

1. Con ayuda de DataCleaner, analiza cada campo del archivo proporcionado para conocer:
 - a. Significado del campo:
 - i. **RFC del vendedor:** RFC del vendedor donde contiene 10 caracteres, de los cuales los primeros 4 son letras y los 6 restantes son números.
 - ii. **Nombre:** Nombre del vendedor.
 - iii. **Apaterno:** Apellido paterno del vendedor.
 - iv. **Amaterno:** Apellido materno del vendedor.
 - v. **Correo electrónico:** Correo electrónico del vendedor.
 - vi. **Genero del vendedor:** Género del vendedor, ya sea femenino o masculino.
 - vii. **Telefono:** Teléfono del vendedor con 10 dígitos.
 - viii. **Anio_Nacimiento:** Año de nacimiento del vendedor.
 - b. Tipo: numérico o nominal.
 - i. **RFC del vendedor:** Nominal.
 - ii. **Nombre:** Nominal.
 - iii. **Apaterno:** Nominal.
 - iv. **Amaterno:** Nominal.
 - v. **Correo electrónico:** Nominal.
 - vi. **Genero del vendedor:** Nominal.
 - vii. **Telefono:** Nominal.
 - viii. **Anio_Nacimiento:** Numérico.
 - c. Número de registros.
 - i. **RFC del vendedor:** 3123
 - ii. **Nombre:** 3123
 - iii. **Apaterno:** 3123
 - iv. **Amaterno:** 3123
 - v. **Correo electrónico:** 3123
 - vi. **Genero del vendedor:** 3123
 - vii. **Telefono:** 3123
 - viii. **Anio_Nacimiento:** 3123
 - d. Número de valores nulos.
 - i. **RFC del vendedor:** 0
 - ii. **Nombre:** 0
 - iii. **Apaterno:** 0
 - iv. **Amaterno:** 0
 - v. **Correo electrónico:** 0
 - vi. **Genero del vendedor:** 0
 - vii. **Telefono:** 0
 - viii. **Anio_Nacimiento:** 0.
 - e. Para campos numéricos obtener:
 - i. **Máximo y mínimo valor.**
 1. 1.999 y 0, respectivamente.
 - ii. **Media.**
 1. 1.486,806
 - iii. **Mediana.**
 1. 1.969
 - iv. **Desviación estándar.**
 1. 748,38
 - f. Para campos nominales:
 - i. **Moda(sólo si los valores del dominio pueden repetirse en varios registros).**

1. **Nombre:** Jose
2. **Apaterno:** Angel
3. **Amaterno:** Lorenzo
4. **Correo electrónico:** jctf@outlook.com
5. **Genero del vendedor:** True
6. **Telefono:** 4777810603

ii. **Máximo número de caracteres.**

1. **RFC del vendedor:** 10
2. **Nombre:** 9
3. **Apaterno:** 9
4. **Amaterno:** 9
5. **Correo electrónico:** 18
6. **Genero del vendedor:** 12
7. **Telefono:** 10

2. Después de este análisis sobre los campos, aplica, si es necesario y con ayuda de Spoon, la o las técnicas de limpieza que consideres apropiadas para obtener datos de calidad. Estas pueden ser:

a. **Normalización/Estandarización.**

- i. Se cambió el nombre del género de Masculino a Hombre y de Femenino a Mujer. Ya que son los valores más usados dentro de su contexto y pueden identificar correctamente el género.
- ii. Los correos solo deben llevar minúsculas.
 1. Se paso todo el correo a minúsculas, así aseguramos estandarización.

b. **Valores Perdidos.**

- i. Hay espacios en blanco, es decir, hay **datos incompletos** en "Género del vendedor"
 1. Se usará la moda para sustituir este dato incompleto, es decir, se usará: True
- ii. Hay espacios en blanco, es decir, hay **datos incompletos** en "Teléfono"
 1. Se usa una constante global "0000000000" para representar este dato perdido ya que es imposible saber el dato.

c. **Valores Atípicos(ruido).**

- i. Hay datos con año 0, 20, 1565, 750, 800, 1856.
 1. Se cambia estos datos por la moda la cual es 1980.
- ii. Hay datos sin coherencia en telefono:
 1. Solo es uno, 2017.12.17. Entonces se optara por usar la constante global "0000000000" ya que no se puede usar la moda, ya que este valor corresponde al vendedor, se podría decir que es único.
- iii. Hay errores de captura en "Género del vendedor" pero al ser pocas las filas que tienen este error se pueden modificar manualmente ya que se sabe el valor del género a partir del nombre del vendedor, estos valores son:
 1. "33 3640 4" por Mujer
 2. "4" por Hombre
- iv. Errores ortográficos en Apaterno.
 1. Tan solo se cambia la primer letra del nombre por mayúscula.
- v. Errores ortográficos en Amaterno.
 1. Tan solo se cambia la primer letra del nombre por mayúscula.