1. Describe los problemas: para cada uno de los elementos susceptibles de ser limpiados, indica:

1.1.
- Tres filas con CrimeId 160913455:
  0028 => 160913455 Vandalism 2016-03-31 0:00:00 20:53 2016-03-31 20:53:00 ND 1600 Block Of Sunnydale Av San Francisco CA 1 Premise Address
  1709 => 160913455 Susp 2016-04-01 0:00:00 18:29 2016-04-01 18:29:00 GOA Geary St/larkin St San Francisco CA 1 Intersection
  3794 => 160913455 Passing Call 2016-04-02 0:00:00 17:11 2016-04-02 17:11:00 Not recorded 900 Block Of Market St San Francisco CA 1 Premise Address
  Tres filas con CrimeId 160950496:
  7047 => 160950496 Passing Call 2016-04-04 0:00:00 6:51 2016-04-04 6:51:00 HAN University St/felton St San Francisco CA 1 Intersection
  7048 => 160950496 Suspicious Vehicle 2016-04-04 0:00:00 6:51 2016-04-04 6:51:00 ND 1400 Block Of Cabrillo St San Francisco CA 1 Premise Address
  7049 => 160950496 Trespasser 2016-04-04 0:00:00 6:51 2016-04-04 6:51:00 CAN Block Of Hampshire St San Francisco CA 1 Premise Address

- Detectado mediante estadísticas de la columna.
- No debería haber Ids repetidos, ya que no se estarían identificando unívocamente cada actuación.
- Ver arriba.

▸ Debería revisarse el método de asignado del Id ya que está fallando y genera repeticiones. En los casos detectados habría que asignar Ids distintos para eliminar las repeticiones.

1.2.
- La Columna OriginalCrimeTypeName parece que no está muy estandarizada y me parece poco útil ya que entradas como "Child", "Family", "Female" o "Hot" no tengo claro que identifiquen el tipo de actuación para cualquier usuario de la BD.
- Detectado mediante estadísticas de la columna.
- En bastantes casos aporta muy poca información.
- 2195 => 160923876 Female 2016-04-01 0:00:00 22:55 2016-04-01 22:55:00 ND 1000 Block Of Van Ness Av San Francisco CA 1 Premise Address

▸ Habría que acotar los posibles valores que pueda tomar la columna a una lista cerrada de actuaciones, dejando quizás la opción de "Others" para casos no contemplados, solicitando una descripción del caso cuando se opte por dicha catalogación. En los casos existentes no hay mucho que hacer.

1.3.
- 321 filas con la columna "City" vacía.
- Detectado mediante estadísticas de la columna.
- Faltan datos aunque la mayoría de las veces serían San Francisco.
- 88 => 160920001 Traffic Stop 2016-04-01 0:00:00 0:00 2016-04-01 0:00:00 ADM 16th/ deharo CA 1 Geo-Override

▸ No permitir dejar la columna vacía. Implementar un sistema automático de cumplimentación que permita seleccionar una dirección y se autocompleten los datos que dependan de ella como "City", "State" y puede que "AddressType". En los casos detectados habría que corregir manualmente.

1.4.
- 3 filas con la columna "State" vacía.
- Detectado mediante estadísticas de la columna.
- Parece que los datos están desplados a la derecha, estando el dato de "State" en "AgencyId" y el de esta en "AddressType", habiéndose perdido el dato de esta última.
- 160942112 Auto Boost / Strip 2016-04-03 0:00:00 14:30 2016-04-03 14:30:00 REP Martin Luther King Dr/bowling Green Dr CA 1

▸ Cumplimentación automática como comento en el punto anterior. Quizás también sobraría la columna, dependiendo del tratamiento de los datos, de si se agregan a nivel nacional y hay que identificar el estado o no. En los casos detectados habría que corregir manualmente.

1.5.
- Columna Range sin valores.
- Detectado mediante estadísticas de la columna.
- Si no se cumplimenta nunca, la columna sobraría.
- Todas.

▸ Averiguar porqué existe la columna y se deja en blanco y planterse su eliminación.

1.6.
- Las columnas "OffenseDate" y "CallTime" ofrecen la misma informaciñon que "CallDateTime".
- A simple vista.
- Datos redundantes.
- Todas.

▸ Averiguar porqué se da está redundancia y si no hay un motivo de peso, eliminarla.

2. Propón un formato JSON equivalente al CSV proporcionado. Cada documento JSON debe contener toda la información contenida en la fila correspondiente del documento CSV. Aprovecha esta conversión de formato (CSV a JSON) para eliminar los campos redundantes, los problemas del punto 1, etc. Describe todos estos cambios o transformaciones para que el profesor conozca el proceso utilizado.

```
{
    "CrimeId": Int, //Entero autoincrementable
    "OriginalCrimeTypeName": String,
    "OffenseDate": Date,
    "CallTime": Time,
    "Disposition": String,
    "Address": String,
    "City": String,
    "AddressType": String
} // CallDateTime, State, AgencyId y Range las elimino
```

3. Describe una breve metodología (o aproximación de un método), clara y replicable que, basándose en tu proceso de limpieza, pueda ser aplicada sobre un nuevo dataset similar al que has tratado. El fin es poder afirmar que, usando tu metodología, el nuevo catálogo de datos estará limpio y listo para ser analizado. La metodología debe ser un listado de pasos fáciles de seguir, no ambiguos. Se valorará la claridad de la metodología y su fácil aplicación.

En primer lugar hacer un análisis de la estructura de los datos para entender el dataset y ver posibles fallos de diseño. A coontinuación utilizar la herramienta de "Estadísticas de columna" para detectar valores extraños, campos vacíos, valores muy repetidos, etc... y con la ayuda de filtros de columna revisar los casos particulares.

4. Propón al menos dos mejoras en el conjunto de datos. Después de conocer el conjunto de datos utilizado, qué dos (o más) mejoras propondrías para garantizar la calidad de los datos. Ten en cuenta que estas mejoras pueden afectar al procedimiento de captura de datos o a la estructura de la base de datos, por ello explica tanto la mejora como el procedimiento para llevarla a cabo.

Automatizar o mejorar la asignación de "CrimeId" para evitar los duplicados. Acotar los valores de "OriginalCrimeTypeName" para clarificar los datos y poder acceder mejor a ellos, realizar estadísticas más fiables, etc... Eliminar la columna "CallDateTime" por contener información redundante para aligerar el tamaño de la BD. Implementar un sistema para introducir las direcciones que cumplimente automáticamente los campos relacionados para evitar errores y valorar eliminar el campo "State". Eliminar "AgencyId" y "Range".