilizarse para futuros cruces con otras fuentes de información que serán utilizadas para el análisis.



Se reorganiza el orden de las columnas con el fin de obtener una estructura final.

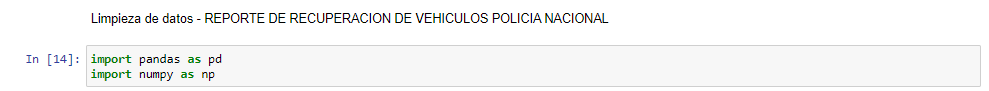


Como ultimo paso de esta limpieza del archivo se inicia una exportación de los datos y asi definir la información que se trabajara en fases posteriores.



**RECUPERACION VEHICULOS POLICIA NACIONAL**

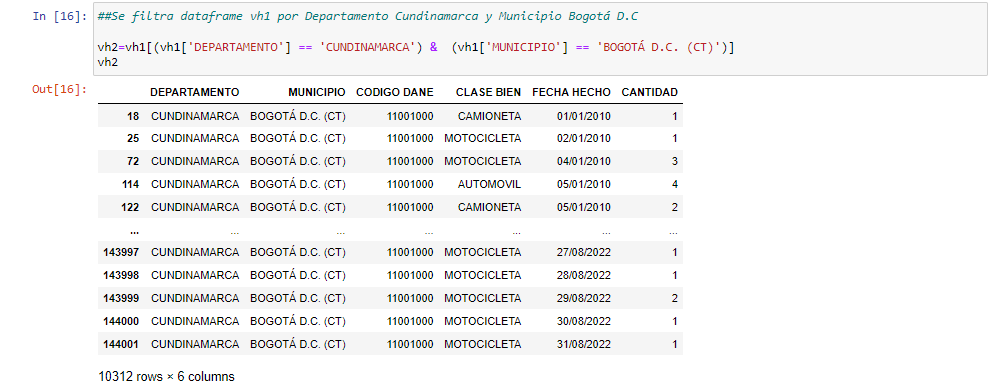
Se realiza la importación de las librerías Pandas y Numpy las cuales se tendrán presentes en el desarrollo de esta fase.

****

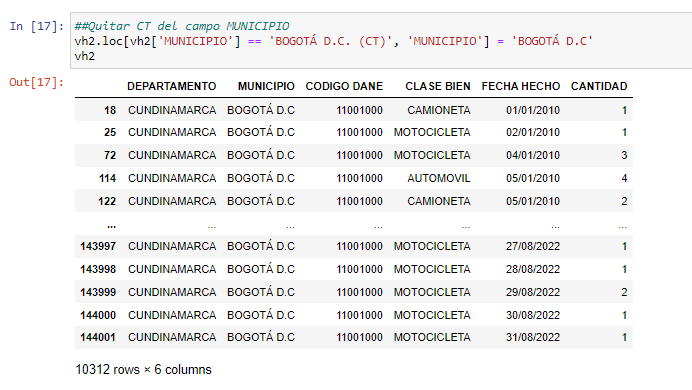
Se realiza la importación del archivo “RECUPERACION VEHICULOS POLICIA NACIONAL.csv” con el propósito de iniciar la limpieza de datos y definir la información a trabajar.

****

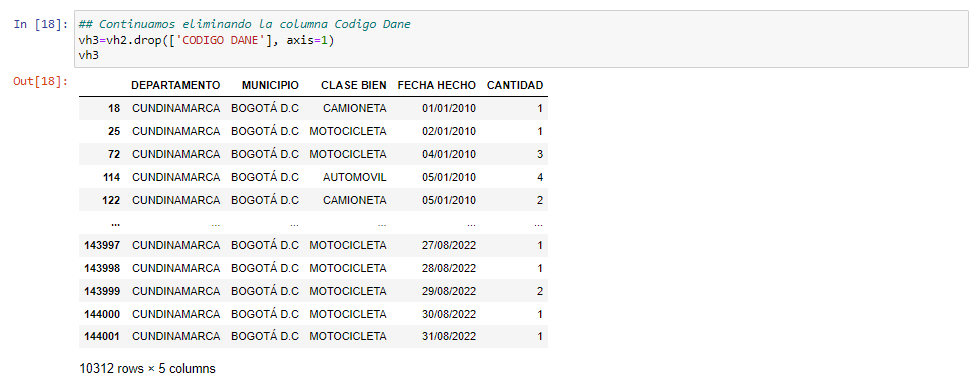
Se realiza el filtro de los campos “DEPARTAMENTO” y “MUNICIPIO” con el objetivo de tener la recuperación de vehículos pertenecientes a la ciudad de “BOGOTÁ D.C. (CT)”

****

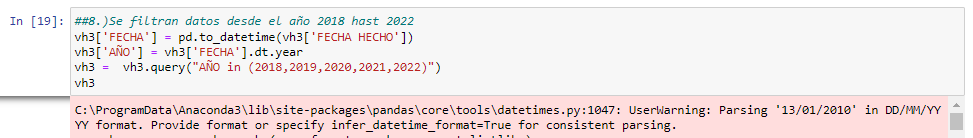
Para continuar con la limpieza de la información se inicia una homologación respecto al campo “MUNICIPIO” teniendo en cuenta que el valor actual es “BOGOTÁ D.C. (CT) “el anterior valor se reemplaza por “BOGOTÁ D.C” con el fin de obtener una data mas clara.

****

Validando la información de cada columna se encuentra el campo de “CODIGO DANE”, el cual se procede a eliminar de la base debido a que es una variable que no se utilizara para el enfoque de la manipulación de datos.

****

Se toman los registros reportados entre el año 2018 al 2022 con el motivo de obtener la línea de tiempo y las fases que serán empleadas en el proyecto.

****

Se crea una columna cuyo nombre será “LLAVE\_ARTICULO” para poder identificar el delito que se está manejando.

****

Finalizando la limpieza de datos de esta base se realiza una exportación de información para tener presente las variables definitivas que serán manipuladas en fases posteriores.

****

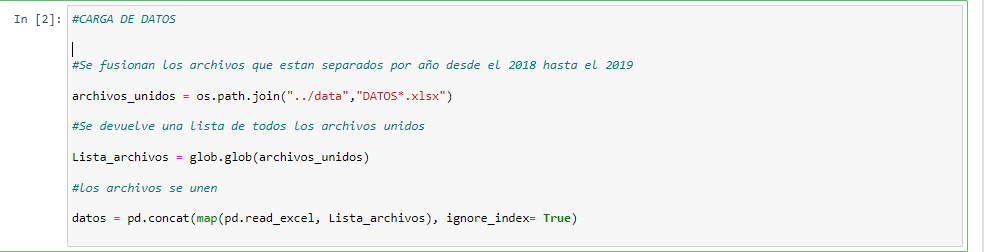
**DATOS – SECRETARIA DE SEGURIDAD CONVIVENCIA Y JUSTICIA**

Se obtienen las fuentes de información de los delitos reportados en la ciudad de Bogotá para los años 2018,2019, 2020, 2021 y 2022 con corte del mes de septiembre.

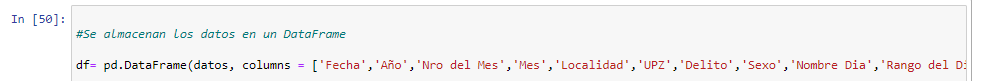
Se importan las librerías que se van a utilizar para el proceso de transformación y limpieza de los datos.



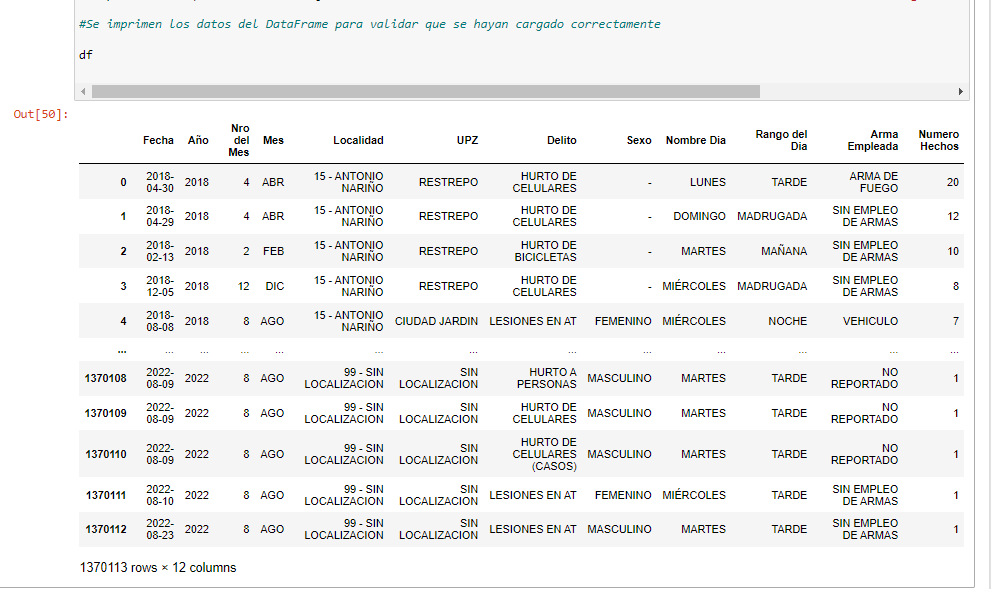
Se cargan los datos que están distribuidos por cinco archivos diferentes por cada año, luego de eso se fusionan en una sola variable la cual después se convierte en una lista de todos los archivos unidos, por ultimo se combina la data de los archivos.



Los datos que se combinaron anteriormente se guardan en un Dataframe para empezar hacer el proceso de limpieza



Se imprime la información del Dataframe para validar que se cargo correctamente y se evidencia que son en total 1.370.113 registros.



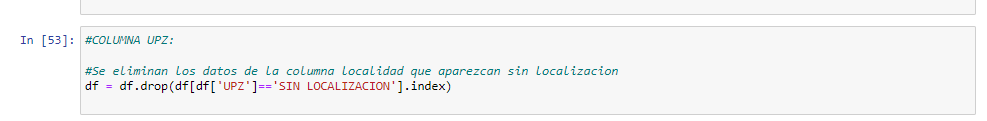
En la columna “LOCALIDAD” se eliminan los datos que aparecen “99 - SIN LOCALIZACION”, adicionalmente se transforman los datos para que aparezca solo el nombre de la localidad sin el código.



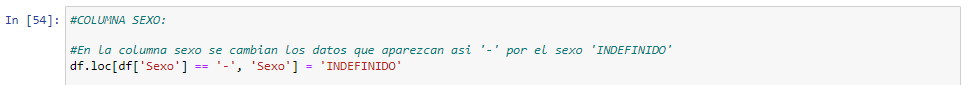
En la columna “DELITOS” se homologa los datos “HURTO DE CELULARES (CASOS) y se unifica con “HURTO DE CELULARES”, adicionalmente se eliminan los delitos con los cuales no vamos a trabajar.



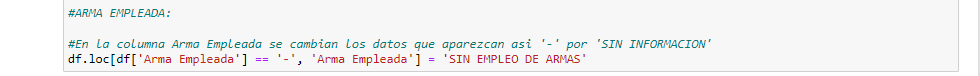
En la columna “UPZ” se eliminan los datos que aparecen “SIN LOCALIZACION”



En la columna “SEXO” se reemplaza el carácter “ - “ por el sexo “INDEFINIDO”



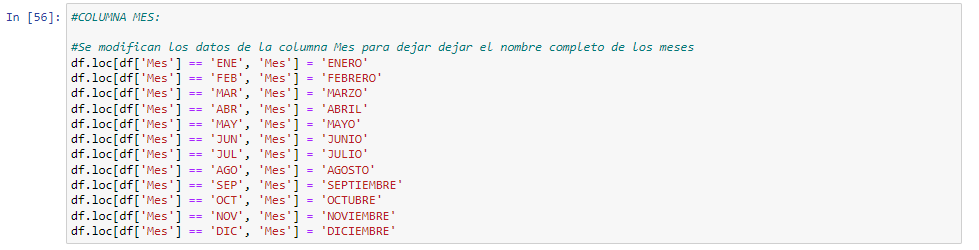
En la columna “ARMA EMPLEADA” se reemplaza el carácter “ - “ por la opción “SIN EMPLEO DE ARMAS”



En la columna “FECHA” se modifica el tipo de dato Date a String con el fin de dejarlo un tipo de dato texto y permita su manipulación en fases posteriores.



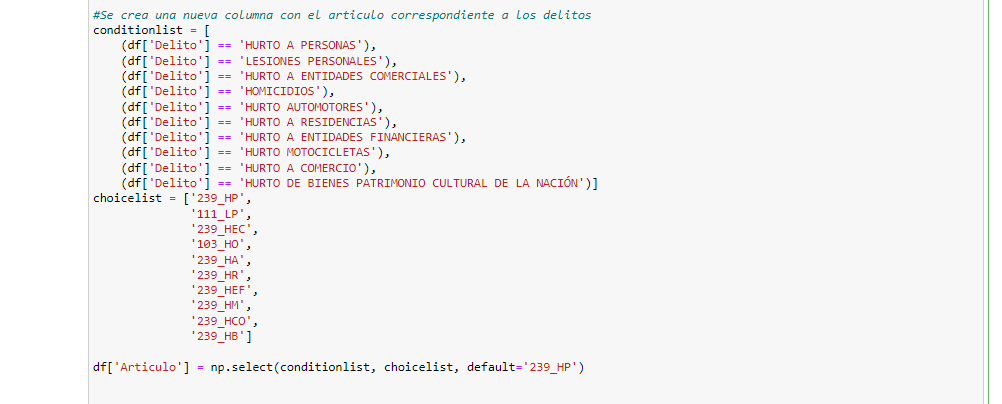
En la columna “MES” se modifican los meses para que queden con el nombre completo.

****

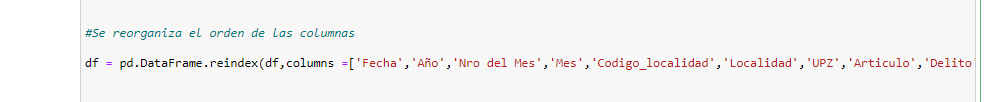
Se crea una nueva columna llamada “CODIGO \_LOCALIDAD” con el fin de guardar los códigos de cada una de las localidades.

****

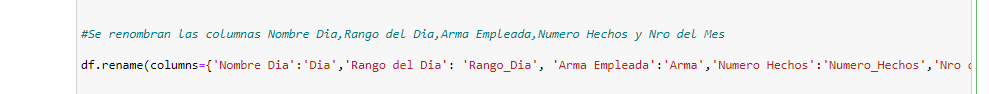
Se crea una nueva columna llamada “ARTICULO” en la cual se identifican los delitos con su articulo correspondiente.

****

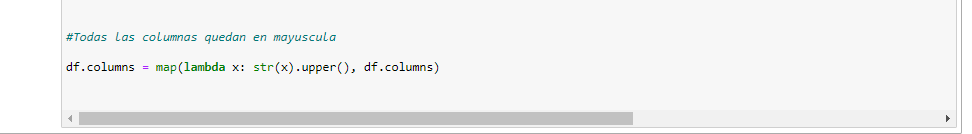
Se reorganiza el orden de las columnas para que la columna “CODIGO\_LOCALIDAD” quede al lado de la columna “LOCALIDAD” y la columna “ARTICULO” quede al lado de la columna “DELITO”.

****

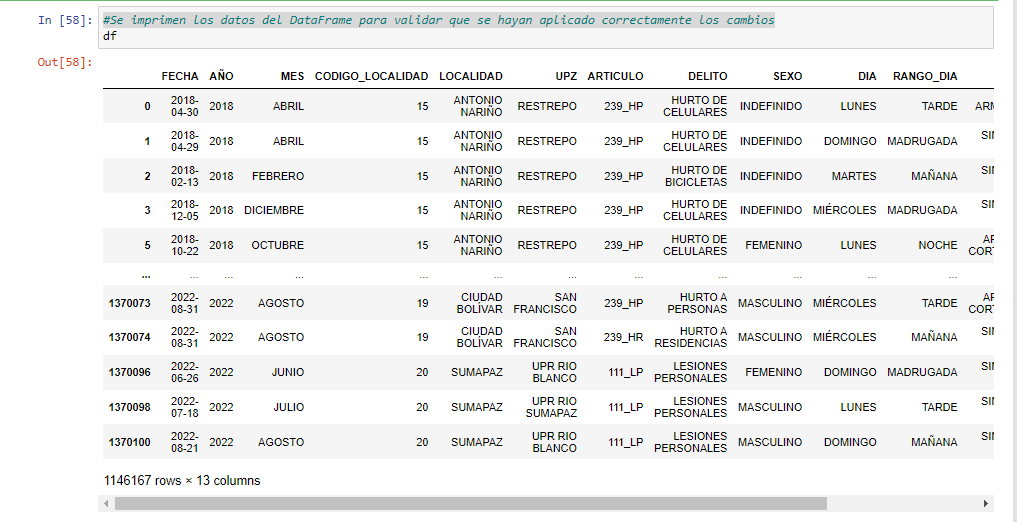
Se renombraron las columnas “NOMBRE DIA”, “RANGO DEL DIA”, “ARMA EMPLEADA”, “NUMERO HECHOS” y “NRO DEL MES” reemplazando los espacios por el carácter “\_” para que sea más fácil su uso en las siguientes fases.

****

Se transforman los nombres de las columnas para que queden en mayúscula.

****

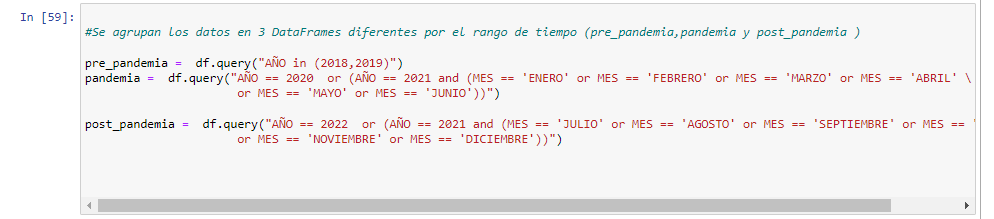
Se vuelven a imprimir los datos para validar que los cambios se hayan aplicado correctamente, se evidencia que después de hacer la limpieza y la transformación de la información quedaron en total 1.146.167 registros.

****

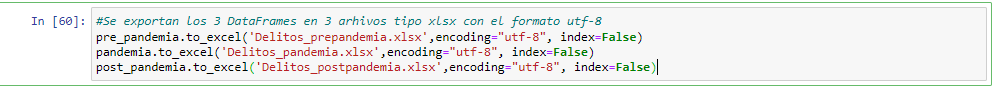
Se divide la información en tres Dataframe diferentes, los cuales se agrupan los datos por diferentes periodos de tiempo:

* Prepandemia: Los datos entre el año 2018 y 2019.
* Pandemia: Los datos del año 2020 y 2021 hasta el mes de junio.
* Postpandemia: Los datos del año 2021 desde julio hasta septiembre de 2022.

Esto se hace con el fin de trabajar la información en las siguientes fases con esta segmentación por periodos de tiempo.

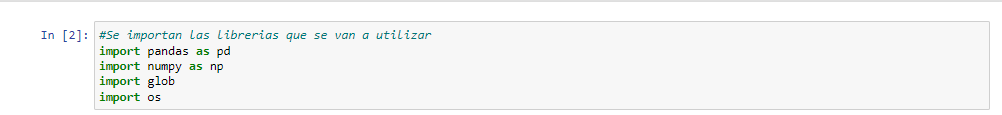
****

Se exportan los tres Dataframes cada uno en un archivo de tipo .xlsx con el formato utf-8 con el fin de que la información se guarde correctamente

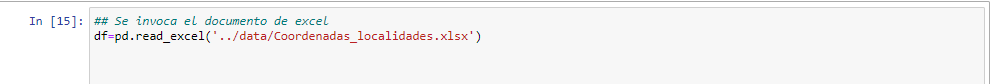
****

**COORDENADAS LOCALIDADES**

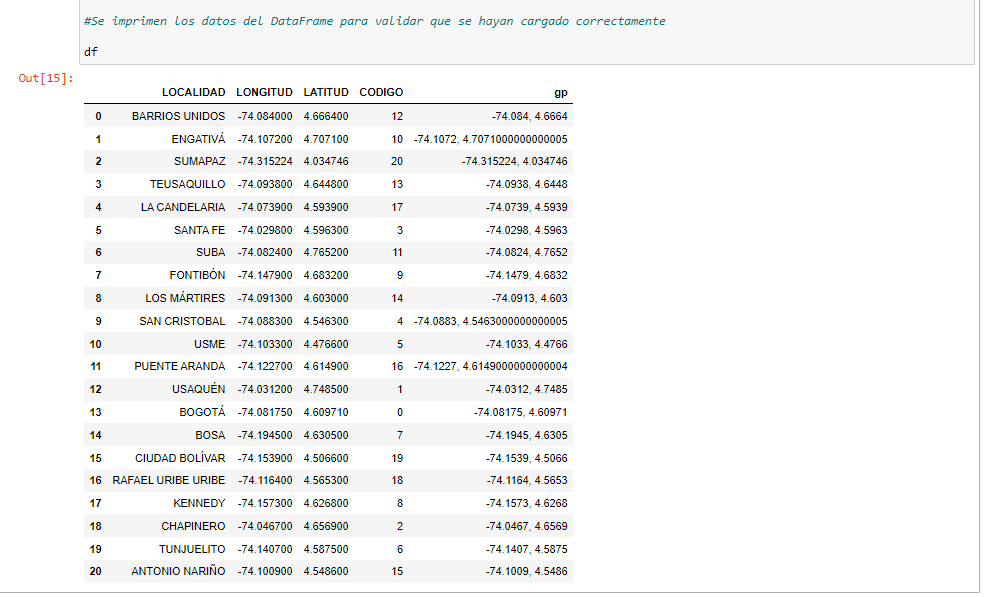
Se importan las librerías que se van a utilizar para el proceso de transformación y limpieza de los datos.

****

Se importa el archivo “Coordenadas\_localidades.xlsx” en el cual esta almacenada la información de las coordenadas geográficas de cada una de las localidades de la ciudad de Bogotá.

****

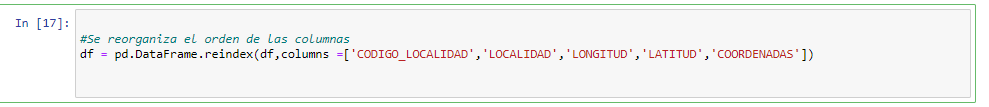
Se imprimen los datos del Dataframe para validar que se hayan cargado correctamente.

****

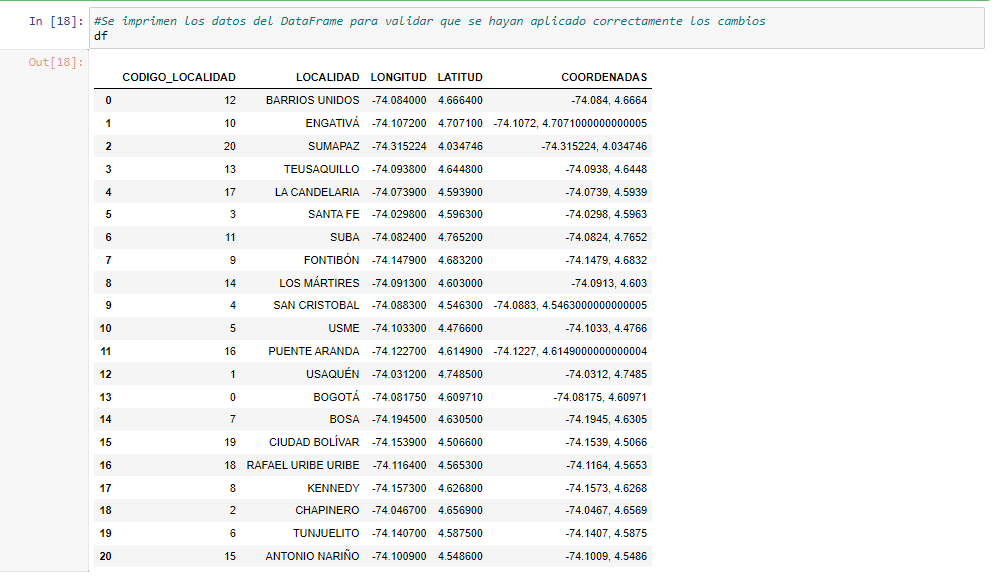
Se renombra la columna “GP” por “COORDENADAS” y la columna “CODIGO” por “CODIGO\_LOCALIDAD”.

****

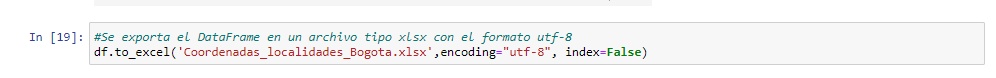
Se reorganiza el orden de las columnas para que la columna “CODIGO\_LOCALIDAD” quede al lado de la columna “LOCALIDAD”.

****

Se imprimen los datos del Dataframe para validad que los cambios se hayan aplicado correctamente.

****

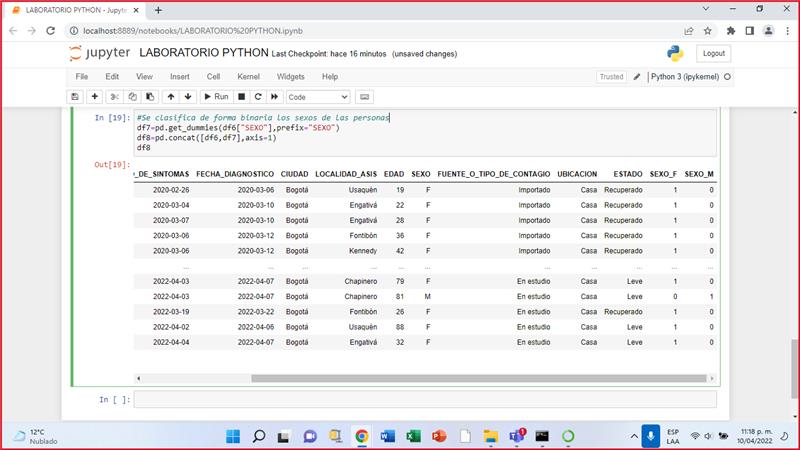
Se exporta los datos del Dataframe en un archivo de tipo .xlsx con el formato utf-8 con el fin de que la información se guarde correctamente.

****

1. **Fase 3: Análisis de los datos y selección de características**

Se toman las tablas, registros y atributos donde se emplea la transformación y limpieza de datos.

De acuerdo al resultado de la fase dos perteneciente a la metodología CRISP-DM, se adquiere una nueva base de datos con 12 columnas y 1.244.510 registros en total, contemplando la integración de las variables sexo\_f y sexo\_m, las cuales generan un valor numérico que clasifican el tipo de sexo en valor binario para análisis estadísticos.

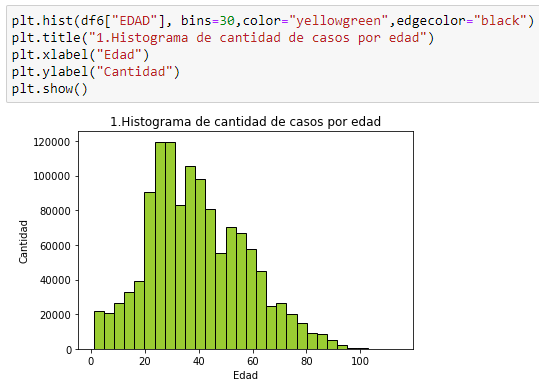


A continuación, se incluyen las nuevas variables categóricas (sexo\_f y sexo\_m), las cuales permitirán agrupar la información y a su vez realizar operaciones matemáticas:

|  |  |
| --- | --- |
| Característica 1. | Variables |
|  | Ciudad |
| Localidad |
| Edad |
| Sexo |
| Sexo\_f |
| Sexo\_m |

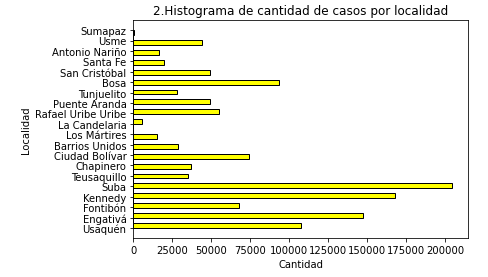
1. **Fase 4: Modelado**

Se seleccionan y aplican las técnicas de modelado que sean pertinentes al problema, cuantas más mejor, y se calibran sus parámetros a valores óptimos.

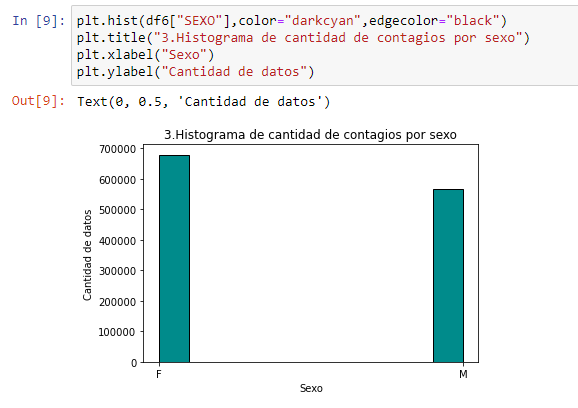


Se evidencia que el máximo de contagios es de 120.000 casos entre el rango de edades de los 20 y 40 años, reflejando un foco de contagio entre la población juvenil de la ciudad de Bogotá.



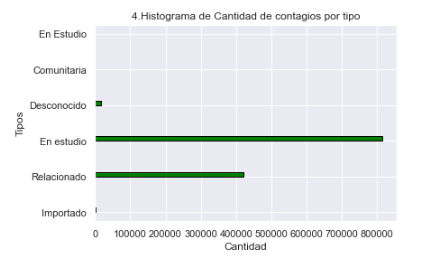


De acuerdo al análisis se evidencia que la localidad de Suba presenta una mayor participación de contagios encabezando con mas de 200.000 casos, debido a que esta localidad reside un mayor número de familias en comparación a la localidad de La Candelaria que es una zona turística y cultural.

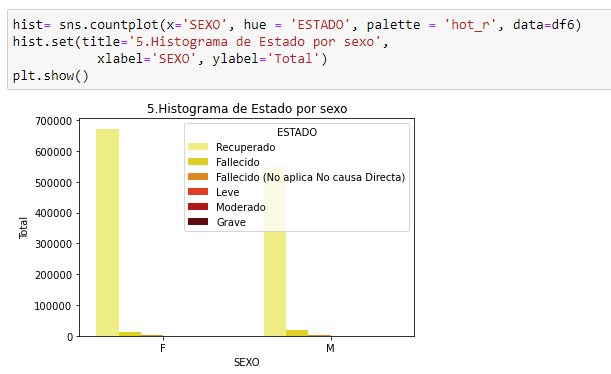


Considerando la cantidad de contagios que han sobresalido en el transcurso de la pandemia, es mas probable que una mujer adquiera coronavirus a comparación de un hombre por sus múltiples ocupaciones, sin embargo, la diferencia en contagios no es superior al 10% frente al total de los casos.

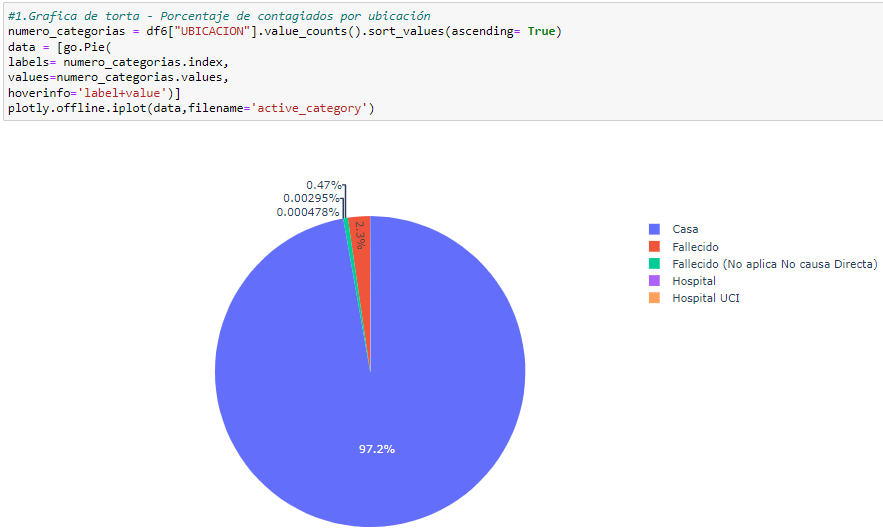




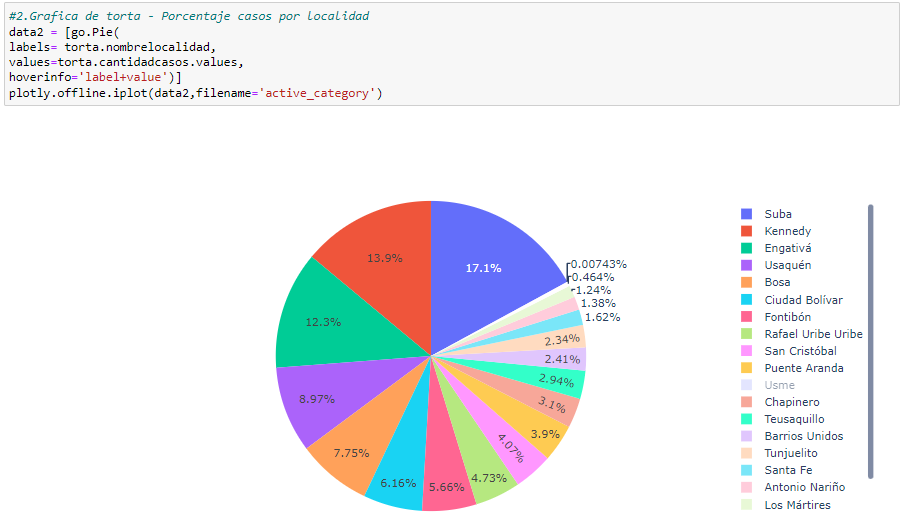
Referente a los casos reportados en la ciudad de Bogotá, la mayoría se encuentran en estudio por el tiempo de aplicaciones en las pruebas PCR y antígenos, casos no reportados en el entorno y validación de posibles prexistencias.



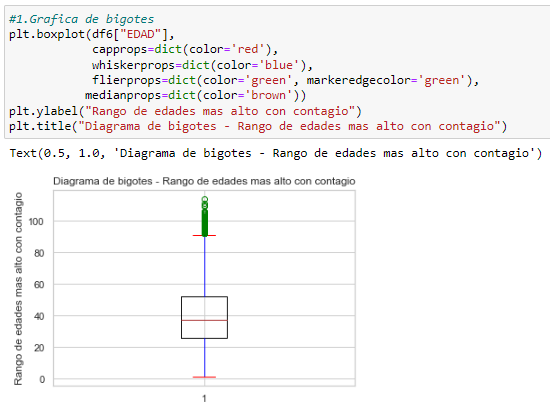
A pesar de que hay un mayor contagio en el genero femenino se evidencia que mueren mas hombres en la ciudad de Bogotá, con un contraste de recuperación mas alto en mujeres sobrepasando mas de 600.000 recuperaciones.



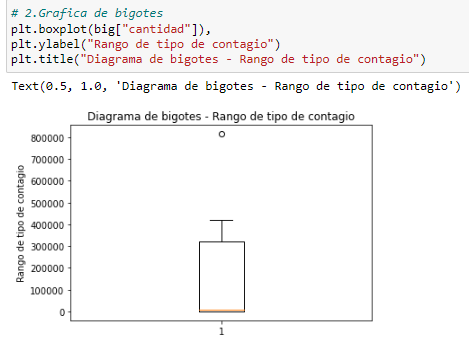
El gran porcentaje de personas contagiadas a pesar que tuvieron síntomas se encuentran en casa, en recuperación o estables, tan solo un 2,3% de personas con coronavirus fallecieron a causa del virus, mientras que el 0,47% corresponde a los fallecimientos con enfermedades asociadas.



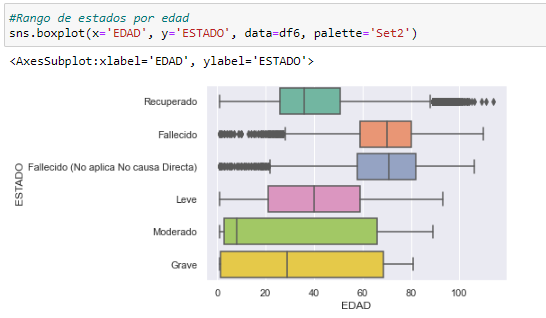
De acuerdo a la ubicación de las localidades se evidencia que hay un mayor porcentaje de contagios sobre la zona norte con tendencia en Suba, Engativá, Usaquén alcanzando un porcentaje del 38,37% de la ciudad de Bogotá.



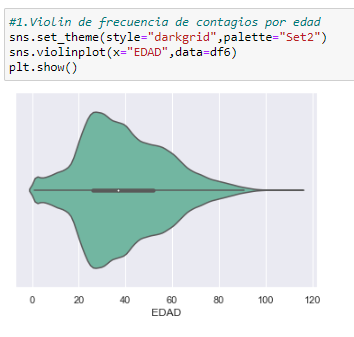
En la población juvenil entre el rango de 20 y 40 años de edad, se destaca que las personas que tienen aproximadamente entre 38 y 39 años son los más vulnerables para contraer el coronavirus, debido a que es una edad con inicio de enfermedades patológicas.



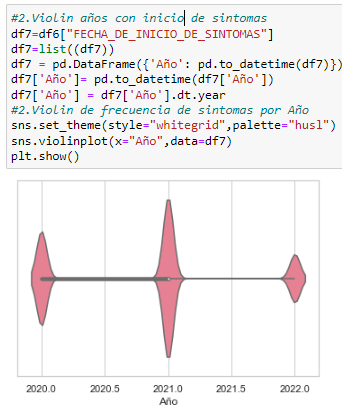
El valor mínimo en el tipo de contagio es 18, con un máximo de 815.178 y una media equivalente al 209.287.



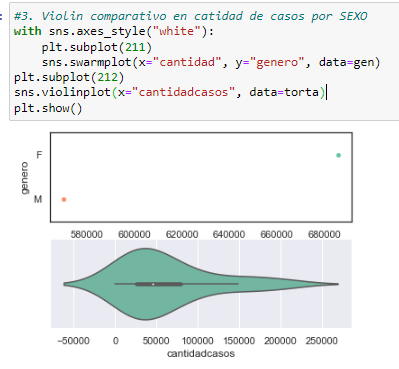
De acuerdo a los casos reportados en la ciudad de Bogotá, hay una mayor concentración de casos graves entre los 0 y 68 años, con una tendencia en los 28 años; los fallecimientos por causas directas o indirectas al COVID se presentan entre los 60 y 80 años siento este un factor de adultos mayores donde se han diagnosticado diferentes enfermedades de base, que generan mayor riesgo de muerte.



La frecuencia con mayor número de contagios se encuentra entre el rango de 18 a 28 años, a partir de los 29 años empieza a descender el riesgo de contagio de acuerdo a la edad.



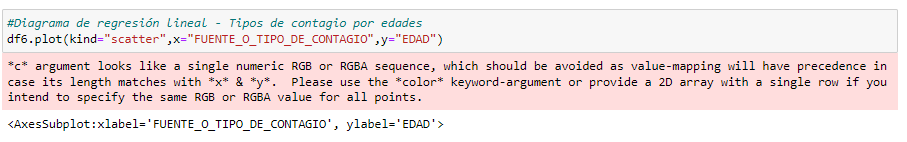
Al iniciar la pandemia de COVID en el año 2020 se evidencia que el volumen de contagios fue promedio de los 3 años de estudio, con un incremento potencial en el año 2021 por la reactivación de la economía e inicio de vacunación. Una vez la mayoría de la población se encuentra vacunada en el año 2022 se evidencia que los contagios disminuyen notablemente.

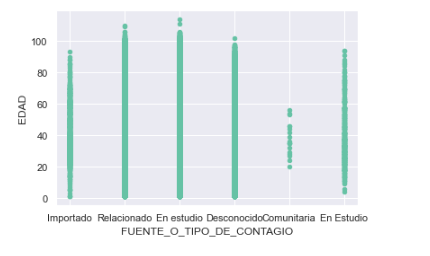


El mayor número de casos se presenta en el género femenino superando los 680.000 casos en comparación del género masculino con menos de 580.000 casos debido a causas anteriormente mencionadas.



Se evidencia una focalización de contagios entre 0 a 50.000 sobre la mayoría de las localidades, lo cual indica que los casos de coronavirus disminuirán en un gran porcentaje en la ciudad de Bogotá, para pocas localidades habrá un aumento significativo.





De acuerdo a la fuente de contagio, se evidencia que entre la edad 0 a 100 años no se ha encontrado la causa por la cual las personas contraen el coronavirus, pero entre la edad de 20 a 60 años el contagio es comunitario, debido a que esta población está más expuesta al momento de salir de su casa para realizar diferentes actividades donde deban transitar en espacios públicos.

1. **ANALISIS Y RESULTADOS**
2. **CONCLUSIONES**
3. **BIBLIOGRAFIA**