Proyecto_Final_Parte_I

November 16, 2022

1 Proyecto Final Parte I: Limpieza, análisis, visualización y kmeans

1.1 Ciencia y analítica de datos (Gpo 10)

1.1.1 Alumnos:

- Armando Bringas Corpus (A01200230),
- Walter André Hauri Rosales (A01794237)

1.1.2 Profesores:

- Dra. María de la Paz Rico Fernández
- Mtra. Victoria Guerrero Orozco

1.1.3 Fecha: 16 de noviembre de 2022

1.2 Importación de los Datos

[5]:
```
      CLAVE                                          SITIO  \
0    DLAGU8  PRESA EL SAUCILLO 100M AGUAS ARRIBA DE LA CORTINA
1  DLBAJ100                         LOS CABOS SEG 22, 2 ISA10B
2  DLBAJ101                         LOS CABOS SEG 22, 1 ISA10B
3  DLBAJ102                                      LOS CABOS 3
4  DLBAJ103                                      LOS CABOS 1

            ORGANISMO_DE_CUENCA              ESTADO        MUNICIPIO  \
0       LERMA SANTIAGO PACIFICO       AGUASCALIENTES  RINCON DE ROMOS
1  PENINSULA DE BAJA CALIFORNIA  BAJA CALIFORNIA SUR        LOS CABOS
2  PENINSULA DE BAJA CALIFORNIA  BAJA CALIFORNIA SUR        LOS CABOS
3  PENINSULA DE BAJA CALIFORNIA  BAJA CALIFORNIA SUR        LOS CABOS
4  PENINSULA DE BAJA CALIFORNIA  BAJA CALIFORNIA SUR        LOS CABOS

              CUENCA     CUERPO DE AGUA     TIPO     SUBTIPO   LONGITUD  \
0      RIO SAN PEDRO  PRESA EL SAUCILLO  LENTICO       PRESA -102.33911
```

```
1  SAN JOSE DEL CABO    OCEANO PACIFICO  COSTERO  OCEANO-MAR -109.84290
2          SAN LUCAS    OCEANO PACIFICO  COSTERO  OCEANO-MAR -109.86442
3          SAN LUCAS    BAHIA SAN LUCAS  COSTERO       BAHIA -109.88604
4          SAN LUCAS    BAHIA SAN LUCAS  COSTERO       BAHIA -109.89657

    LATITUD  PERIODO DBO_mg/L    CALIDAD_DBO DQO_mg/L  CALIDAD_DQO SST_mg/L  \
0  22.24730   2020.0        6  Buena calidad    54.08  Contaminada    13.75
1  22.90473   2020.0      NaN            NaN      NaN          NaN      <10
2  22.89880   2020.0      NaN            NaN      NaN          NaN      <10
3  22.89609   2020.0      NaN            NaN      NaN          NaN  13.9667
4  22.87694   2020.0      NaN            NaN      NaN          NaN      <10

  CALIDAD_SST COLI_FEC_NMP_100mL CALIDAD_COLI_FEC E_COLI_NMP_100mL  \
0   Excelente               1162      Contaminada               98
1   Excelente                NaN              NaN              NaN
2   Excelente                NaN              NaN              NaN
3   Excelente                NaN              NaN              NaN
4   Excelente                NaN              NaN              NaN

  CALIDAD_E_COLI ENTEROC_NMP_100mL CALIDAD_ENTEROC OD_PORC CALIDAD_OD_PORC  \
0      Excelente               NaN             NaN     NaN             NaN
1            NaN                20       Excelente     NaN             NaN
2            NaN                <3       Excelente     NaN             NaN
3            NaN                <3       Excelente     NaN             NaN
4            NaN                30       Excelente     NaN             NaN

  OD_PORC_SUP CALIDAD_OD_PORC_SUP OD_PORC_MED CALIDAD_OD_PORC_MED OD_PORC_FON  \
0        46.8           Aceptable         NaN                 NaN         NaN
1          92           Excelente        95.4           Excelente        92.2
2          92           Excelente        95.4           Excelente        92.2
3         NaN                 NaN         NaN                 NaN        86.7
4        96.2           Excelente        95.9           Excelente        95.5

  CALIDAD_OD_PORC_FON TOX_D_48_UT CALIDAD_TOX_D_48 TOX_V_15_UT  \
0                 NaN         NaN              NaN         NaN
1           Excelente         NaN              NaN         NaN
2           Excelente         NaN              NaN         NaN
3           Excelente         NaN              NaN         NaN
4           Excelente         NaN              NaN         NaN

  CALIDAD_TOX_V_15 TOX_D_48_SUP_UT CALIDAD TOX_D_48_SUP  TOX_D_48_FON_UT  \
0              NaN              <1            No Toxico              NaN
1              NaN             NaN                  NaN              NaN
2              NaN             NaN                  NaN              NaN
3              NaN             NaN                  NaN              NaN
4              NaN             NaN                  NaN              NaN
```

```
   CALIDAD_TOX_D_48_FON TOX_FIS_SUP_15_UT CALIDAD_TOX_FIS_SUP_15  \
0                   NaN                <1              No Toxico
1                   NaN                <1              No Toxico
2                   NaN                <1              No Toxico
3                   NaN                <1              No Toxico
4                   NaN                <1              No Toxico

   TOX_FIS_FON_15_UT  CALIDAD_TOX_FIS_FON_15 SEMAFORO CONTAMINANTES  \
0                NaN                     NaN     Rojo       DQO,CF,
1                NaN                     NaN    Verde           NaN
2                NaN                     NaN    Verde           NaN
3                NaN                     NaN    Verde           NaN
4                NaN                     NaN    Verde           NaN

  CUMPLE_CON_DBO CUMPLE_CON_DQO CUMPLE_CON_SST CUMPLE_CON_CF  \
0             SI             NO             SI            NO
1             ND             ND             SI            ND
2             ND             ND             SI            ND
3             ND             ND             SI            ND
4             ND             ND             SI            ND

  CUMPLE_CON_E_COLI CUMPLE_CON_ENTEROC CUMPLE_CON_OD CUMPLE_CON_TOX    GRUPO
0                SI                 ND            SI             SI  LENTICO
1                ND                 SI            SI             SI  COSTERO
2                ND                 SI            SI             SI  COSTERO
3                ND                 SI            SI             SI  COSTERO
4                ND                 SI            SI             SI  COSTERO
```

[6]:

```
     CLAVE SITIO ORGANISMO_DE_CUENCA ESTADO MUNICIPIO CUENCA CUERPO DE AGUA  \
4136   NaN   NaN                 NaN    NaN       NaN    NaN            NaN
4137   NaN   NaN                 NaN    NaN       NaN    NaN            NaN
4138   NaN   NaN                 NaN    NaN       NaN    NaN            NaN
4139   NaN   NaN                 NaN    NaN       NaN    NaN            NaN
4140   NaN   NaN                 NaN    NaN       NaN    NaN            NaN

     TIPO SUBTIPO  LONGITUD  LATITUD  PERIODO DBO_mg/L CALIDAD_DBO DQO_mg/L  \
4136  NaN     NaN       NaN      NaN      NaN      NaN         NaN      NaN
4137  NaN     NaN       NaN      NaN      NaN      NaN         NaN      NaN
4138  NaN     NaN       NaN      NaN      NaN      NaN         NaN      NaN
4139  NaN     NaN       NaN      NaN      NaN      NaN         NaN      NaN
4140  NaN     NaN       NaN      NaN      NaN      NaN         NaN      NaN

     CALIDAD_DQO SST_mg/L CALIDAD_SST COLI_FEC_NMP_100mL CALIDAD_COLI_FEC  \
4136         NaN      NaN         NaN                NaN              NaN
4137         NaN      NaN         NaN                NaN              NaN
4138         NaN      NaN         NaN                NaN              NaN
4139         NaN      NaN         NaN                NaN              NaN
```

```
4140         NaN      NaN         NaN                NaN               NaN

     E_COLI_NMP_100mL CALIDAD_E_COLI ENTEROC_NMP_100mL CALIDAD_ENTEROC  \
4136              NaN            NaN               NaN             NaN
4137              NaN            NaN               NaN             NaN
4138              NaN            NaN               NaN             NaN
4139              NaN            NaN               NaN             NaN
4140              NaN            NaN               NaN             NaN

     OD_PORC CALIDAD_OD_PORC OD_PORC_SUP CALIDAD_OD_PORC_SUP OD_PORC_MED  \
4136     NaN             NaN         NaN                 NaN         NaN
4137     NaN             NaN         NaN                 NaN         NaN
4138     NaN             NaN         NaN                 NaN         NaN
4139     NaN             NaN         NaN                 NaN         NaN
4140     NaN             NaN         NaN                 NaN         NaN

     CALIDAD_OD_PORC_MED OD_PORC_FON CALIDAD_OD_PORC_FON TOX_D_48_UT  \
4136                 NaN         NaN                 NaN         NaN
4137                 NaN         NaN                 NaN         NaN
4138                 NaN         NaN                 NaN         NaN
4139                 NaN         NaN                 NaN         NaN
4140                 NaN         NaN                 NaN         NaN

     CALIDAD_TOX_D_48 TOX_V_15_UT CALIDAD_TOX_V_15 TOX_D_48_SUP_UT  \
4136              NaN         NaN              NaN             NaN
4137              NaN         NaN              NaN             NaN
4138              NaN         NaN              NaN             NaN
4139              NaN         NaN              NaN             NaN
4140              NaN         NaN              NaN             NaN

     CALIDAD TOX_D_48_SUP  TOX_D_48_FON_UT  CALIDAD_TOX_D_48_FON  \
4136                 NaN              NaN                   NaN
4137                 NaN              NaN                   NaN
4138                 NaN              NaN                   NaN
4139                 NaN              NaN                   NaN
4140                 NaN              NaN                   NaN

     TOX_FIS_SUP_15_UT CALIDAD_TOX_FIS_SUP_15  TOX_FIS_FON_15_UT  \
4136               NaN                    NaN                NaN
4137               NaN                    NaN                NaN
4138               NaN                    NaN                NaN
4139               NaN                    NaN                NaN
4140               NaN                    NaN                NaN

      CALIDAD_TOX_FIS_FON_15 SEMAFORO CONTAMINANTES CUMPLE_CON_DBO  \
4136                    NaN      NaN           NaN            NaN
4137                    NaN      NaN           NaN            NaN
```

```
4138                    NaN      NaN           NaN            NaN
4139                    NaN      NaN           NaN            NaN
4140                    NaN      NaN           NaN            NaN

     CUMPLE_CON_DQO CUMPLE_CON_SST CUMPLE_CON_CF CUMPLE_CON_E_COLI  \
4136            NaN            NaN           NaN               NaN
4137            NaN            NaN           NaN               NaN
4138            NaN            NaN           NaN               NaN
4139            NaN            NaN           NaN               NaN
4140            NaN            NaN           NaN               NaN

     CUMPLE_CON_ENTEROC CUMPLE_CON_OD CUMPLE_CON_TOX GRUPO
4136                NaN           NaN            NaN   NaN
4137                NaN           NaN            NaN   NaN
4138                NaN           NaN            NaN   NaN
4139                NaN           NaN            NaN   NaN
4140                NaN           NaN            NaN   NaN
```

[7]:
```
227755
```

[8]:
```
Index(['CLAVE', 'SITIO', 'ORGANISMO_DE_CUENCA', 'ESTADO', 'MUNICIPIO',
       'CUENCA', 'CUERPO DE AGUA', 'TIPO', 'SUBTIPO', 'LONGITUD', 'LATITUD',
       'PERIODO', 'DBO_mg/L', 'CALIDAD_DBO', 'DQO_mg/L', 'CALIDAD_DQO',
       'SST_mg/L', 'CALIDAD_SST', 'COLI_FEC_NMP_100mL', 'CALIDAD_COLI_FEC',
       'E_COLI_NMP_100mL', 'CALIDAD_E_COLI', 'ENTEROC_NMP_100mL',
       'CALIDAD_ENTEROC', 'OD_PORC', 'CALIDAD_OD_PORC', 'OD_PORC_SUP',
       'CALIDAD_OD_PORC_SUP', 'OD_PORC_MED', 'CALIDAD_OD_PORC_MED',
       'OD_PORC_FON', 'CALIDAD_OD_PORC_FON', 'TOX_D_48_UT', 'CALIDAD_TOX_D_48',
       'TOX_V_15_UT', 'CALIDAD_TOX_V_15', 'TOX_D_48_SUP_UT',
       'CALIDAD TOX_D_48_SUP', 'TOX_D_48_FON_UT', 'CALIDAD_TOX_D_48_FON',
       'TOX_FIS_SUP_15_UT', 'CALIDAD_TOX_FIS_SUP_15', 'TOX_FIS_FON_15_UT',
       'CALIDAD_TOX_FIS_FON_15', 'SEMAFORO', 'CONTAMINANTES', 'CUMPLE_CON_DBO',
       'CUMPLE_CON_DQO', 'CUMPLE_CON_SST', 'CUMPLE_CON_CF',
       'CUMPLE_CON_E_COLI', 'CUMPLE_CON_ENTEROC', 'CUMPLE_CON_OD',
       'CUMPLE_CON_TOX', 'GRUPO'],
      dtype='object')
```

[9]:
```
(55,)
```

1.3 Limpieza de los Datos

Analizamos la distribución de los datos nulos, es decir, cuántos atributos se tienen en nulos.

[10]:
```
16    1703
55     648
25     351
17     319
23     228
```

```
12     178
13     139
21     131
24     128
14     123
15      75
22      38
18      32
20      23
27      15
19       7
26       2
29       1
dtype: int64
```

Encontramos que 648 registros tienen todos los valores faltantes, de manera que procedemos a eliminarlos.

Eliminamos todos los rergistros que no tengan un valor del semáforo válido, ya que es nuestra variable a predecir.

Procedemos a analizar el número de registros con datos nulos por cada atributo.

```
Número de columnas con valores NAN: 36
```

[13]:
```
CUENCA                     1
CUERPO DE AGUA            14
SUBTIPO                   14
DBO_mg/L                 912
CALIDAD_DBO              912
DQO_mg/L                 912
CALIDAD_DQO              912
SST_mg/L                   4
CALIDAD_SST                4
COLI_FEC_NMP_100mL       911
CALIDAD_COLI_FEC         911
E_COLI_NMP_100mL         911
CALIDAD_E_COLI           911
ENTEROC_NMP_100mL       2589
CALIDAD_ENTEROC         2589
OD_PORC                 1696
CALIDAD_OD_PORC         1696
OD_PORC_SUP             1874
CALIDAD_OD_PORC_SUP     1874
OD_PORC_MED             3006
CALIDAD_OD_PORC_MED     3006
OD_PORC_FON             2547
CALIDAD_OD_PORC_FON     2547
TOX_D_48_UT             1677
```

```
CALIDAD_TOX_D_48          1677
TOX_V_15_UT               1674
CALIDAD_TOX_V_15          1674
TOX_D_48_SUP_UT           2731
CALIDAD TOX_D_48_SUP      2731
TOX_D_48_FON_UT           3493
CALIDAD_TOX_D_48_FON      3493
TOX_FIS_SUP_15_UT         1819
CALIDAD_TOX_FIS_SUP_15    1819
TOX_FIS_FON_15_UT         3493
CALIDAD_TOX_FIS_FON_15    3493
CONTAMINANTES             1267
dtype: int64
```

[14]:

	CUENCA	CUERPO DE AGUA	SUBTIPO	DBO_mg/L	\
0	RIO SAN PEDRO	PRESA EL SAUCILLO	PRESA	6	
1	SAN JOSE DEL CABO	OCEANO PACIFICO	OCEANO-MAR	NaN	
2	SAN LUCAS	OCEANO PACIFICO	OCEANO-MAR	NaN	
3	SAN LUCAS	BAHIA SAN LUCAS	BAHIA	NaN	
4	SAN LUCAS	BAHIA SAN LUCAS	BAHIA	NaN	
…	…	…	…	…	
3488	RIO BRAVO 11	RIO SALADO	RIO	<2	
3489	RIO BRAVO 11	EL LAGUITO	LAGO	39.09	
3490	RIO BRAVO 11	RIO BRAVO	RIO	4.4	
3491	RIO BRAVO 10	RIO BRAVO	RIO	<2	
3492	RIO BRAVO 10	RIO BRAVO	RIO	<2	

	CALIDAD_DBO	DQO_mg/L	CALIDAD_DQO	SST_mg/L	CALIDAD_SST	\
0	Buena calidad	54.08	Contaminada	13.75	Excelente	
1	NaN	NaN	NaN	<10	Excelente	
2	NaN	NaN	NaN	<10	Excelente	
3	NaN	NaN	NaN	13.9667	Excelente	
4	NaN	NaN	NaN	<10	Excelente	
…	…	…	…	…	…	
3488	Excelente	<10	Excelente	122	Aceptable	
3489	Contaminada	115.88	Contaminada	54	Buena calidad	
3490	Buena calidad	87.64	Contaminada	70	Buena calidad	
3491	Excelente	35.92	Aceptable	56	Buena calidad	
3492	Excelente	34.6	Aceptable	45	Buena calidad	

	COLI_FEC_NMP_100mL	CALIDAD_COLI_FEC	E_COLI_NMP_100mL	\
0	1162	Contaminada	98	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	NaN	NaN	NaN	
4	NaN	NaN	NaN	
…	…	…	…	

```
3488              2400      Contaminada             2400
3489              2400      Contaminada                4
3490               240        Aceptable               15
3491              2400      Contaminada               11
3492              2400      Contaminada               15

              CALIDAD_E_COLI ENTEROC_NMP_100mL CALIDAD_ENTEROC OD_PORC  \
0                  Excelente               NaN            NaN     NaN
1                        NaN                20       Excelente     NaN
2                        NaN                <3       Excelente     NaN
3                        NaN                <3       Excelente     NaN
4                        NaN                30       Excelente     NaN
…                          …               …             …     …
3488  Fuertemente contaminada               NaN            NaN    69.1
3489                Excelente               NaN            NaN     NaN
3490                Excelente               NaN            NaN    78.8
3491                Excelente               NaN            NaN    76.2
3492                Excelente               NaN            NaN    85.4

     CALIDAD_OD_PORC OD_PORC_SUP CALIDAD_OD_PORC_SUP OD_PORC_MED  \
0                NaN        46.8           Aceptable         NaN
1                NaN          92           Excelente        95.4
2                NaN          92           Excelente        95.4
3                NaN         NaN                 NaN         NaN
4                NaN        96.2           Excelente        95.9
…                  …           …                   …           …
3488   Buena calidad         NaN                 NaN         NaN
3489             NaN        91.6           Excelente         NaN
3490       Excelente         NaN                 NaN         NaN
3491       Excelente         NaN                 NaN         NaN
3492       Excelente         NaN                 NaN         NaN

     CALIDAD_OD_PORC_MED OD_PORC_FON CALIDAD_OD_PORC_FON TOX_D_48_UT  \
0                    NaN         NaN                 NaN         NaN
1              Excelente        92.2           Excelente         NaN
2              Excelente        92.2           Excelente         NaN
3                    NaN        86.7           Excelente         NaN
4              Excelente        95.5           Excelente         NaN
…                    …           …                   …           …
3488                 NaN         NaN                 NaN          <1
3489                 NaN         NaN                 NaN         NaN
3490                 NaN         NaN                 NaN          <1
3491                 NaN         NaN                 NaN          <1
3492                 NaN         NaN                 NaN          <1

     CALIDAD_TOX_D_48 TOX_V_15_UT CALIDAD_TOX_V_15 TOX_D_48_SUP_UT  \
0                 NaN         NaN              NaN              <1
```

```
1                NaN         NaN              NaN             NaN
2                NaN         NaN              NaN             NaN
3                NaN         NaN              NaN             NaN
4                NaN         NaN              NaN             NaN
…                …           …                …               …
3488      No Toxico          <1       No Toxico              NaN
3489            NaN         NaN             NaN               <1
3490      No Toxico          <1       No Toxico              NaN
3491      No Toxico          <1       No Toxico              NaN
3492      No Toxico          <1       No Toxico              NaN

     CALIDAD TOX_D_48_SUP  TOX_D_48_FON_UT  CALIDAD_TOX_D_48_FON  \
0               No Toxico              NaN                   NaN
1                     NaN              NaN                   NaN
2                     NaN              NaN                   NaN
3                     NaN              NaN                   NaN
4                     NaN              NaN                   NaN
…                     …                …                     …
3488                  NaN              NaN                   NaN
3489            No Toxico              NaN                   NaN
3490                  NaN              NaN                   NaN
3491                  NaN              NaN                   NaN
3492                  NaN              NaN                   NaN

     TOX_FIS_SUP_15_UT CALIDAD_TOX_FIS_SUP_15  TOX_FIS_FON_15_UT  \
0                   <1              No Toxico                NaN
1                   <1              No Toxico                NaN
2                   <1              No Toxico                NaN
3                   <1              No Toxico                NaN
4                   <1              No Toxico                NaN
…                   …                       …                …
3488               NaN                    NaN              NaN
3489                <1              No Toxico                NaN
3490               NaN                    NaN              NaN
3491               NaN                    NaN              NaN
3492               NaN                    NaN              NaN

      CALIDAD_TOX_FIS_FON_15 CONTAMINANTES
0                        NaN       DQO,CF,
1                        NaN           NaN
2                        NaN           NaN
3                        NaN           NaN
4                        NaN           NaN
…                        …             …
3488                     NaN    CF,E_COLI,
3489                     NaN   DBO,DQO,CF,
3490                     NaN          DQO,
```

```
3491                    NaN          CF,
3492                    NaN          CF,

[3493 rows x 36 columns]
```

1.3.1 Cuenca

[15]:

```
           CLAVE                                SITIO ORGANISMO_DE_CUENCA  \
820  DLMIC1546W1  DESEMBOCADURA RIO CARRIZAL DE ARTEAGA              BALSAS

                 ESTADO        MUNICIPIO CUENCA        CUERPO DE AGUA  \
820  MICHOACAN DE OCAMPO  LAZARO CARDENAS    NaN  TRANSICION RIO - MAR

       TIPO SUBTIPO   LONGITUD   LATITUD  PERIODO DBO_mg/L CALIDAD_DBO  \
820  COSTERO  ESTERO -102.37853  17.98233   2020.0      NaN         NaN

    DQO_mg/L CALIDAD_DQO SST_mg/L CALIDAD_SST COLI_FEC_NMP_100mL  \
820      NaN         NaN     10.8   Excelente                NaN

    CALIDAD_COLI_FEC E_COLI_NMP_100mL CALIDAD_E_COLI ENTEROC_NMP_100mL  \
820              NaN              NaN            NaN                63

    CALIDAD_ENTEROC OD_PORC CALIDAD_OD_PORC OD_PORC_SUP CALIDAD_OD_PORC_SUP  \
820       Excelente     NaN             NaN        91.7           Excelente

    OD_PORC_MED CALIDAD_OD_PORC_MED OD_PORC_FON CALIDAD_OD_PORC_FON  \
820         NaN                 NaN         NaN                 NaN

    TOX_D_48_UT CALIDAD_TOX_D_48 TOX_V_15_UT CALIDAD_TOX_V_15 TOX_D_48_SUP_UT  \
820         NaN              NaN         NaN              NaN             NaN

    CALIDAD TOX_D_48_SUP  TOX_D_48_FON_UT  CALIDAD_TOX_D_48_FON  \
820                  NaN              NaN                   NaN

    TOX_FIS_SUP_15_UT CALIDAD_TOX_FIS_SUP_15  TOX_FIS_FON_15_UT  \
820                <1              No Toxico                NaN

     CALIDAD_TOX_FIS_FON_15 SEMAFORO CONTAMINANTES CUMPLE_CON_DBO  \
820                     NaN    Verde           NaN             ND

    CUMPLE_CON_DQO CUMPLE_CON_SST CUMPLE_CON_CF CUMPLE_CON_E_COLI  \
820             ND             SI            ND                ND

    CUMPLE_CON_ENTEROC CUMPLE_CON_OD CUMPLE_CON_TOX    GRUPO
820                 SI            SI             SI  COSTERO
```

[16]:

	CLAVE	SITIO
816	DLAGU51M1	PRESA ABELARDO RODRIGUEZ 100M AGUAS ARRIBA DE …
817	DLMIC1543W1	LAGUNA COSTERA EL CAIMAN 4
818	DLMIC1544W1	LAGUNA COSTERA EL CAIMAN 5
819	DLMIC1545	DESEMBOCADURA RIO CARRIZAL DE ARTEAGA 2
820	DLMIC1546W1	DESEMBOCADURA RIO CARRIZAL DE ARTEAGA
821	DLMIC1548	RIO BALSAS CORRIENTE IZQ PUENTE
822	DLMIC1549	RIO BALSAS 5
823	DLMIC1553	RIO BALSAS PUENTE VIEJO

	ORGANISMO_DE_CUENCA	ESTADO	MUNICIPIO
816	LERMA SANTIAGO PACIFICO	AGUASCALIENTES	JESUS MARIA
817	BALSAS	MICHOACAN DE OCAMPO	LAZARO CARDENAS
818	BALSAS	MICHOACAN DE OCAMPO	LAZARO CARDENAS
819	BALSAS	MICHOACAN DE OCAMPO	LAZARO CARDENAS
820	BALSAS	MICHOACAN DE OCAMPO	LAZARO CARDENAS
821	BALSAS	MICHOACAN DE OCAMPO	LAZARO CARDENAS
822	BALSAS	MICHOACAN DE OCAMPO	LAZARO CARDENAS
823	BALSAS	MICHOACAN DE OCAMPO	LAZARO CARDENAS

	CUENCA	CUERPO DE AGUA
816	PRESA EL NIAGARA	PRESA ABELARDO RODRIGUEZ
817	RIO ACAPILCAN	LAGUNA COSTERA EL CAIMAN
818	RIO ACAPILCAN	LAGUNA COSTERA EL CAIMAN
819	RIO ACAPILCAN	TRANSICION RIO - MAR
820	NaN	TRANSICION RIO - MAR
821	RIO BAJO BALSAS	RIO BALSAS CORRIENTE IZQ PUENTE
822	RIO BAJO BALSAS	RIO BALSAS
823	RIO BAJO BALSAS	RIO BALSAS PUENTE VIEJO

	TIPO	SUBTIPO	LONGITUD	LATITUD	PERIODO
816	LENTICO (HUMEDAL)	PRESA	-102.42860	21.91633	2020.0
817	LENTICO - COSTERO (HUMEDAL)	LAGUNA	-102.32441	17.97750	2020.0
818	LENTICO - COSTERO (HUMEDAL)	LAGUNA	-102.32790	17.98442	2020.0
819	LOTICO - COSTERO	ESTERO	-102.37526	17.98857	2020.0
820	COSTERO	ESTERO	-102.37853	17.98233	2020.0
821	LOTICO - COSTERO (HUMEDAL)	RIO	-102.19769	17.97331	2020.0
822	LOTICO	RIO	-102.20828	18.00848	2020.0
823	LENTICO (HUMEDAL)	LAGO	-102.19073	18.02432	2020.0

	DBO_mg/L	CALIDAD_DBO	DQO_mg/L	CALIDAD_DQO	SST_mg/L	CALIDAD_SST
816	<2	Excelente	28.83	Aceptable	10.5	Excelente
817	NaN	NaN	NaN	NaN	19.5	Excelente
818	<2	Excelente	29.68	Aceptable	<10	Excelente
819	4.9	Buena calidad	26.88	Aceptable	<10	Excelente
820	NaN	NaN	NaN	NaN	10.8	Excelente
821	<2	Excelente	11.592	Buena calidad	<10	Excelente

```
822       <2      Excelente    18.22  Buena calidad      <10   Excelente
823       <2      Excelente      <10      Excelente      <10   Excelente

    COLI_FEC_NMP_100mL         CALIDAD_COLI_FEC E_COLI_NMP_100mL  \
816                379                Aceptable               10
817                NaN                      NaN              NaN
818              24196  Fuertemente contaminada            24196
819               6867              Contaminada              384
820                NaN                      NaN              NaN
821              24196  Fuertemente contaminada            17329
822               6867              Contaminada             4611
823              24196  Fuertemente contaminada               75

             CALIDAD_E_COLI ENTEROC_NMP_100mL          CALIDAD_ENTEROC  \
816               Excelente               NaN                      NaN
817                     NaN             17329  Fuertemente contaminada
818  Fuertemente contaminada               NaN                      NaN
819            Buena calidad               NaN                      NaN
820                     NaN                63                Excelente
821  Fuertemente contaminada               NaN                      NaN
822  Fuertemente contaminada               NaN                      NaN
823               Excelente               NaN                      NaN

    OD_PORC CALIDAD_OD_PORC OD_PORC_SUP CALIDAD_OD_PORC_SUP OD_PORC_MED  \
816     NaN             NaN       100.7           Excelente        80.4
817     NaN             NaN        26.2         Contaminada         NaN
818     NaN             NaN          26         Contaminada         NaN
819    76.5       Excelente         NaN                 NaN         NaN
820     NaN             NaN        91.7           Excelente         NaN
821    98.1       Excelente         NaN                 NaN         NaN
822    63.5   Buena calidad         NaN                 NaN         NaN
823     NaN             NaN          52       Buena calidad         NaN

    CALIDAD_OD_PORC_MED OD_PORC_FON CALIDAD_OD_PORC_FON TOX_D_48_UT  \
816           Excelente        50.2       Buena calidad         NaN
817                 NaN         NaN                 NaN         NaN
818                 NaN         NaN                 NaN         NaN
819                 NaN         NaN                 NaN          <1
820                 NaN         NaN                 NaN         NaN
821                 NaN         NaN                 NaN          <1
822                 NaN         NaN                 NaN          <1
823                 NaN         NaN                 NaN         NaN

    CALIDAD_TOX_D_48 TOX_V_15_UT CALIDAD_TOX_V_15 TOX_D_48_SUP_UT  \
816              NaN         NaN              NaN              <1
817              NaN         NaN              NaN             NaN
818              NaN         NaN              NaN              <1
```

```
819        No Toxico          <1        No Toxico             NaN
820              NaN         NaN              NaN             NaN
821        No Toxico          <1        No Toxico             NaN
822        No Toxico          <1        No Toxico             NaN
823              NaN         NaN              NaN              <1

    CALIDAD TOX_D_48_SUP  TOX_D_48_FON_UT  CALIDAD_TOX_D_48_FON  \
816           No Toxico              NaN                   NaN
817                 NaN              NaN                   NaN
818           No Toxico              NaN                   NaN
819                 NaN              NaN                   NaN
820                 NaN              NaN                   NaN
821                 NaN              NaN                   NaN
822                 NaN              NaN                   NaN
823           No Toxico              NaN                   NaN

    TOX_FIS_SUP_15_UT CALIDAD_TOX_FIS_SUP_15  TOX_FIS_FON_15_UT  \
816                <1             No Toxico                NaN
817                <1             No Toxico                NaN
818                <1             No Toxico                NaN
819               NaN                   NaN                NaN
820                <1             No Toxico                NaN
821               NaN                   NaN                NaN
822               NaN                   NaN                NaN
823                <1             No Toxico                NaN

     CALIDAD_TOX_FIS_FON_15  SEMAFORO    CONTAMINANTES CUMPLE_CON_DBO  \
816                     NaN     Verde              NaN             SI
817                     NaN      Rojo    ENT_FEC,OD%S,             ND
818                     NaN  Amarillo  CF,E_COLI,OD%S,             SI
819                     NaN  Amarillo              CF,             SI
820                     NaN     Verde              NaN             ND
821                     NaN  Amarillo       CF,E_COLI,             SI
822                     NaN  Amarillo       CF,E_COLI,             SI
823                     NaN  Amarillo              CF,             SI

    CUMPLE_CON_DQO CUMPLE_CON_SST CUMPLE_CON_CF CUMPLE_CON_E_COLI  \
816             SI             SI            SI                SI
817             ND             SI            ND                ND
818             SI             SI            NO                NO
819             SI             SI            NO                SI
820             ND             SI            ND                ND
821             SI             SI            NO                NO
822             SI             SI            NO                NO
823             SI             SI            NO                SI

    CUMPLE_CON_ENTEROC CUMPLE_CON_OD CUMPLE_CON_TOX    GRUPO
```

```
816                ND           SI            SI  LENTICO
817                NO           NO            SI  COSTERO
818                ND           NO            SI  COSTERO
819                ND           SI            SI  COSTERO
820                SI           SI            SI  COSTERO
821                ND           SI            SI  COSTERO
822                ND           SI            SI   LOTICO
823                ND           SI            SI  LENTICO
```

Podemos observar que los registros eestán acomodados de forma que las cuencas, municipios y estados aparezcan seriados. Partiendo de esta premisa, el valor faltante se encuentra en Michoacán, en el municipio de Lázaro Cárdenas y tiene una mayor similitud con el registro superior a él. De forma que se colocará la misma cuenca (*RIO ACAPILCAN*).

1.3.2 Cuerpo de agua

[18]:

```
          CLAVE      SITIO ORGANISMO_DE_CUENCA   ESTADO  MUNICIPIO   CUENCA  \
1573  DLTAB5552   MANATI 1        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1574  DLTAB5553   MANATI 2        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1575  DLTAB5554   MANATI 3        FRONTERA SUR  TABASCO     JONUTA  Chilapa
1576  DLTAB5555   MANATI 4        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1577  DLTAB5556   MANATI 5        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1578  DLTAB5557   MANATI 6        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1579  DLTAB5558   MANATI 7        FRONTERA SUR  TABASCO     JONUTA  Chilapa
1580  DLTAB5559   MANATI 8        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1581  DLTAB5560   MANATI 9        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1582  DLTAB5561  MANATI 10        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1583  DLTAB5562  MANATI 11        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1584  DLTAB5563  MANATI 12        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1585  DLTAB5564  MANATI 13        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1586  DLTAB5565  MANATI 14        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa

     CUERPO DE AGUA    TIPO SUBTIPO  LONGITUD   LATITUD  PERIODO DBO_mg/L  \
1573            NaN  LOTICO     NaN -92.33349  18.06207   2020.0       <2
1574            NaN  LOTICO     NaN -92.33424  18.07569   2020.0       <2
1575            NaN  LOTICO     NaN -92.30066  18.09384   2020.0       <2
1576            NaN  LOTICO     NaN -92.30488  18.09933   2020.0       <2
1577            NaN  LOTICO     NaN -92.35835  18.02775   2020.0       <2
1578            NaN  LOTICO     NaN -92.35786  18.03186   2020.0       <2
1579            NaN  LOTICO     NaN -92.31493  17.95932   2020.0       <2
1580            NaN  LOTICO     NaN -92.32552  17.80520   2020.0       <2
1581            NaN  LOTICO     NaN -92.31082  18.00212   2020.0       <2
1582            NaN  LOTICO     NaN -92.30760  17.92084   2020.0       <2
1583            NaN  LOTICO     NaN -92.23665  17.86673   2020.0       <2
1584            NaN  LOTICO     NaN -92.25317  17.83635   2020.0       <2
1585            NaN  LOTICO     NaN -92.39570  18.03142   2020.0       <2
1586            NaN  LOTICO     NaN -92.32300  18.06031   2020.0       <2
```

	CALIDAD_DBO	DQO_mg/L	CALIDAD_DQO	SST_mg/L	CALIDAD_SST
1573	Excelente	<10	Excelente	<10	Excelente
1574	Excelente	<10	Excelente	<10	Excelente
1575	Excelente	12.54	Buena calidad	<10	Excelente
1576	Excelente	28.42	Aceptable	<10	Excelente
1577	Excelente	<10	Excelente	<10	Excelente
1578	Excelente	13.38	Buena calidad	44.4	Buena calidad
1579	Excelente	21.74	Aceptable	<10	Excelente
1580	Excelente	<10	Excelente	33.13	Buena calidad
1581	Excelente	20.06	Aceptable	<10	Excelente
1582	Excelente	<10	Excelente	<10	Excelente
1583	Excelente	30.18	Aceptable	<10	Excelente
1584	Excelente	14.21	Buena calidad	<10	Excelente
1585	Excelente	<10	Excelente	49.3	Buena calidad
1586	Excelente	<10	Excelente	<10	Excelente

	COLI_FEC_NMP_100mL	CALIDAD_COLI_FEC	E_COLI_NMP_100mL	CALIDAD_E_COLI
1573	332	Aceptable	20	Excelente
1574	241	Aceptable	52	Excelente
1575	432	Aceptable	86	Excelente
1576	1086	Contaminada	203	Buena calidad
1577	216	Aceptable	85	Excelente
1578	359	Aceptable	145	Buena calidad
1579	1872	Contaminada	833	Aceptable
1580	1198	Contaminada	728	Aceptable
1581	1274	Contaminada	161	Buena calidad
1582	1872	Contaminada	457	Buena calidad
1583	538	Aceptable	109	Excelente
1584	670	Aceptable	259	Buena calidad
1585	305	Aceptable	145	Buena calidad
1586	728	Aceptable	109	Excelente

	ENTEROC_NMP_100mL	CALIDAD_ENTEROC	OD_PORC	CALIDAD_OD_PORC
1573	NaN	NaN	NaN	NaN
1574	NaN	NaN	NaN	NaN
1575	NaN	NaN	NaN	NaN
1576	NaN	NaN	<10	Fuertemente contaminada
1577	NaN	NaN	48.7	Aceptable
1578	NaN	NaN	53.1	Buena calidad
1579	NaN	NaN	<10	Fuertemente contaminada
1580	NaN	NaN	67	Buena calidad
1581	NaN	NaN	NaN	NaN
1582	NaN	NaN	NaN	NaN
1583	NaN	NaN	<10	Fuertemente contaminada
1584	NaN	NaN	36.1	Aceptable
1585	NaN	NaN	51.8	Buena calidad
1586	NaN	NaN	NaN	NaN

	OD_PORC_SUP	CALIDAD_OD_PORC_SUP	OD_PORC_MED	CALIDAD_OD_PORC_MED
1573	NaN	NaN	NaN	NaN
1574	NaN	NaN	NaN	NaN
1575	NaN	NaN	NaN	NaN
1576	NaN	NaN	NaN	NaN
1577	NaN	NaN	NaN	NaN
1578	NaN	NaN	NaN	NaN
1579	NaN	NaN	NaN	NaN
1580	NaN	NaN	NaN	NaN
1581	NaN	NaN	NaN	NaN
1582	NaN	NaN	NaN	NaN
1583	NaN	NaN	NaN	NaN
1584	NaN	NaN	NaN	NaN
1585	NaN	NaN	NaN	NaN
1586	NaN	NaN	NaN	NaN

	OD_PORC_FON	CALIDAD_OD_PORC_FON	TOX_D_48_UT	CALIDAD_TOX_D_48	TOX_V_15_UT
1573	NaN	NaN	<1	No Toxico	<1
1574	NaN	NaN	<1	No Toxico	<1
1575	NaN	NaN	<1	No Toxico	<1
1576	NaN	NaN	<1	No Toxico	<1
1577	NaN	NaN	<1	No Toxico	<1
1578	NaN	NaN	<1	No Toxico	<1
1579	NaN	NaN	<1	No Toxico	<1
1580	NaN	NaN	<1	No Toxico	<1
1581	NaN	NaN	<1	No Toxico	<1
1582	NaN	NaN	<1	No Toxico	<1
1583	NaN	NaN	<1	No Toxico	<1
1584	NaN	NaN	<1	No Toxico	<1
1585	NaN	NaN	<1	No Toxico	<1
1586	NaN	NaN	<1	No Toxico	<1

	CALIDAD_TOX_V_15	TOX_D_48_SUP_UT	CALIDAD TOX_D_48_SUP	TOX_D_48_FON_UT
1573	No Toxico	NaN	NaN	NaN
1574	No Toxico	NaN	NaN	NaN
1575	No Toxico	NaN	NaN	NaN
1576	No Toxico	NaN	NaN	NaN
1577	No Toxico	NaN	NaN	NaN
1578	No Toxico	NaN	NaN	NaN
1579	No Toxico	NaN	NaN	NaN
1580	No Toxico	NaN	NaN	NaN
1581	No Toxico	NaN	NaN	NaN
1582	No Toxico	NaN	NaN	NaN
1583	No Toxico	NaN	NaN	NaN
1584	No Toxico	NaN	NaN	NaN
1585	No Toxico	NaN	NaN	NaN

```
1586       No Toxico             NaN                  NaN             NaN

      CALIDAD_TOX_D_48_FON TOX_FIS_SUP_15_UT CALIDAD_TOX_FIS_SUP_15  \
1573                  NaN               NaN                    NaN
1574                  NaN               NaN                    NaN
1575                  NaN               NaN                    NaN
1576                  NaN               NaN                    NaN
1577                  NaN               NaN                    NaN
1578                  NaN               NaN                    NaN
1579                  NaN               NaN                    NaN
1580                  NaN               NaN                    NaN
1581                  NaN               NaN                    NaN
1582                  NaN               NaN                    NaN
1583                  NaN               NaN                    NaN
1584                  NaN               NaN                    NaN
1585                  NaN               NaN                    NaN
1586                  NaN               NaN                    NaN

      TOX_FIS_FON_15_UT  CALIDAD_TOX_FIS_FON_15  SEMAFORO CONTAMINANTES  \
1573               NaN                     NaN     Verde           NaN
1574               NaN                     NaN     Verde           NaN
1575               NaN                     NaN     Verde           NaN
1576               NaN                     NaN  Amarillo      CF,OD%L,
1577               NaN                     NaN     Verde           NaN
1578               NaN                     NaN     Verde           NaN
1579               NaN                     NaN  Amarillo      CF,OD%L,
1580               NaN                     NaN  Amarillo           CF,
1581               NaN                     NaN  Amarillo           CF,
1582               NaN                     NaN  Amarillo           CF,
1583               NaN                     NaN  Amarillo         OD%L,
1584               NaN                     NaN     Verde           NaN
1585               NaN                     NaN     Verde           NaN
1586               NaN                     NaN     Verde           NaN

     CUMPLE_CON_DBO CUMPLE_CON_DQO CUMPLE_CON_SST CUMPLE_CON_CF  \
1573             SI             SI              SI            SI
1574             SI             SI              SI            SI
1575             SI             SI              SI            SI
1576             SI             SI              SI            NO
1577             SI             SI              SI            SI
1578             SI             SI              SI            SI
1579             SI             SI              SI            NO
1580             SI             SI              SI            NO
1581             SI             SI              SI            NO
1582             SI             SI              SI            NO
1583             SI             SI              SI            SI
1584             SI             SI              SI            SI
```

```
1585             SI             SI             SI            SI
1586             SI             SI             SI            SI

     CUMPLE_CON_E_COLI CUMPLE_CON_ENTEROC CUMPLE_CON_OD CUMPLE_CON_TOX   GRUPO
1573                SI                 ND            ND             SI  LOTICO
1574                SI                 ND            ND             SI  LOTICO
1575                SI                 ND            ND             SI  LOTICO
1576                SI                 ND            NO             SI  LOTICO
1577                SI                 ND            SI             SI  LOTICO
1578                SI                 ND            SI             SI  LOTICO
1579                SI                 ND            NO             SI  LOTICO
1580                SI                 ND            SI             SI  LOTICO
1581                SI                 ND            ND             SI  LOTICO
1582                SI                 ND            ND             SI  LOTICO
1583                SI                 ND            NO             SI  LOTICO
1584                SI                 ND            SI             SI  LOTICO
1585                SI                 ND            SI             SI  LOTICO
1586                SI                 ND            ND             SI  LOTICO
```

[19]:
```
array(['PRESA EL SAUCILLO', 'OCEANO PACIFICO', 'BAHIA SAN LUCAS', …,
       'PRESA MARTE R. GOMEZ', 'PRESA INTERNACIONAL FALCON', 'EL LAGUITO'],
      dtype=object)
```

Podemos observar que la columna *CUERPO DE AGUA* hace referencia al nombre de pila, y que los valores faltantes corresponden a un único cuerpo que no tiene el nombre definido. Por tanto, colocaremos en todos ellos un nombre genérico *Desconocido*, para evitar la presencia de valores nulos.

1.3.3 Subtipo

[21]:
```
          CLAVE      SITIO ORGANISMO_DE_CUENCA   ESTADO  MUNICIPIO   CUENCA  \
1573  DLTAB5552   MANATI 1        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1574  DLTAB5553   MANATI 2        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1575  DLTAB5554   MANATI 3        FRONTERA SUR  TABASCO     JONUTA  Chilapa
1576  DLTAB5555   MANATI 4        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1577  DLTAB5556   MANATI 5        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1578  DLTAB5557   MANATI 6        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1579  DLTAB5558   MANATI 7        FRONTERA SUR  TABASCO     JONUTA  Chilapa
1580  DLTAB5559   MANATI 8        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1581  DLTAB5560   MANATI 9        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1582  DLTAB5561  MANATI 10        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1583  DLTAB5562  MANATI 11        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1584  DLTAB5563  MANATI 12        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1585  DLTAB5564  MANATI 13        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa
1586  DLTAB5565  MANATI 14        FRONTERA SUR  TABASCO  MACUSPANA  Chilapa

     CUERPO DE AGUA    TIPO SUBTIPO  LONGITUD   LATITUD  PERIODO DBO_mg/L  \
1573            NaN  LOTICO     NaN -92.33349  18.06207   2020.0       <2
```

```
1574           NaN  LOTICO     NaN -92.33424  18.07569   2020.0       <2
1575           NaN  LOTICO     NaN -92.30066  18.09384   2020.0       <2
1576           NaN  LOTICO     NaN -92.30488  18.09933   2020.0       <2
1577           NaN  LOTICO     NaN -92.35835  18.02775   2020.0       <2
1578           NaN  LOTICO     NaN -92.35786  18.03186   2020.0       <2
1579           NaN  LOTICO     NaN -92.31493  17.95932   2020.0       <2
1580           NaN  LOTICO     NaN -92.32552  17.80520   2020.0       <2
1581           NaN  LOTICO     NaN -92.31082  18.00212   2020.0       <2
1582           NaN  LOTICO     NaN -92.30760  17.92084   2020.0       <2
1583           NaN  LOTICO     NaN -92.23665  17.86673   2020.0       <2
1584           NaN  LOTICO     NaN -92.25317  17.83635   2020.0       <2
1585           NaN  LOTICO     NaN -92.39570  18.03142   2020.0       <2
1586           NaN  LOTICO     NaN -92.32300  18.06031   2020.0       <2

     CALIDAD_DBO DQO_mg/L    CALIDAD_DQO SST_mg/L    CALIDAD_SST  \
1573   Excelente      <10      Excelente      <10      Excelente
1574   Excelente      <10      Excelente      <10      Excelente
1575   Excelente    12.54  Buena calidad      <10      Excelente
1576   Excelente    28.42      Aceptable      <10      Excelente
1577   Excelente      <10      Excelente      <10      Excelente
1578   Excelente    13.38  Buena calidad     44.4  Buena calidad
1579   Excelente    21.74      Aceptable      <10      Excelente
1580   Excelente      <10      Excelente    33.13  Buena calidad
1581   Excelente    20.06      Aceptable      <10      Excelente
1582   Excelente      <10      Excelente      <10      Excelente
1583   Excelente    30.18      Aceptable      <10      Excelente
1584   Excelente    14.21  Buena calidad      <10      Excelente
1585   Excelente      <10      Excelente     49.3  Buena calidad
1586   Excelente      <10      Excelente      <10      Excelente

     COLI_FEC_NMP_100mL CALIDAD_COLI_FEC E_COLI_NMP_100mL CALIDAD_E_COLI  \
1573                332        Aceptable               20      Excelente
1574                241        Aceptable               52      Excelente
1575                432        Aceptable               86      Excelente
1576               1086      Contaminada              203  Buena calidad
1577                216        Aceptable               85      Excelente
1578                359        Aceptable              145  Buena calidad
1579               1872      Contaminada              833      Aceptable
1580               1198      Contaminada              728      Aceptable
1581               1274      Contaminada              161  Buena calidad
1582               1872      Contaminada              457  Buena calidad
1583                538        Aceptable              109      Excelente
1584                670        Aceptable              259  Buena calidad
1585                305        Aceptable              145  Buena calidad
1586                728        Aceptable              109      Excelente

     ENTEROC_NMP_100mL CALIDAD_ENTEROC OD_PORC          CALIDAD_OD_PORC  \
```

```
1573              NaN             NaN     NaN                       NaN
1574              NaN             NaN     NaN                       NaN
1575              NaN             NaN     NaN                       NaN
1576              NaN             NaN     <10  Fuertemente contaminada
1577              NaN             NaN    48.7                Aceptable
1578              NaN             NaN    53.1            Buena calidad
1579              NaN             NaN     <10  Fuertemente contaminada
1580              NaN             NaN      67            Buena calidad
1581              NaN             NaN     NaN                       NaN
1582              NaN             NaN     NaN                       NaN
1583              NaN             NaN     <10  Fuertemente contaminada
1584              NaN             NaN    36.1                Aceptable
1585              NaN             NaN    51.8            Buena calidad
1586              NaN             NaN     NaN                       NaN

     OD_PORC_SUP CALIDAD_OD_PORC_SUP OD_PORC_MED CALIDAD_OD_PORC_MED  \
1573         NaN                 NaN         NaN                 NaN
1574         NaN                 NaN         NaN                 NaN
1575         NaN                 NaN         NaN                 NaN
1576         NaN                 NaN         NaN                 NaN
1577         NaN                 NaN         NaN                 NaN
1578         NaN                 NaN         NaN                 NaN
1579         NaN                 NaN         NaN                 NaN
1580         NaN                 NaN         NaN                 NaN
1581         NaN                 NaN         NaN                 NaN
1582         NaN                 NaN         NaN                 NaN
1583         NaN                 NaN         NaN                 NaN
1584         NaN                 NaN         NaN                 NaN
1585         NaN                 NaN         NaN                 NaN
1586         NaN                 NaN         NaN                 NaN

     OD_PORC_FON CALIDAD_OD_PORC_FON TOX_D_48_UT CALIDAD_TOX_D_48 TOX_V_15_UT  \
1573         NaN                 NaN          <1        No Toxico          <1
1574         NaN                 NaN          <1        No Toxico          <1
1575         NaN                 NaN          <1        No Toxico          <1
1576         NaN                 NaN          <1        No Toxico          <1
1577         NaN                 NaN          <1        No Toxico          <1
1578         NaN                 NaN          <1        No Toxico          <1
1579         NaN                 NaN          <1        No Toxico          <1
1580         NaN                 NaN          <1        No Toxico          <1
1581         NaN                 NaN          <1        No Toxico          <1
1582         NaN                 NaN          <1        No Toxico          <1
1583         NaN                 NaN          <1        No Toxico          <1
1584         NaN                 NaN          <1        No Toxico          <1
1585         NaN                 NaN          <1        No Toxico          <1
1586         NaN                 NaN          <1        No Toxico          <1
```

	CALIDAD_TOX_V_15	TOX_D_48_SUP_UT	CALIDAD TOX_D_48_SUP	TOX_D_48_FON_UT	\
1573	No Toxico	NaN	NaN	NaN	
1574	No Toxico	NaN	NaN	NaN	
1575	No Toxico	NaN	NaN	NaN	
1576	No Toxico	NaN	NaN	NaN	
1577	No Toxico	NaN	NaN	NaN	
1578	No Toxico	NaN	NaN	NaN	
1579	No Toxico	NaN	NaN	NaN	
1580	No Toxico	NaN	NaN	NaN	
1581	No Toxico	NaN	NaN	NaN	
1582	No Toxico	NaN	NaN	NaN	
1583	No Toxico	NaN	NaN	NaN	
1584	No Toxico	NaN	NaN	NaN	
1585	No Toxico	NaN	NaN	NaN	
1586	No Toxico	NaN	NaN	NaN	

	CALIDAD_TOX_D_48_FON	TOX_FIS_SUP_15_UT	CALIDAD_TOX_FIS_SUP_15	\
1573	NaN	NaN	NaN	
1574	NaN	NaN	NaN	
1575	NaN	NaN	NaN	
1576	NaN	NaN	NaN	
1577	NaN	NaN	NaN	
1578	NaN	NaN	NaN	
1579	NaN	NaN	NaN	
1580	NaN	NaN	NaN	
1581	NaN	NaN	NaN	
1582	NaN	NaN	NaN	
1583	NaN	NaN	NaN	
1584	NaN	NaN	NaN	
1585	NaN	NaN	NaN	
1586	NaN	NaN	NaN	

	TOX_FIS_FON_15_UT	CALIDAD_TOX_FIS_FON_15	SEMAFORO	CONTAMINANTES	\
1573	NaN	NaN	Verde	NaN	
1574	NaN	NaN	Verde	NaN	
1575	NaN	NaN	Verde	NaN	
1576	NaN	NaN	Amarillo	CF,OD%L,	
1577	NaN	NaN	Verde	NaN	
1578	NaN	NaN	Verde	NaN	
1579	NaN	NaN	Amarillo	CF,OD%L,	
1580	NaN	NaN	Amarillo	CF,	
1581	NaN	NaN	Amarillo	CF,	
1582	NaN	NaN	Amarillo	CF,	
1583	NaN	NaN	Amarillo	OD%L,	
1584	NaN	NaN	Verde	NaN	
1585	NaN	NaN	Verde	NaN	
1586	NaN	NaN	Verde	NaN	

	CUMPLE_CON_DBO	CUMPLE_CON_DQO	CUMPLE_CON_SST	CUMPLE_CON_CF	\
1573	SI	SI	SI	SI	
1574	SI	SI	SI	SI	
1575	SI	SI	SI	SI	
1576	SI	SI	SI	NO	
1577	SI	SI	SI	SI	
1578	SI	SI	SI	SI	
1579	SI	SI	SI	NO	
1580	SI	SI	SI	NO	
1581	SI	SI	SI	NO	
1582	SI	SI	SI	NO	
1583	SI	SI	SI	SI	
1584	SI	SI	SI	SI	
1585	SI	SI	SI	SI	
1586	SI	SI	SI	SI	

	CUMPLE_CON_E_COLI	CUMPLE_CON_ENTEROC	CUMPLE_CON_OD	CUMPLE_CON_TOX	GRUPO
1573	SI	ND	ND	SI	LOTICO
1574	SI	ND	ND	SI	LOTICO
1575	SI	ND	ND	SI	LOTICO
1576	SI	ND	NO	SI	LOTICO
1577	SI	ND	SI	SI	LOTICO
1578	SI	ND	SI	SI	LOTICO
1579	SI	ND	NO	SI	LOTICO
1580	SI	ND	SI	SI	LOTICO
1581	SI	ND	ND	SI	LOTICO
1582	SI	ND	ND	SI	LOTICO
1583	SI	ND	NO	SI	LOTICO
1584	SI	ND	SI	SI	LOTICO
1585	SI	ND	SI	SI	LOTICO
1586	SI	ND	ND	SI	LOTICO

Observamos que son los mismos registros que tienen el *CUERPO DE AGUA* faltante, de forma que procedemos a agregar también un *Desconocido*.

1.3.4 DBO_mg/L

[23]:
```
array(['6', nan, '<2', '4.26', '6.4', '4.98', '5.46', '7.1', '4.66',
       '5.3', '5.73', '7.68', '8.06', '11.06', '18.86', '33.42', '27.42',
       '30.66', '9.66', '4.74', '4.77', '3.63', '5.1', '8.49', '4.32',
       '10.14', '12.21', '4.14', '5.52', '3.27', '11.85', '49.6', '6.42',
       '42.9', '127.75', '9.2', '35.5', '4.44', '20.9', '5.04', '5.56',
       '10.7', '5.14', '20.5', '8', '7.94', '45.1', '34.95', '7.58',
       '7.92', '8.9', '5.84', '7.08', '10.1', '5.9', '12.8', '13.2',
       '9.8', '6.14', '9.86', '8.2', '19.5', '6.48', '8.76', '22.2',
       '4.28', '2.34', '4.38', '10.3', '7.82', '4.3', '4.54', '10.2',
```

```
'6.76', '37.2', '4.02', '4.06', '4.52', '5.44', '7.7', '5.02',
'24.61', '45.3', '42.3', '27.9', '54.02', '42.01', '59.27',
'75.31', '45.6', '23.1', '56.12', '63.47', '7.5', '31.5', '25.36',
'26.11', '23.4', '12', '57.47', '7.2', '22.6', '32.7', '11.1',
'10.6', '3.18', '7.66', '4.48', '6.96', '2.6', '11.63', '89.71',
'588.59', '36.5', '61.5', '111.25', '38.66', '4.68', '8.68',
'20.4', '4.36', '38', '34', '29.7', '16.51', '32.72', '9', '5',
'75', '136', '57', '219', '7.44', '4.47', '3.06', '21.3', '24',
'35', '13', '23', '33', '22', '19', '41', '28', '10', '15', '62',
'31', '30', '143', '119', '256', '239', '151', '1015', '915',
'1500', '474', '13.3', '580', '20.3', '15.6', '48', '53.25', '635',
'735', '73.25', '9.72', '15.2', '98.25', '6.02', '5.98', '54.5',
'229.75', '56', '14', '11', '17', '16', '18', '20', '46', '26.15',
'368', '37.1', '6.6', '80', '3.67', '2.75', '2.61', '2.38', '3.8',
'4.35', '4.55', '2.96', '5.58', '5.4', '2.2', '3.4', '2', '2.14',
'2.43', '2.29', '6.2', '3.57', '4.33', '2.13', '4.18', '6.08',
'4.88', '22.7', '7.18', '21.76', '28.07', '24.16', '28.66',
'27.91', '31.2', '65.54', '5.91', '20.66', '26.35', '19.51',
'17.57', '3.72', '3.66', '5.22', '8.85', '15.45', '4.41', '7.35',
'6.45', '34.32', '2.57', '5.55', '4.05', '7.8', '7', '10.53',
'9.36', '28.23', '12.6', '28.35', '9.9', '93.31', '2.22', '7.34',
'4', '407', '59', '23.6', '157.75', '4.58', '30.3', '36.7', '4.9',
'6.51', '2.24', '6.7', '23.5', '19.2', '15.9', '6.3', '16.6',
'3.78', '6.27', '13.6', '6.9', '12.1', '10.9', '16.9', '13.9',
'22.8', '12.3', '21', '15.4', '4.96', '232.25', '890', '29', '25',
'26', '42', '64', '92', '109', '6.34', '61', '287', '159', '98',
'4.92', '4.1', '7.22', '4.16', '3.58', '5.16', '4.72', '8.6',
'5.8', '7.42', '6.56', '2.72', '60', '155', '250', '145', '260',
'66', '90', '130', '240', '50', '26.67', '70', '2.86', '3.33',
'2.67', '3.53', '3.11', '3.5', '3.29', '5.2', '8.67', '58',
'33.33', '16.5', '7.33', '5.33', '32', '11.67', '9.67', '3.17',
'4.83', '350', '27', '125', '43.33', '46.67', '10.33', '150', '44',
'500', '3.43', '16.67', '21.33', '3.38', '3.04', '2.54', '61.75',
'156', '91.25', '87', '895', '121', '127', '346', '366', '208',
'114', '213', '51', '202', '39', '47', '4.2', '6.54', '17.31',
'15.69', '8.25', '21.72', '12.66', '13.01', '14.82', '14.01',
'15.53', '16.23', '15.03', '9.87', '9.54', '10.23', '8.55', '8.1',
'11.28', '10.95', '5.37', '4.34', '12.9', '5.42', '4.99', '2.31',
'5.08', '2.66', '5.64', '9.27', '3', '3.42', '7.32', '5.25',
'2.07', '22.05', '3.54', '2.58', '13.5', '3.12', '86.67', '83.33',
'12.5', '18.67', '11.5', '73.33', '2.63', '4.4', '3.6', '7.75',
'56.67', '53.33', '3.45', '880', '76', '91.5', '25.3', '34.6',
'4.8', '7.49', '86.5', '183.5', '49.42', '44.82', '138.5', '4.07',
'4.73', '5.49', '164.47', '5.09', '15.3', '27.3', '19.62', '164.5',
'32.11', '78.01', '57.32', '108.5', '2.88', '3.69', '4.59',
'126.48', '24.36', '39.9', '67.5', '20.1', '11.4', '15.76', '5.7',
'13.8', '13.7', '8.4', '14.58', '19.08', '7.04', '51.35', '40.9',
```

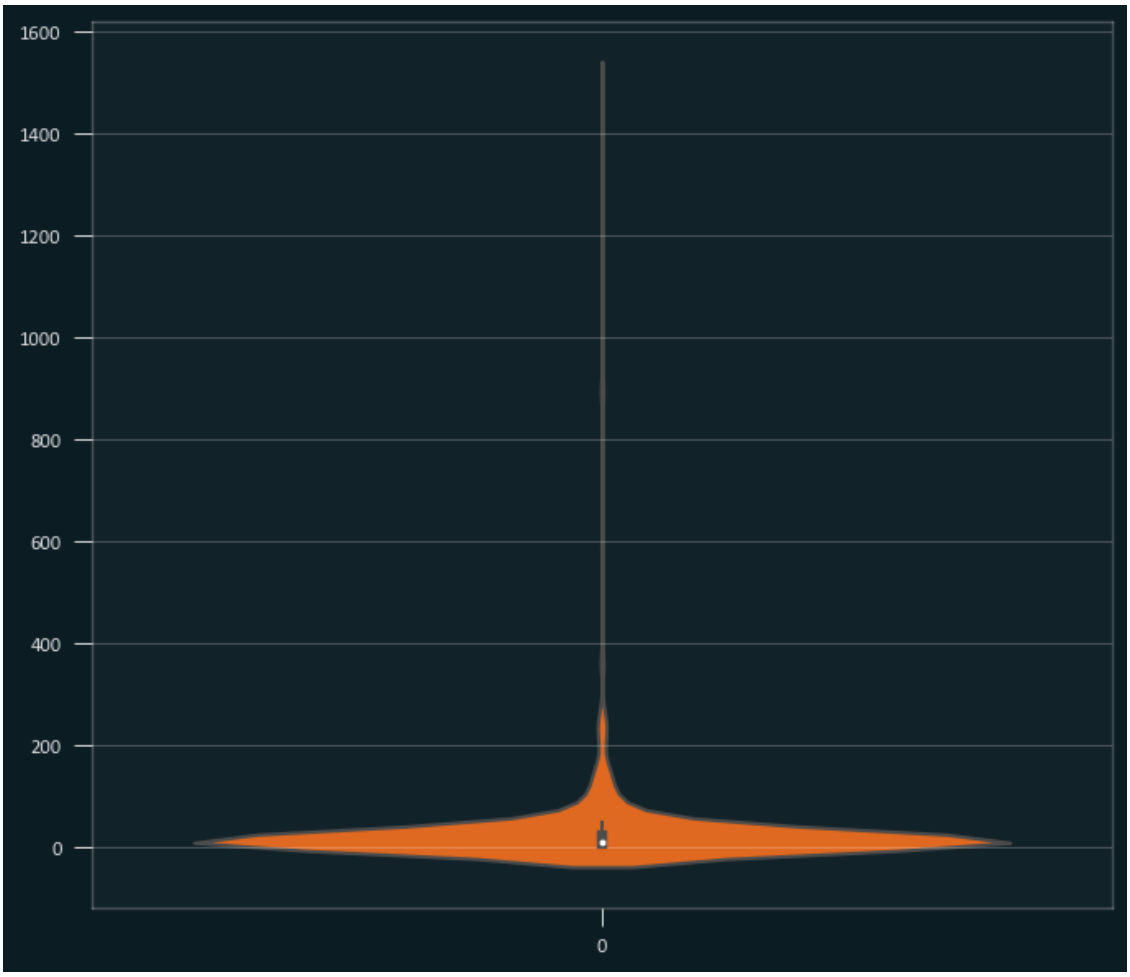
```
'6.95', '14.4', '18.3', '13.4', '18.8', '8.98', '17.7', '15.1',
'4.6', '11.8', '3.71', '3.14', '7.88', '21.2', '2.71', '20.8',
'2.85', '13.29', '2.7', '3.48', '14.1', '7.05', '2.82', '14.34',
'6.17', '11.76', '11.18', '6.84', '8.82', '5.94', '85.23', '16.21',
'8.63', '34.06', '78.64', '2.18', '5.21', '2.78', '11.73', '8.48',
'8.07', '8.01', '2.19', '16.81', '7.14', '13.82', '5.76', '6.36',
'2.79', '2.97', '10.04', '9.08', '11.64', '10.5', '4.62', '8.7',
'11.86', '13.02', '17.04', '21.5', '4.84', '6.04', '5.36', '13.18',
'60.92', '14.7', '2.46', '3.47', '6.67', '6.33', '3.2', '2.17',
'2.4', '9.95', '8.94', '3.93', '6.43', '2.28', '2.94', '2.74',
'2.95', '113.33', '7.91', '3.39', '200', '140', '19.33', '54',
'8.5', '4.67', '96.67', '4.43', '8.33', '10.32', '3.28', '3.68',
'11.37', '12.2', '2.33', '17.03', '6.63', '8.23', '16.87', '3.37',
'7.65', '5.17', '52', '63.33', '2.83', '15.56', '56.71', '2.26',
'2.41', '22.41', '3.35', '11.25', '26.57', '12.44', '11.51',
'10.86', '6.78', '9.58', '2.04', '3.05', '7.4', '14.23', '13.61',
'14.63', '3.08', '21.81', '5.86', '6.03', '21.51', '5.34', '14.8',
'5.83', '10.08', '22.87', '23.23', '22.15', '6.94', '2.9', '4.31',
'5.18', '4.93', '6.79', '5.72', '2.32', '11.43', '2.84', '3.81',
'6.8', '8.71', '3.97', '3.07', '16.27', '18.53', '12.7', '53.55',
'18.9', '18.5', '7.6', '46.5', '32.8', '11.9', '6.98', '6.82',
'6.52', '15.8', '19.1', '8.8', '26.3', '33.1', '12.4', '36',
'23.2', '8.16', '6.1', '19.4', '16.1', '8.66', '11.54', '34.5',
'42.5', '111.75', '4.08', '9.7', '10.4', '5.88', '6.5', '87.75',
'7.9', '9.3', '75.5', '44.5', '4.82', '10.8', '9.22', '4.12',
'52.5', '6.38', '49.5', '9.5', '33.2', '11.6', '113', '44.6', '81',
'38.7', '7.3', '76.5', '25.1', '50.9', '48.5', '9.62', '2.68',
'2.16', '26.6', '28.7', '26.7', '3.84', '3.62', '32.5', '30.7',
'2.06', '3.74', '2.52', '93.5', '2.5', '7.26', '4.42', '7.56',
'5.26', '35.6', '14.5', '2.8', '57.5', '60.5', '6.64', '45.9',
'21.6', '94', '45.5', '220', '170', '40', '2.25', '2.39', '8.88',
'3.347', '5.35', '10.68', '47.6', '7.86', '5.62', '81.25', '7.52',
'38.9', '7.28', '33.3', '6.97', '10.51', '3.94', '6.22', '6.28',
'23.8', '3.03', '5.53', '33.56', '25.47', '42.55', '13.04',
'23.73', '113.2', '39.66', '17.73', '13.79', '95.87', '23.77',
'2.55', '39.09'], dtype=object)
```

Podemos observar que son valores numéricos, excepto por los faltantes y una categoría denominada *<2*. Veamos qué tantos valores entran en esa categoría.

[24]:
```
0.47423479271600155
```

Casi la mitad de los datos entran en esa categoría, de forma que imputar un valor tomaría la máxima relevancia en la distribución de los datos. Podemos analizar cómo se distribuyen el resto de los datos.

[25]:
```
<AxesSubplot:>
```

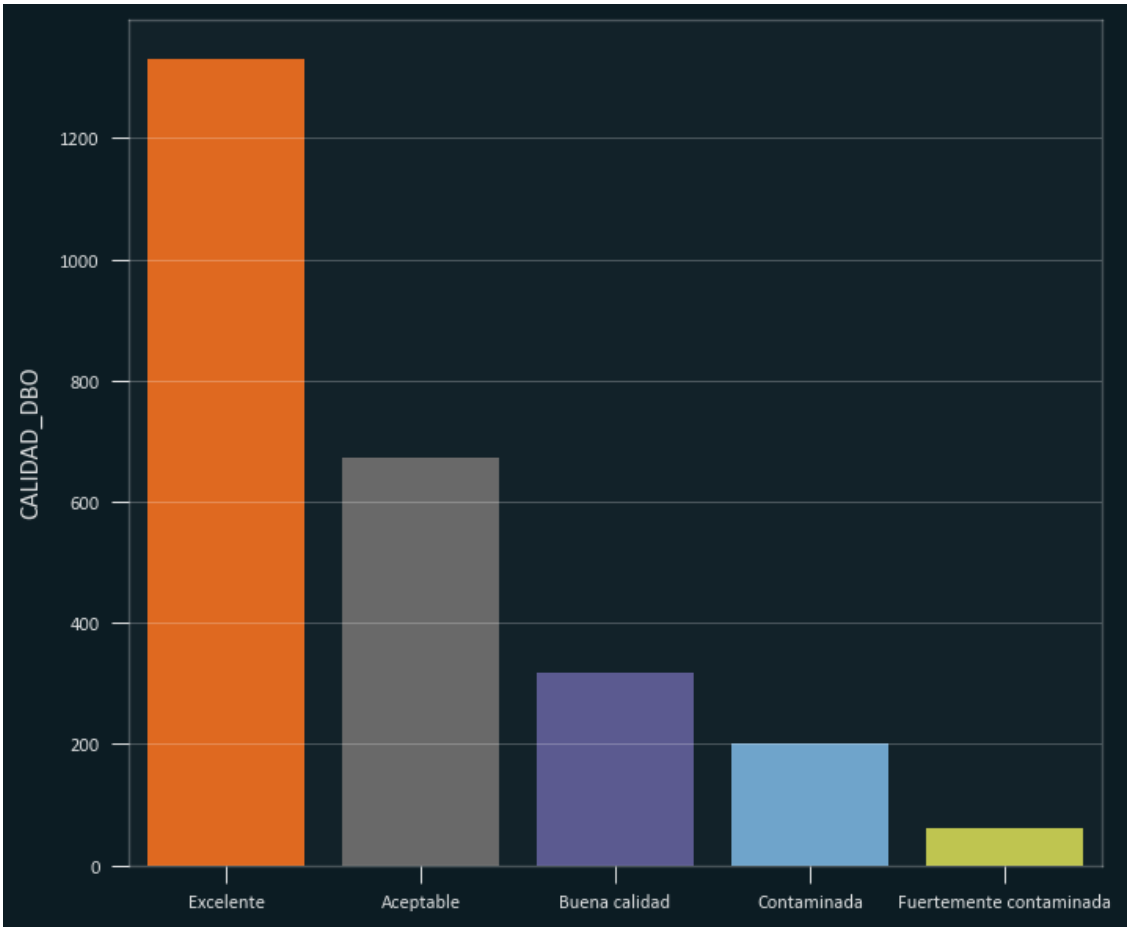


En este caso se puede observar que los valores tienen una distribución centrada en valores chicos (menores a 100), con algunos outliers que superan esos valores. Con eesto en mente, podemos imputar la categoría de *<2* a un único valor 0.

[27]:
```
count    2581.000000
mean       15.938011
std        65.364379
min         0.000000
25%         0.000000
50%         2.630000
75%        10.000000
max      1500.000000
Name: DBO_mg/L, dtype: float64
```

1.3.5 CALIDAD_DBO

[29]:
```
array(['Buena calidad', nan, 'Excelente', 'Aceptable', 'Contaminada',
       'Fuertemente contaminada'], dtype=object)
```



[31]:
```
count          2581
unique            5
top       Excelente
freq           1330
Name: CALIDAD_DBO, dtype: object
```

Se observa que la categoría predominante es *Excelente*, por lo que se puede hacer una imputación por moda.

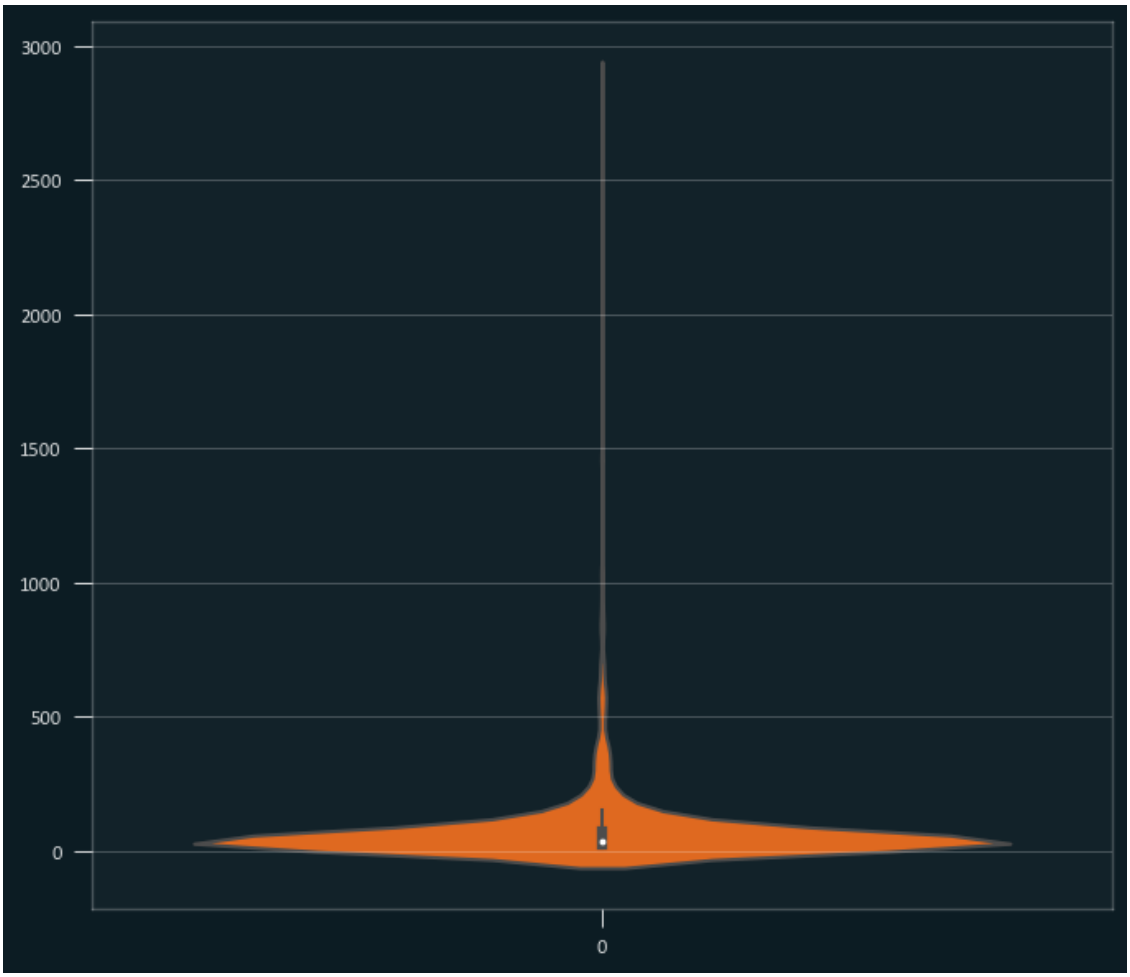
1.3.6 DQO_mg/L

[33]:
```
array(['54.08', nan, '<10', …, '115.88', '35.92', '34.6'], dtype=object)
```

Se observa un comportamiento similar al de DBO, donde hay una categoría marcada como "<10" y el resto tienen valores numéricos. Analizamos esos valores numéricos.

[34]: 0.21658271987601704

[35]: <AxesSubplot:>

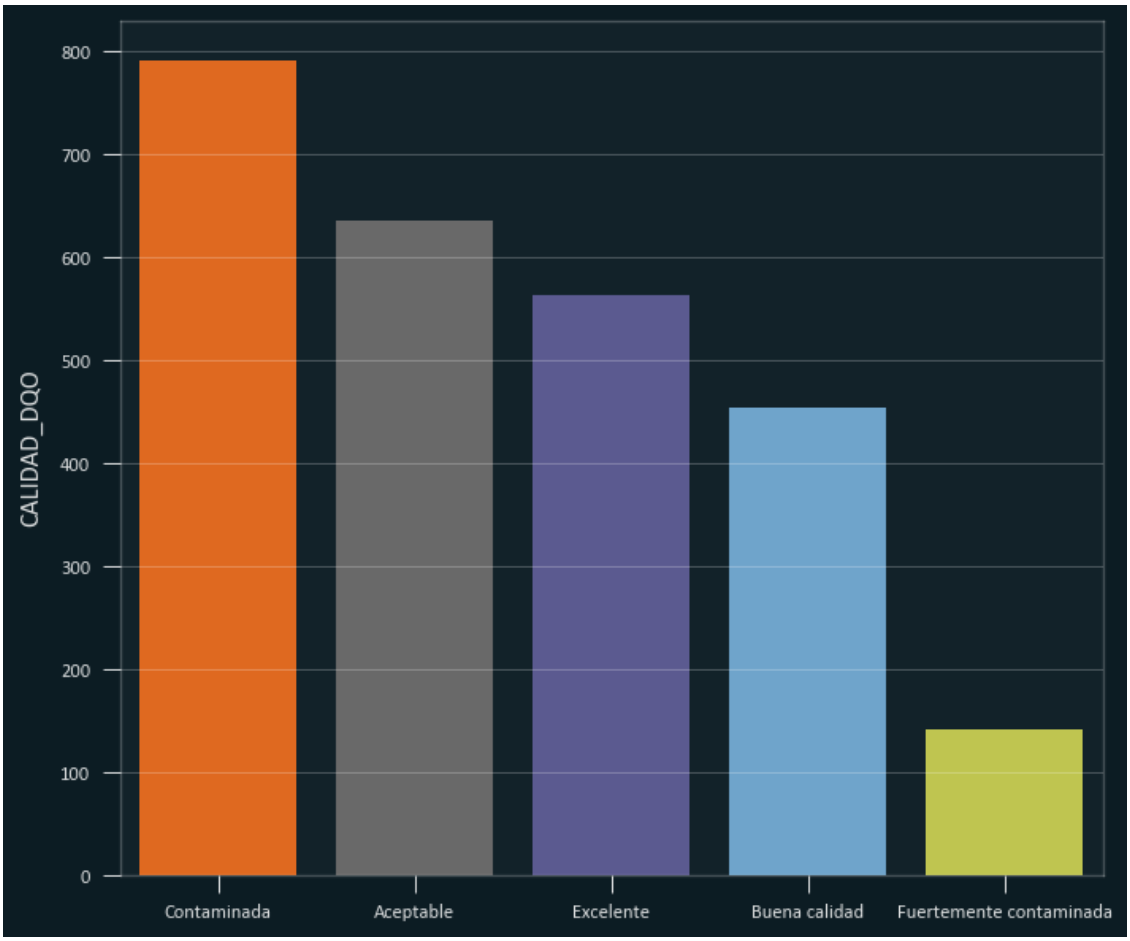


De manera simila, podemos seguir el mismo procedimiento que anteriormente con la variable DBO, usando la mediana para imputar los nulos, y colocar en 0 los de la categoría *<10*.

[36]:
```
count    2581.000000
mean       62.167157
std       150.668059
min         0.000000
25%        11.870000
50%        27.010000
75%        57.000000
max      2871.250000
Name: DQO_mg/L, dtype: float64
```

1.3.7 CALIDAD_DQO

[38]:
```
array(['Contaminada', nan, 'Excelente', 'Aceptable', 'Buena calidad',
       'Fuertemente contaminada'], dtype=object)
```

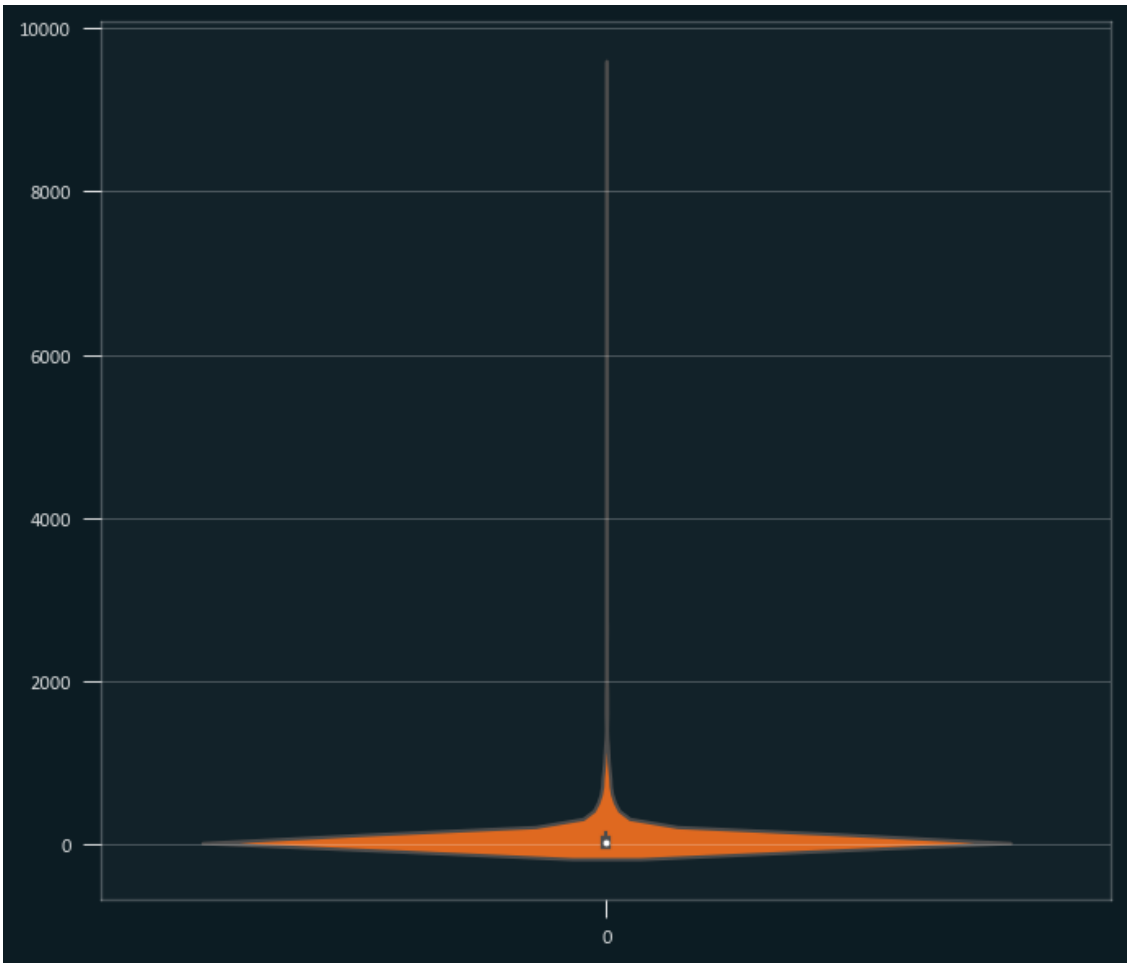


Hagamos una función que nos permita reproducir los métodos anteriormente realizados para las subsecuentes columnas, las cuales o serán numéricas o categóricas.

1.3.8 SST_mg/L

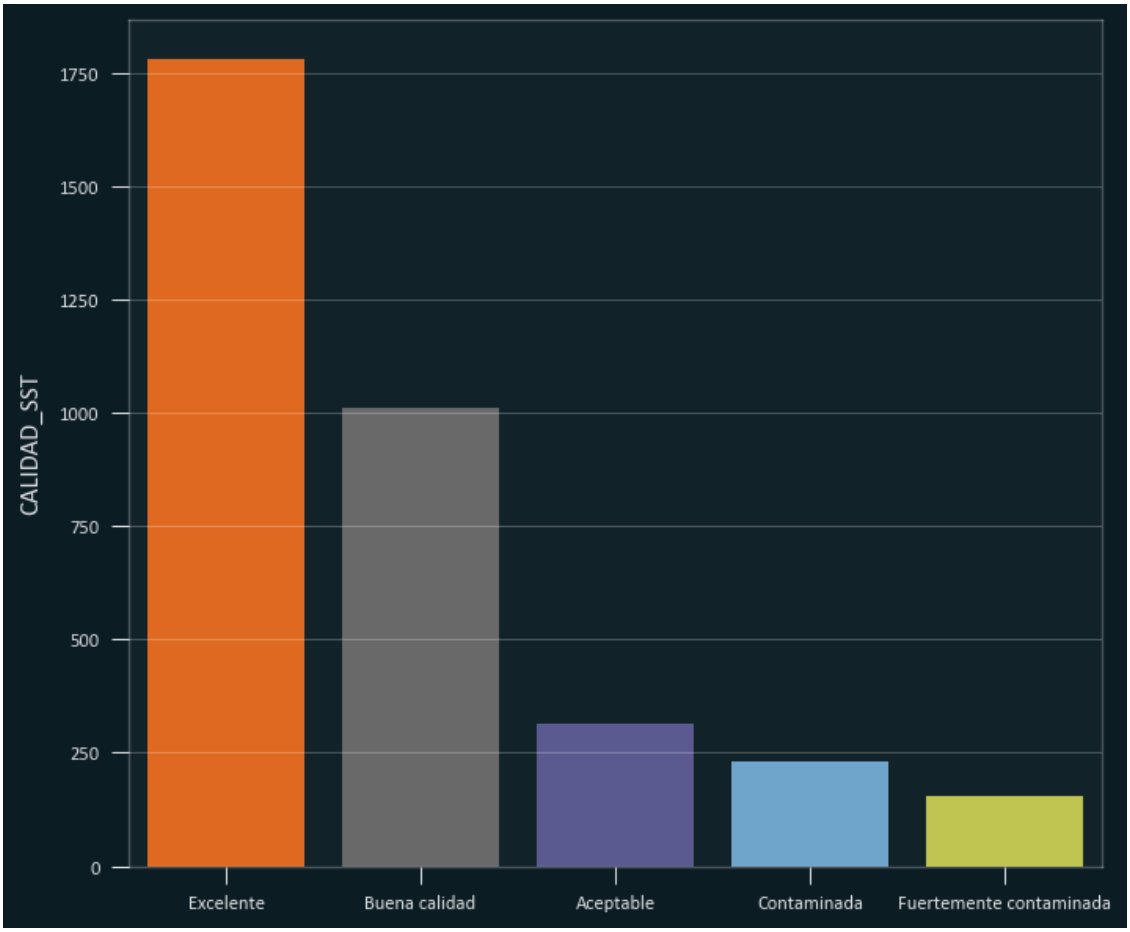
[42]:
```
array(['13.75', '<10', '13.9667', …, '137', '22.09', '119'],
      dtype=object)

Percentage of unique value to total: 0.2522212668386357
Values distribution: count    3489.000000
mean       99.625931
std       442.407559
min         0.000000
25%         0.000000
```

```
50%        24.300000
75%        57.000000
max      9430.000000
Name: SST_mg/L, dtype: float64
```



1.3.9 CALIDAD_SST

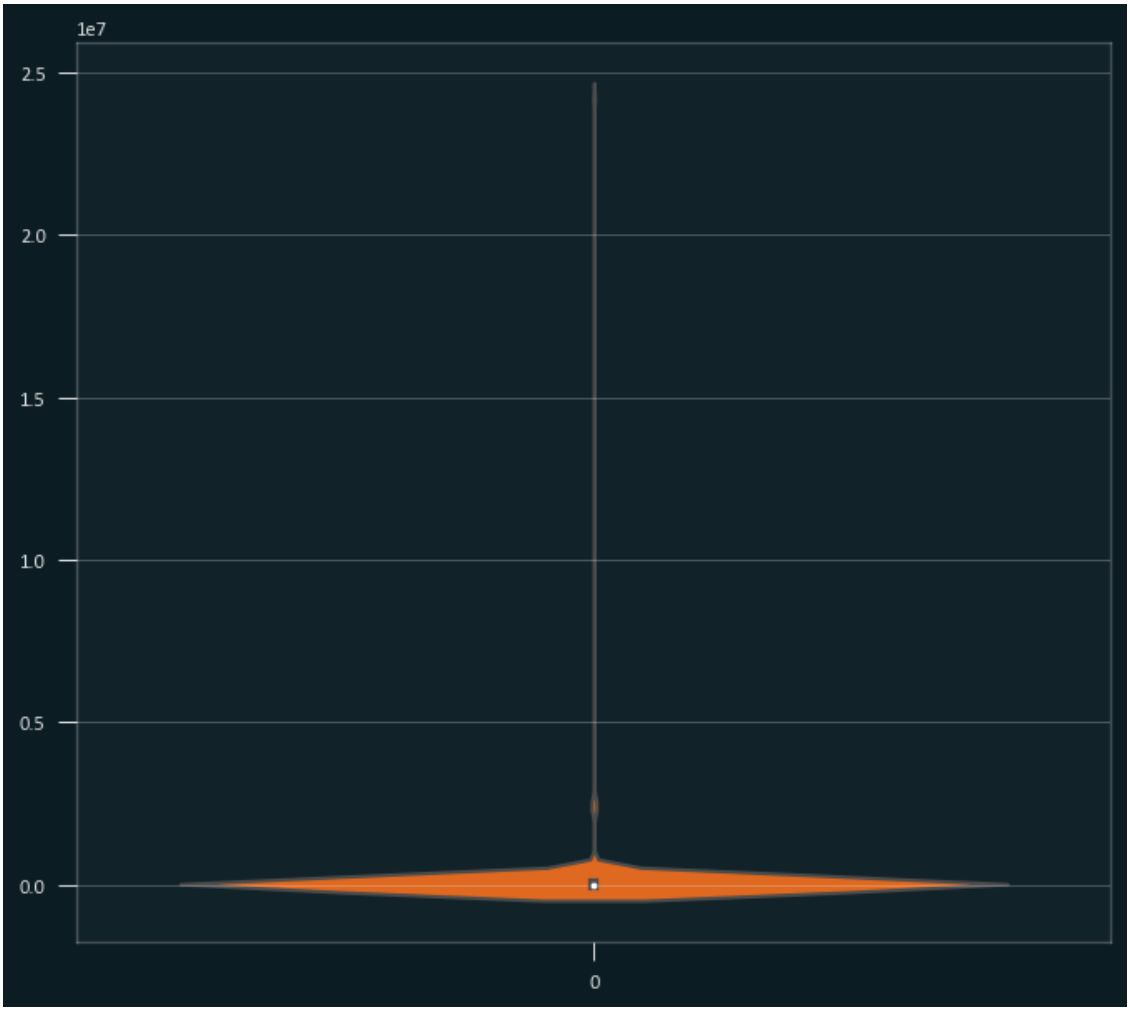
[44]:
```
array(['Excelente', 'Buena calidad', 'Contaminada', 'Aceptable', nan,
       'Fuertemente contaminada'], dtype=object)
```



1.3.10 COLI_FEC_NMP_100mL

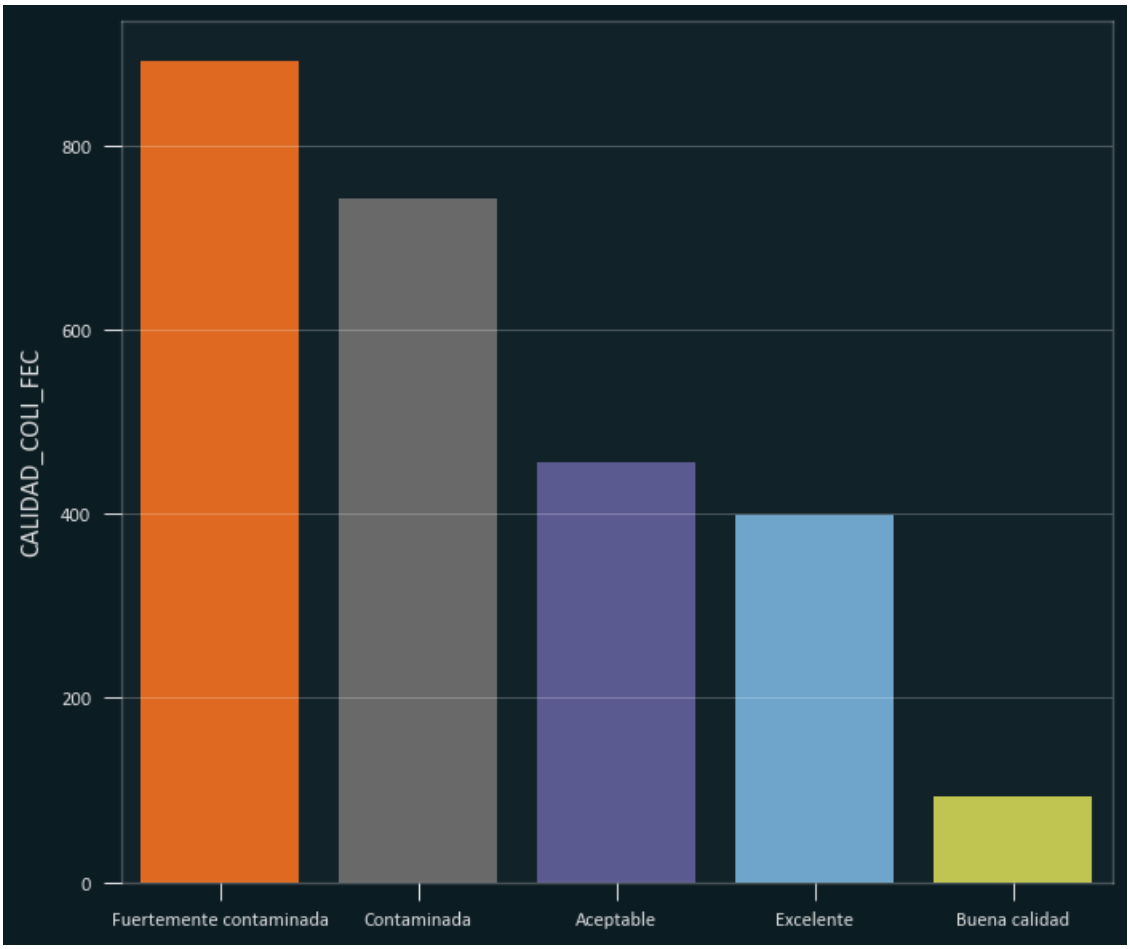
[46]:
```
array(['1162', nan, '3873', '189', '1408', '15531', '10', '24196', '218',
       '663', '14136', '1720', '2613', '495', '1650', '3255', '2481',
       '2046', '17329', '17863', '230', '430', '90', '2400', '70', '40',
       '210', '<3', '175', '241960', '15531000', '148', '52', '4611',
       '31', '364', '1523', '2755', '3448', '4884', '1439', '1450',
       '120330', '480', '3654', '9804', '109', '7270', '1421', '1553',
       '1872', '11199', '4280', '1793', '805', '1266', '1585', '1483',
       '2909', '749', '1430', '187', '41', '98', '183', '1172', '767',
       '216', '504', '20', '141360', '19863', '17230', '19560', '46110',
       '3180', '4950', '2247', '12100', '1220', '12033', '5475', '3076',
       '135', '722', '1229', '1860', '97', '512', '4352', '8164', '8664',
       '63', '448', '1090', '1198', '2014', '331', '1723', '1119900',
       '1725', '110', '132', '833', '75', '417', '785', '299', '471',
       '104', '5748', '122', '1918', '464', '1500', '280', '930', '11000',
       '200', '24000', '2100', '4600', '410', '677', '86', '9208', '5172',
       '2419600', '6488', '6131', '150', '140', '43', '140100', '16100',
```

```
       '792000', '74000', '24196000', '14136000', '1153', '5172000',
       '213', '203', '30', '933', '457', '374', '224700', '173290',
       '282000', '1100', '460', '23', '9', '15', '32550', '61310', '750',
       '960000', '6910', '161', '146', '243', '346', '325500', '1379',
       '441', '36540', '2987', '379', '6867', '5794', '384', '885', '231',
       '1956', '2143', '906', '359', '573', '27550', '11980', '24810',
       '1787', '57940', '41060', '104620', '155310', '4100', '30760',
       '12590', '109200', '19350', '9320', '10760', '34480', '1354',
       '17220', '1935', '6867000', '2359', '4', '39', '93', '7', '28',
       '2800', '240', '3', '14', '120', '21', '11620', '10462', '1664',
       '2098', '931', '350', '395', '684', '855', '1539', '246', '345',
       '341', '228', '602', '650', '1597', '1134', '1224', '1331', '1210',
       '1236', '12997', '86640', '4106', '548', '670', '1086', '762',
       '691', '1446', '4810', '48840', '22470', '26130', '4140', '1143',
       '594', '385', '830', '1081', '1607', '313', '158', '1119', '624',
       '295', '309', '390', '1200', '1025', '520', '1223', '1333', '1722',
       '1334', '17200', '521', '617', '1250', '988', '609', '404', '1658',
       '160', '332', '241', '432', '1274', '538', '305', '728', '98040',
       '81640', '1291', '92080', '54750', '991', '961', '1464', '2790',
       '905', '1350', '198630', '439', '327', '68670', '1309', '727',
       '479', '435', '1017', '576', '568', '744', '1376', '1401', '488',
       '712', '11', '21430', '7890', '85', '1054', '860', '38730', '8780',
       '233', '51720', '648880', '399', '557', '1071', '959', '14300',
       '18720', '279', '171', '852', '565', '676', '72700', '733',
       '14830', '539', '1098', '2187', '743', '1022', '738', '586', '74',
       '282', '355', '129970', '14500', '23590', '9600', '20460', '15390',
       '64880', '2489', '14140', '8050', '16570', '269', '1785', '256',
       '1058', '487', '1904', '73', '134', '826', '323', '77010',
       '613100', '13790', '288', '206', '794', '578', '426', '1789',
       '644', '882', '1076', '368', '7701', '420', '7940', '9106', '4790',
       '6500', '1160', '1529', '2723', '9090', '21870', '16580', '13310',
       '20140', '29090', '173', '1396', '15000', '934', '462', '960',
       '1565', '829', '1267', '1314', '990', '605', '880', '389'],
      dtype=object)

Percentage of unique value to total: 0.04221533694810225
Values distribution: count    2.582000e+03
mean     9.568868e+04
std      1.168887e+06
min      0.000000e+00
25%      3.420000e+02
50%      2.400000e+03
75%      2.400000e+04
max      2.419600e+07
Name: COLI_FEC_NMP_100mL, dtype: float64
```



1.3.11 CALIDAD_COLI_FEC

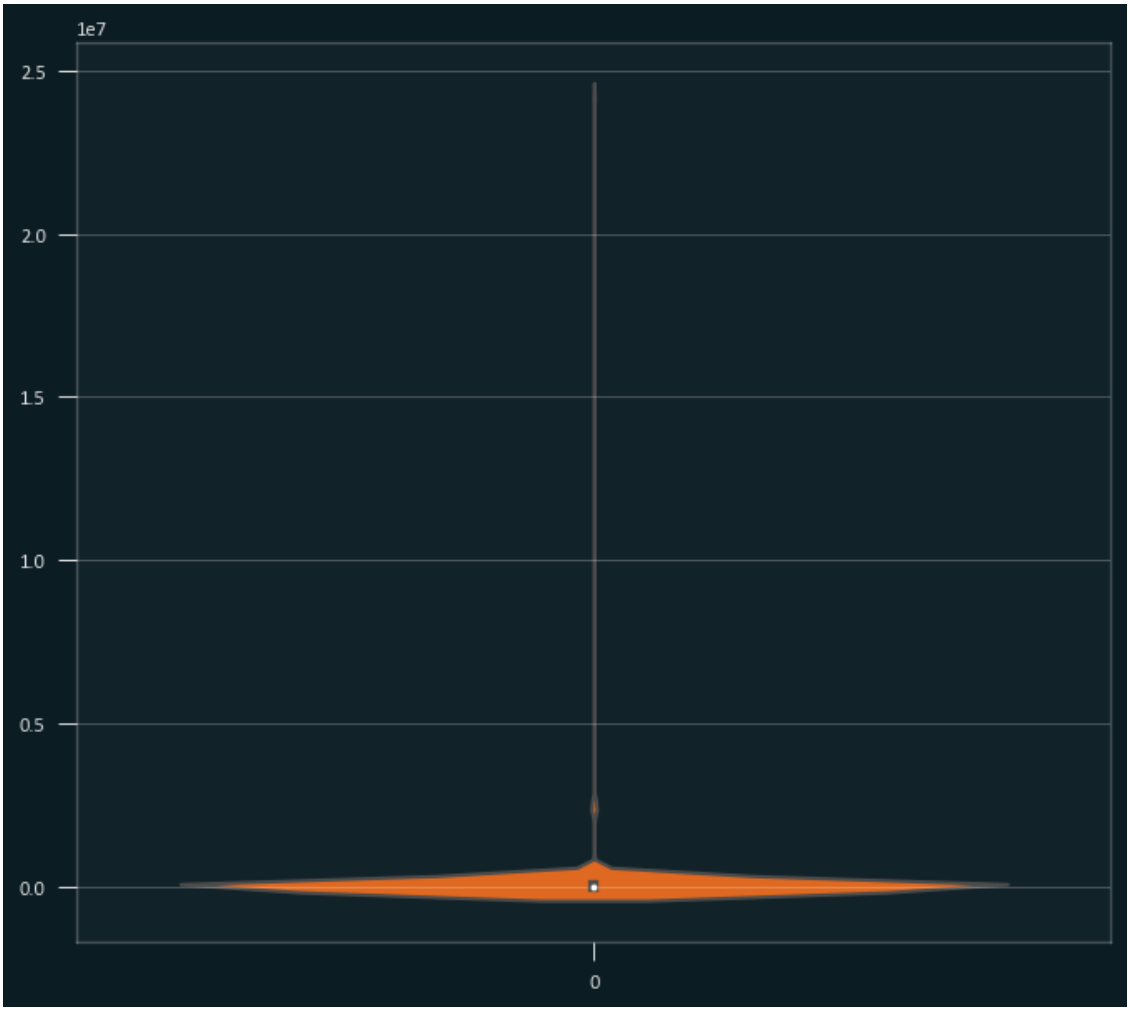
```
['Contaminada' nan 'Buena calidad' 'Fuertemente contaminada' 'Excelente'
 'Aceptable']
```



1.3.12 E_COLI_NMP_100mL

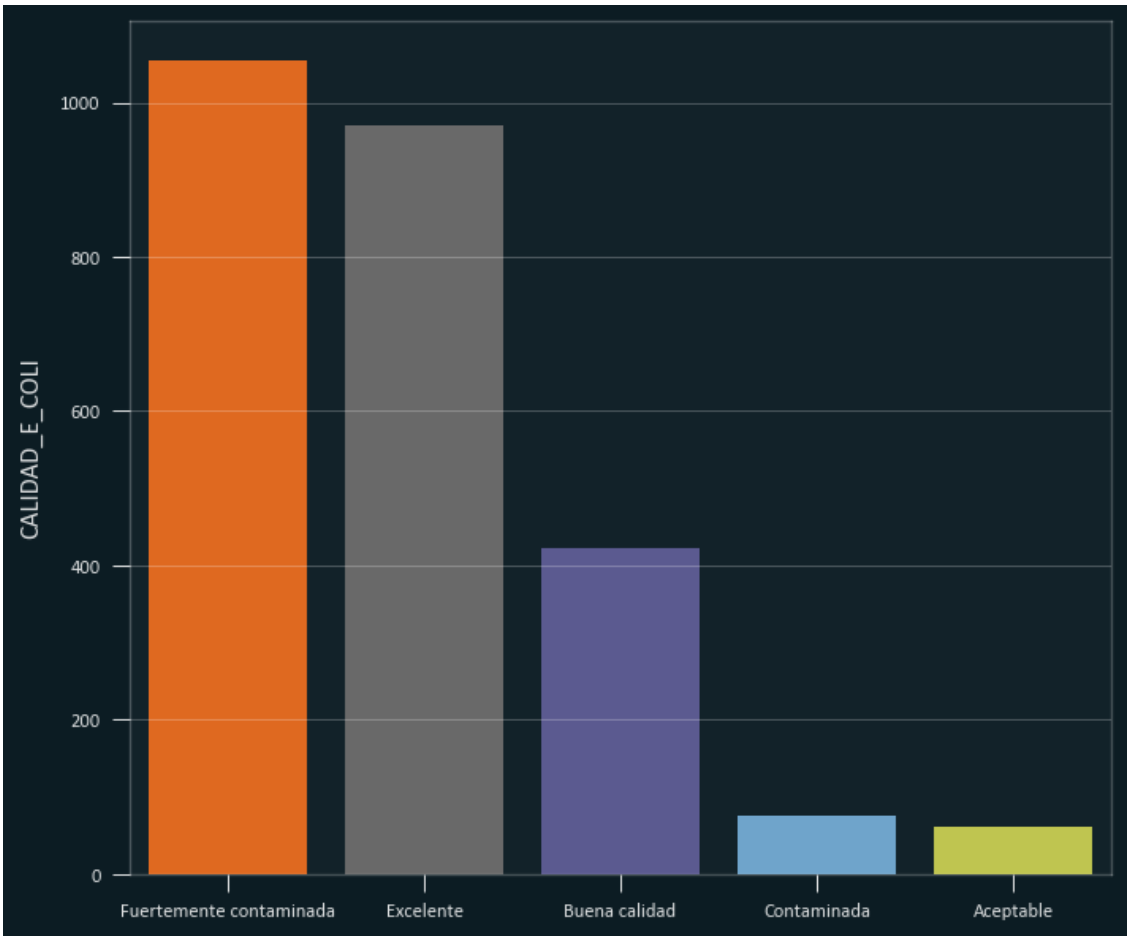
[49]:
```
array(['98', nan, '512', '<3', '84', '538', '14136', '74', '368', '20',
       '10', '384', '131', '3076', '97', '171', '489', '210', '40', '90',
       '230', '2400', '70', '60', '241960', '24196', '12997000', '1789',
       '52', '63', '187', '85', '201', '144', '32550', '8664', '459',
       '256', '142', '100', '121', '120', '86', '73', '249', '933',
       '26130', '1956', '3640', '1560', '19863', '7701', '11199', '15531',
       '1291', '3255', '34480', '108', '959', '3880', '158', '410',
       '5172', '285', '3448', '146', '31', '48040', '9208', '670', '4884',
       '563', '3654', '17329', '1935', '300', '605', '173', '61', '96',
       '110', '341', '9804', '41', '226', '243', '269', '866400', '650',
       '30', '189', '464', '1500', '930', '430', '11000', '200', '24000',
       '280', '2100', '4600', '150', '576', '4611', '4106', '1000',
       '104620', '2419600', '6131', '5794', '43', '460', '135400',
       '14800', '663000', '31000', '24196000', '14136000', '241',
       '1616000', '3076000', '313', '305', '1670', '123600', '86640',
       '211000', '750', '1100', '240', '15', '9', '23', '4', '7',
```

```
       '120330', '77010', '16580', '17820', '238000', '2110', '1200',
       '12033', '546', '231', '496', '224700', '259', '7760', '6488',
       '75', '21430', '20140', '740', '310', '27230', '27550', '38730',
       '41060', '3100', '1421', '11780', '61300', '8820', '3990', '6440',
       '4650', '18500', '441', '4590', '209', '51', '94', '203',
       '6488000', '223', '39', '93', '11', '28', '21', '3', '620', '754',
       '784', '95', '183', '573', '1785', '132', '934', '311', '130',
       '906', '2143', '10462', '5475', '7270', '8164', '22470', '5780',
       '119', '275', '169', '985', '644', '4352', '776', '26030', '21870',
       '13540', '480', '5730', '4410', '1674', '1090', '390', '504',
       '372', '233', '145', '140', '141', '1414', '932', '561', '836',
       '199', '355', '703', '107', '109', '6867', '2460', '4786', '1333',
       '161', '315', '228', '265', '318', '657', '292', '1376', '594',
       '663', '364', '281', '350', '833', '728', '457', '23590', '12997',
       '62', '2247', '155310', '10170', '43520', '198630', '27000',
       '5247', '175', '327', '691', '282', '958', '1585', '272', '448',
       '602', '738', '134', '545', '426', '521', '216', '64', '14',
       '13140', '345', '2430', '8570', '16160', '2280', '213', '141360',
       '36540', '51720', '279', '3170', '1455', '857', '373', '3730',
       '12360', '246', '92080', '29090', '64880', '173290', '15410',
       '7890', '3873', '9060', '238', '135', '10390', '1080', '4080',
       '1990', '1450', '2590', '1210', '630', '397', '1730', '6770',
       '420', '10760', '17850', '7380', '5520', '10310', '889', '336',
       '68670', '10432', '4140', '2142', '428', '677', '455', '537',
       '16640', '1401', '613100', '520', '2909', '905', '993', '759',
       '1723', '2310', '1246', '980', '970', '2014', '262', '72', '122',
       '1067', '2489', '2141', '235', '12230', '12460', '1296', '399',
       '1539', '5380', '6830', '4880', '6630', '206', '11600', '16700',
       '148', '2495', '1043', '422', '1076', '195'], dtype=object)

Percentage of unique value to total: 0.14136328427575523
Values distribution: count    2.582000e+03
mean     7.933726e+04
std      1.051334e+06
min      0.000000e+00
25%      4.000000e+01
50%      4.240000e+02
75%      6.488000e+03
max      2.419600e+07
Name: E_COLI_NMP_100mL, dtype: float64
```



1.3.13 CALIDAD_E_COLI

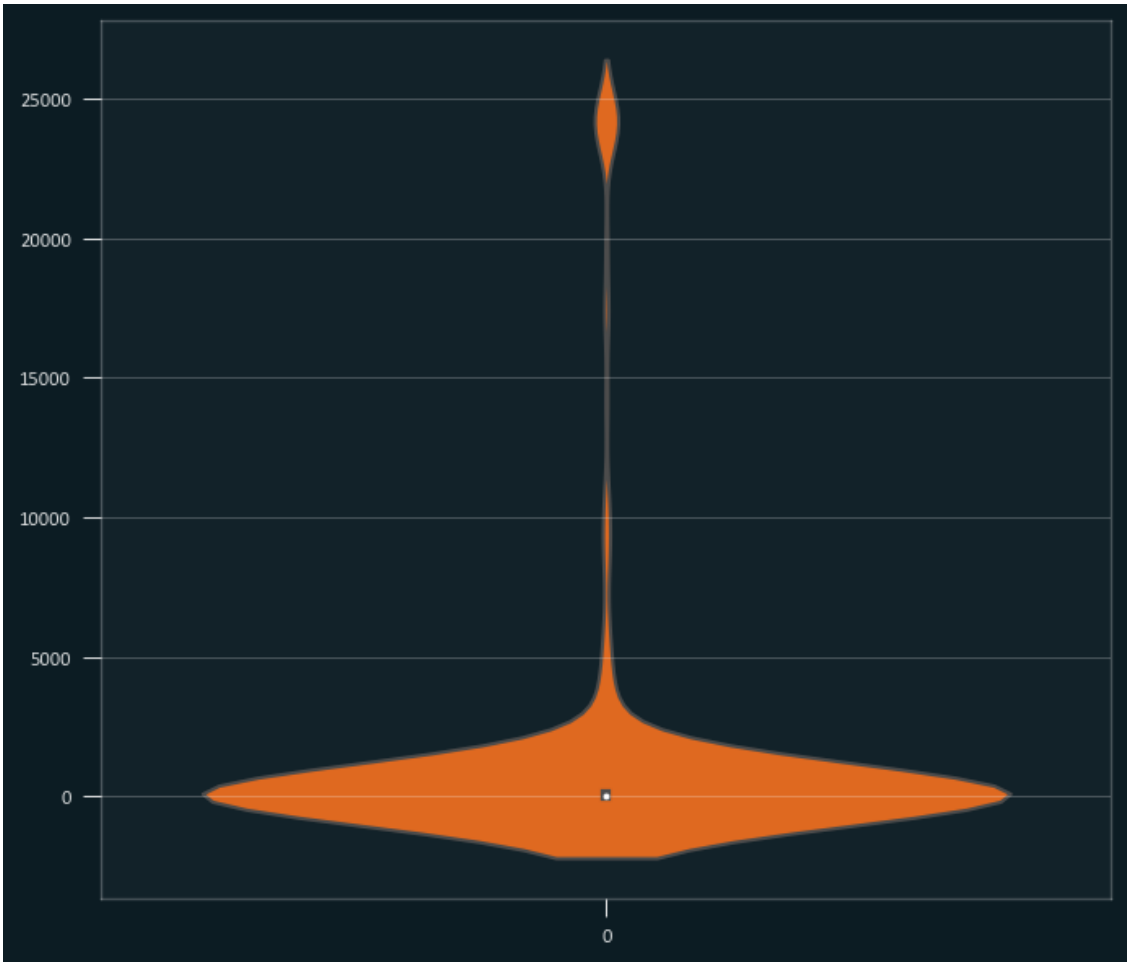
```
['Excelente' nan 'Buena calidad' 'Fuertemente contaminada' 'Contaminada'
 'Aceptable']
```



1.3.14 ENTEROC_NMP_100mL

[52]:
```
array([nan, '20', '<3', '30', '90', '402', '24196', '350', '355', '17329',
       '108', '118', '75', '4', '9', '7', '1722', '250', '556', '62',
       '41', '203', '10', '504', '148', '256', '173', '121', '187',
       '2778', '31', '2603', '5298', '246', '259', '11199', '14136',
       '3255', '5172', '52', '8664', '63', '5748', '109', '457', '1780',
       '73', '437', '865', '2114', '581', '84', '51', '320', '74', '98',
       '122', '4520', '1935', '19863', '422', '763', '608', '9208', '862',
       '114', '3724', '364', '959', '283', '1309', '131', '130', '359',
       '8704', '8164', '1169', '161', '345', '1050', '1211', '5794',
       '6488', '1259', '119', '97', '160', '43', '166', '512', '23', '93',
       '240', '21', '1376', '465', '275', '83', '669', '1723', '412',
       '382', '146', '1947', '5335', '4106', '776', '505', '1014',
       '10462', '3257', '435', '243', '10112', '82', '3873', '325', '155',
       '629', '2723', '3076', '408', '145', '573', '1100', '529', '10.9',
       '2419.6', '365.4', '57.8', '34.5', '437.4', '39.5', '173.3',
       '410.6', '33.6', '31.2', '284.1', '325.7', '105.7', '19.1',
```

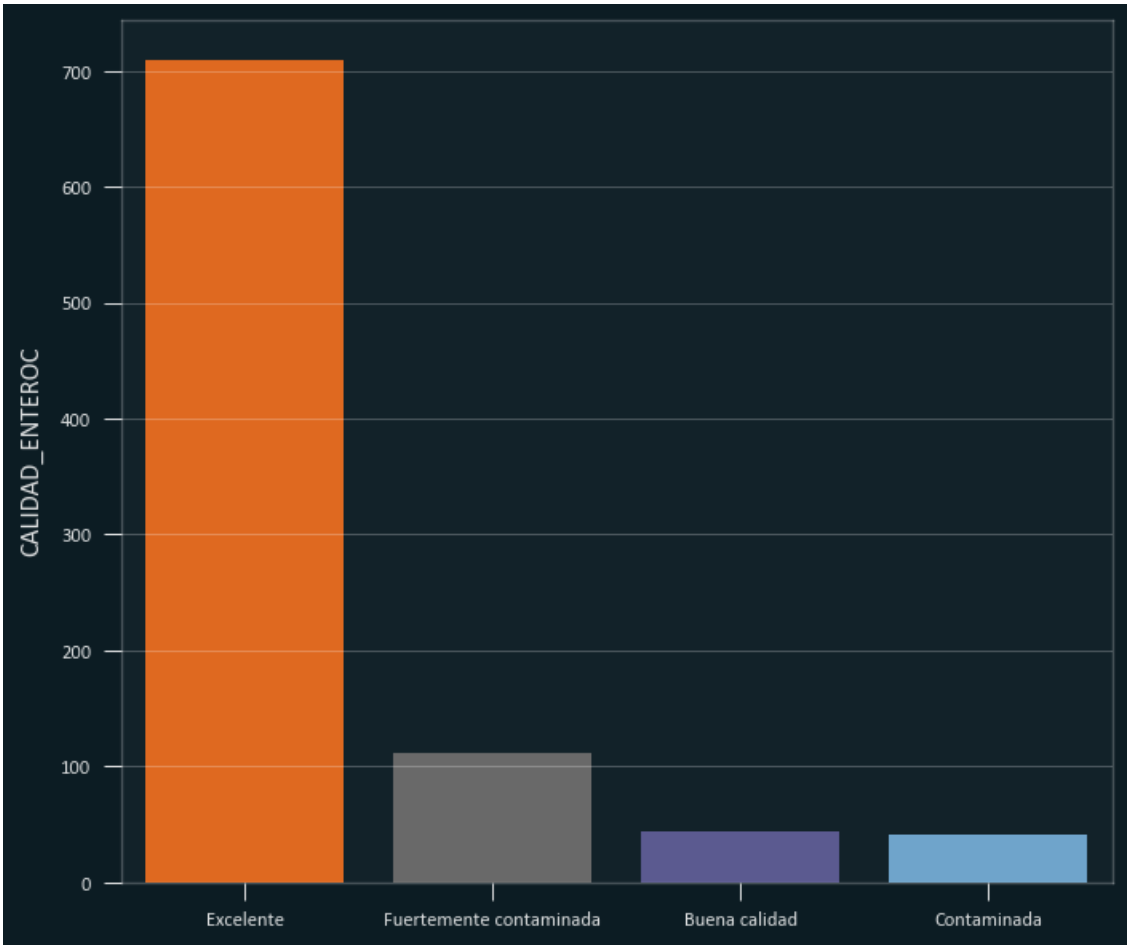
```
       '317.4', '867', '3282', '2012', '197', '488', '14.8', '11', '6.3',
       '1789', '1210', '2098', '85', '9804', '86', '110', '153', '71',
       '677', '135', '61', '1086', '2909', '327', '191', '258', '137',
       '1036', '253', '142', '720', '938', '1833', '2037', '193', '120',
       '95', '1057', '4352', '1250', '3.1', '70.1', '7.5', '5.2', '265'],
      dtype=object)

Percentage of unique value to total: 0.5132743362831859
Values distribution: count      904.000000
mean      1085.170022
std       4306.301542
min          0.000000
25%          0.000000
50%          0.000000
75%         63.000000
max      24196.000000
Name: ENTEROC_NMP_100mL, dtype: float64
```



1.3.15 Calidades

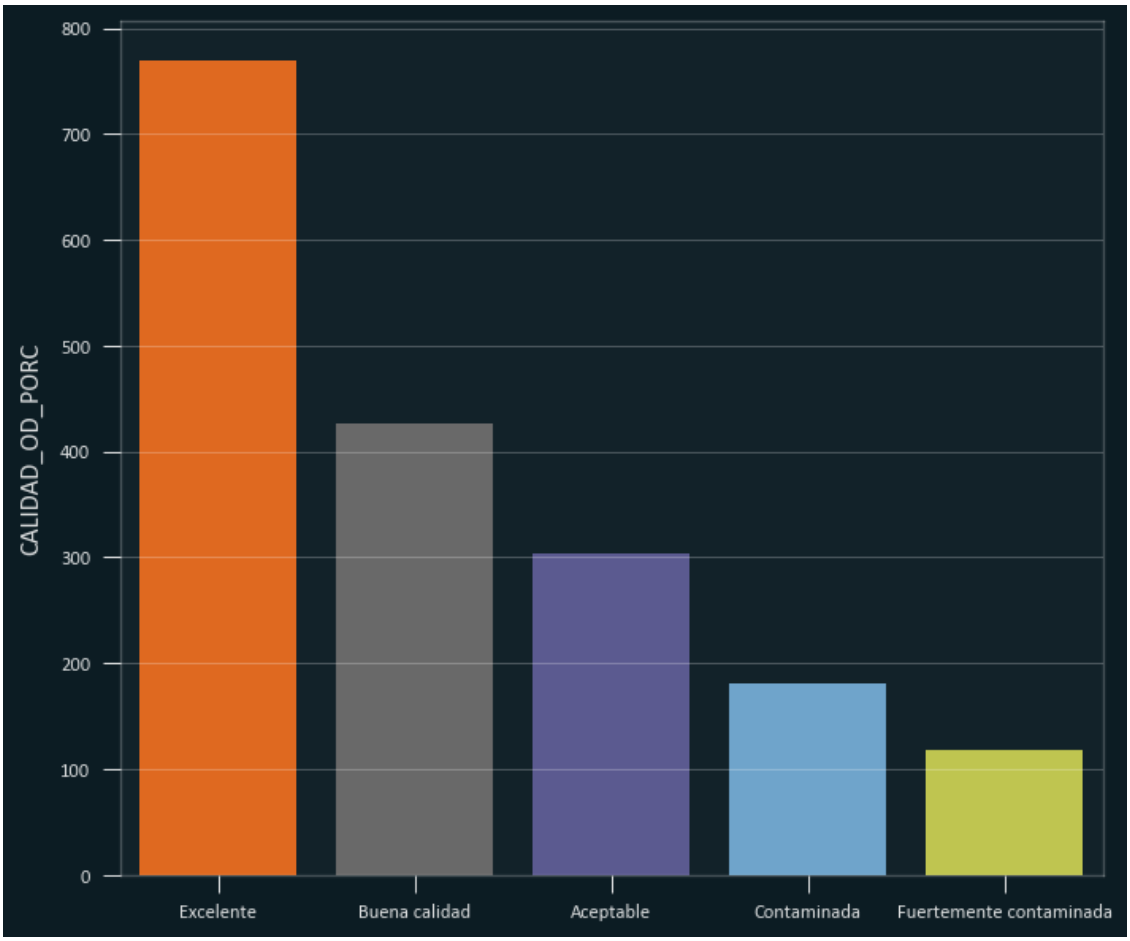
1.3.16 CALIDAD_ENTEROC

```
[nan 'Excelente' 'Contaminada' 'Fuertemente contaminada' 'Buena calidad']
```



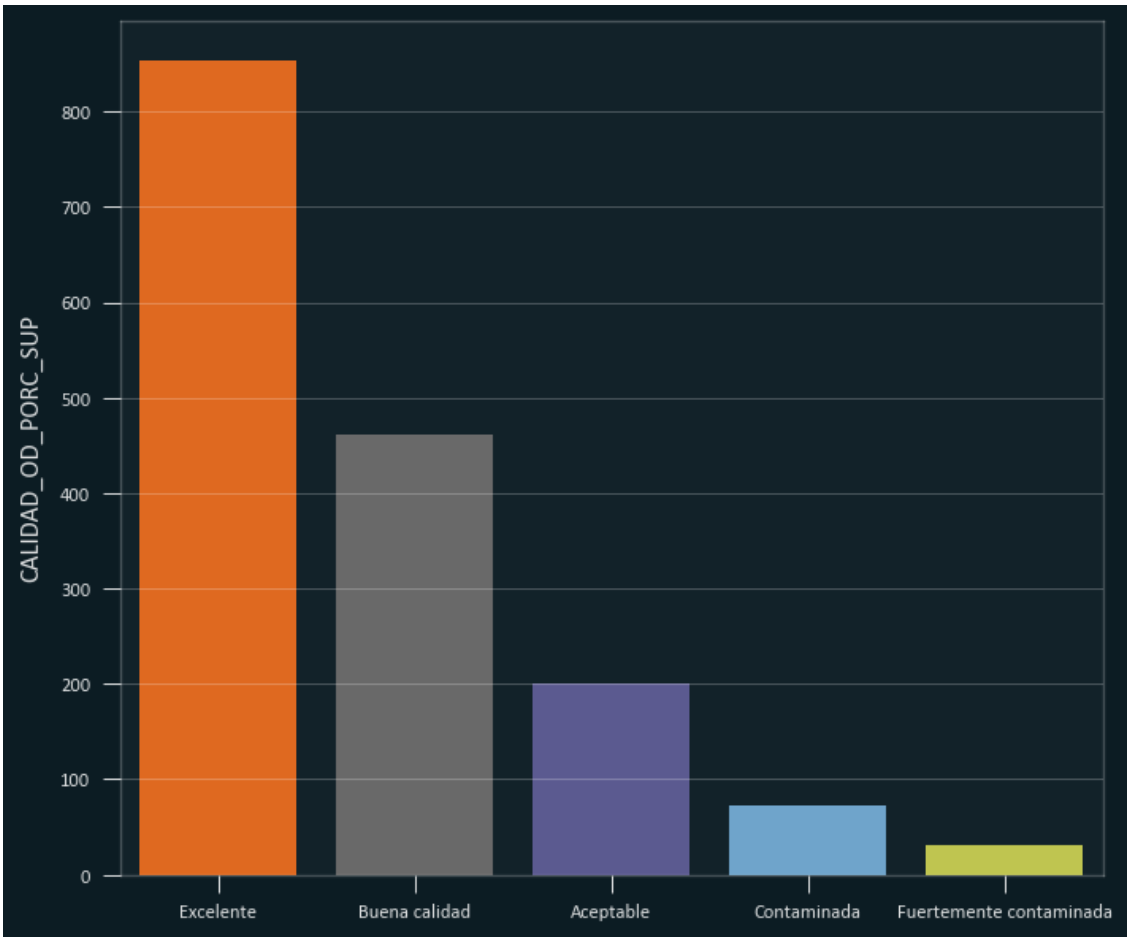
1.3.17 CALIDAD_OD_PORC

```
[nan 'Excelente' 'Contaminada' 'Fuertemente contaminada' 'Aceptable'
 'Buena calidad']
```



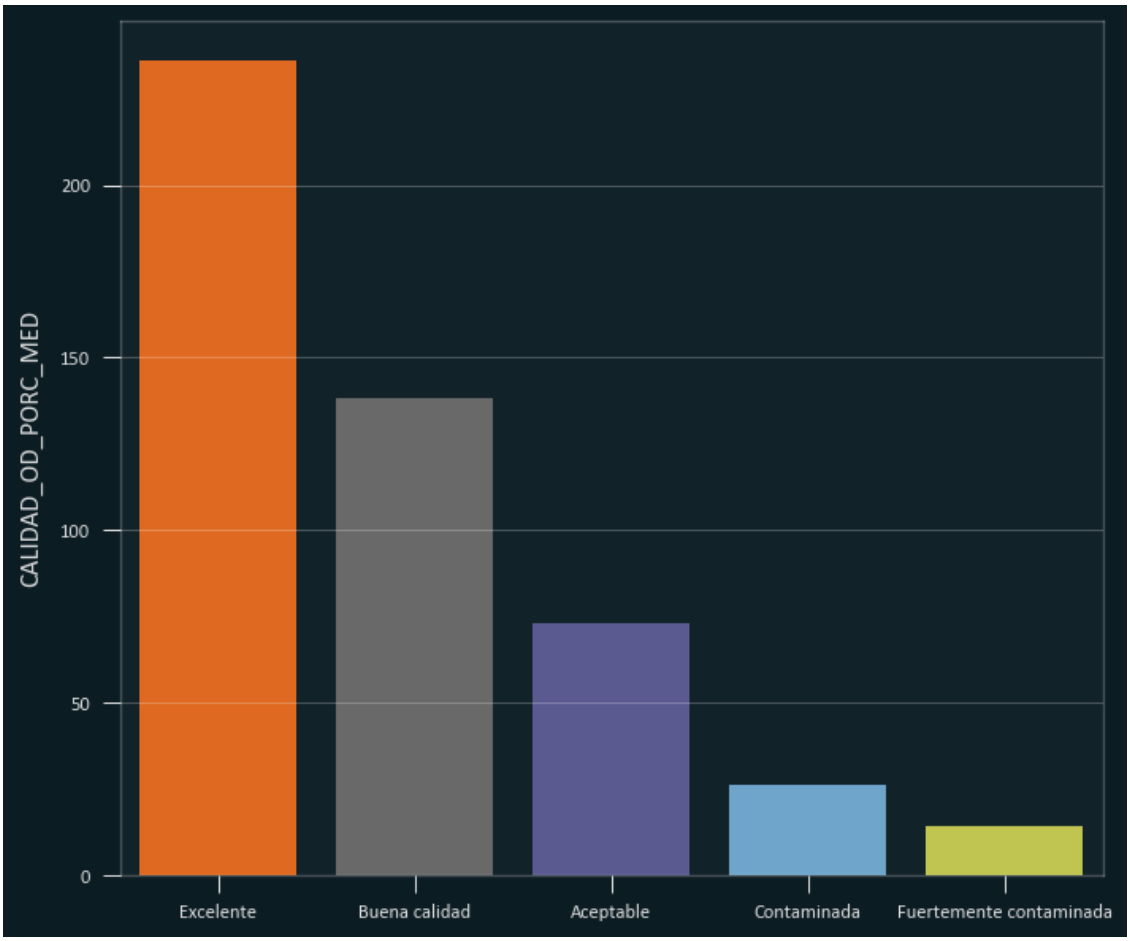
1.3.18 CALIDAD_OD_PORC_SUP

```
['Aceptable' 'Excelente' nan 'Buena calidad' 'Contaminada'
 'Fuertemente contaminada']
```



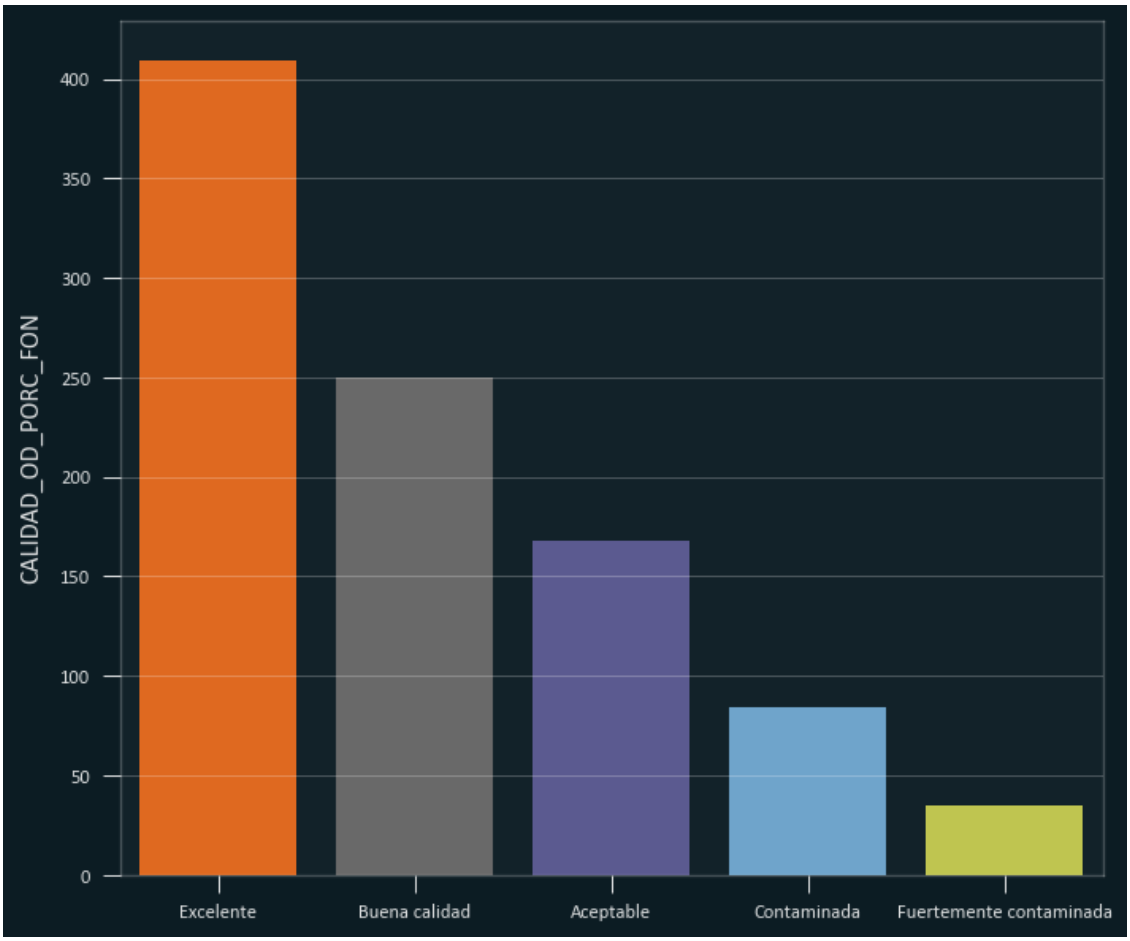
1.3.19 CALIDAD_OD_PORC_MED

```
[nan 'Excelente' 'Buena calidad' 'Contaminada' 'Aceptable'
 'Fuertemente contaminada']
```



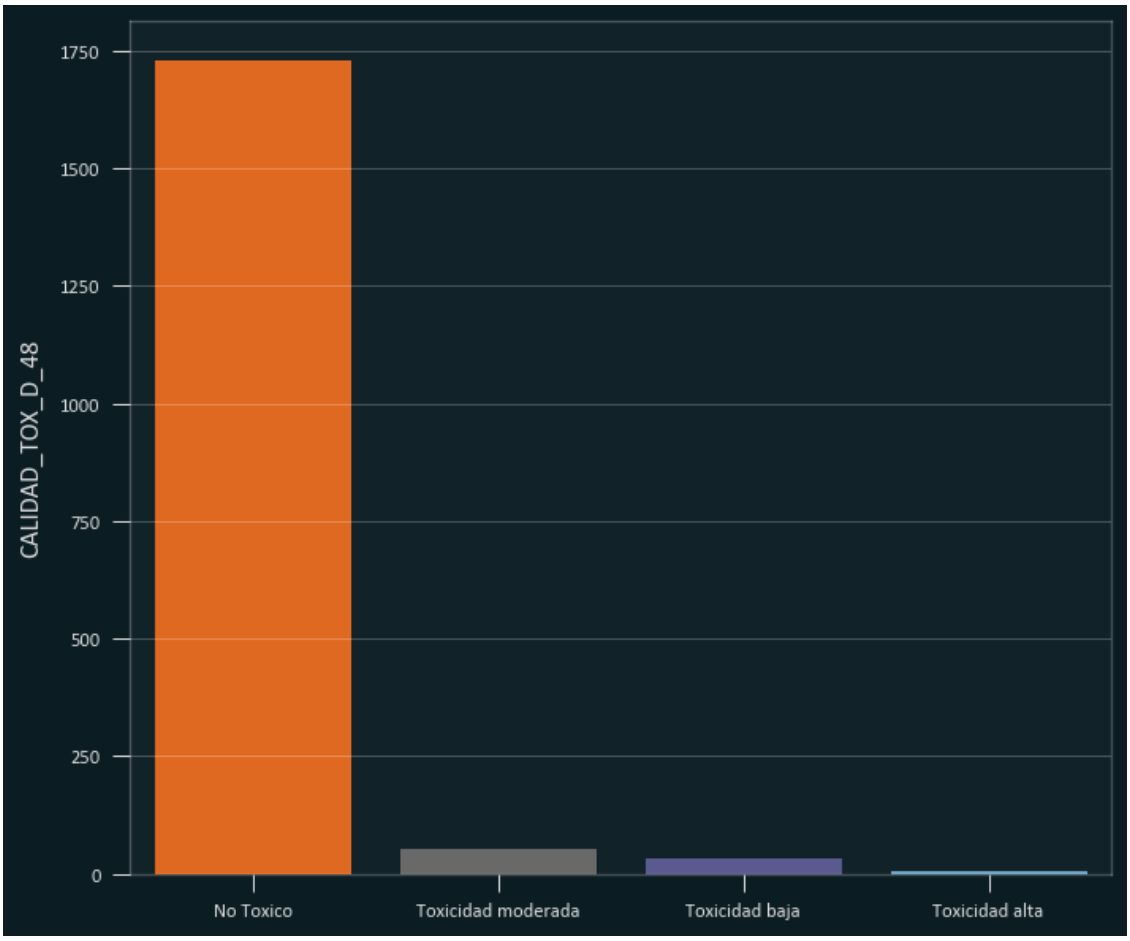
1.3.20 CALIDAD_OD_PORC_FON

```
[nan 'Excelente' 'Contaminada' 'Buena calidad' 'Aceptable'
 'Fuertemente contaminada']
```



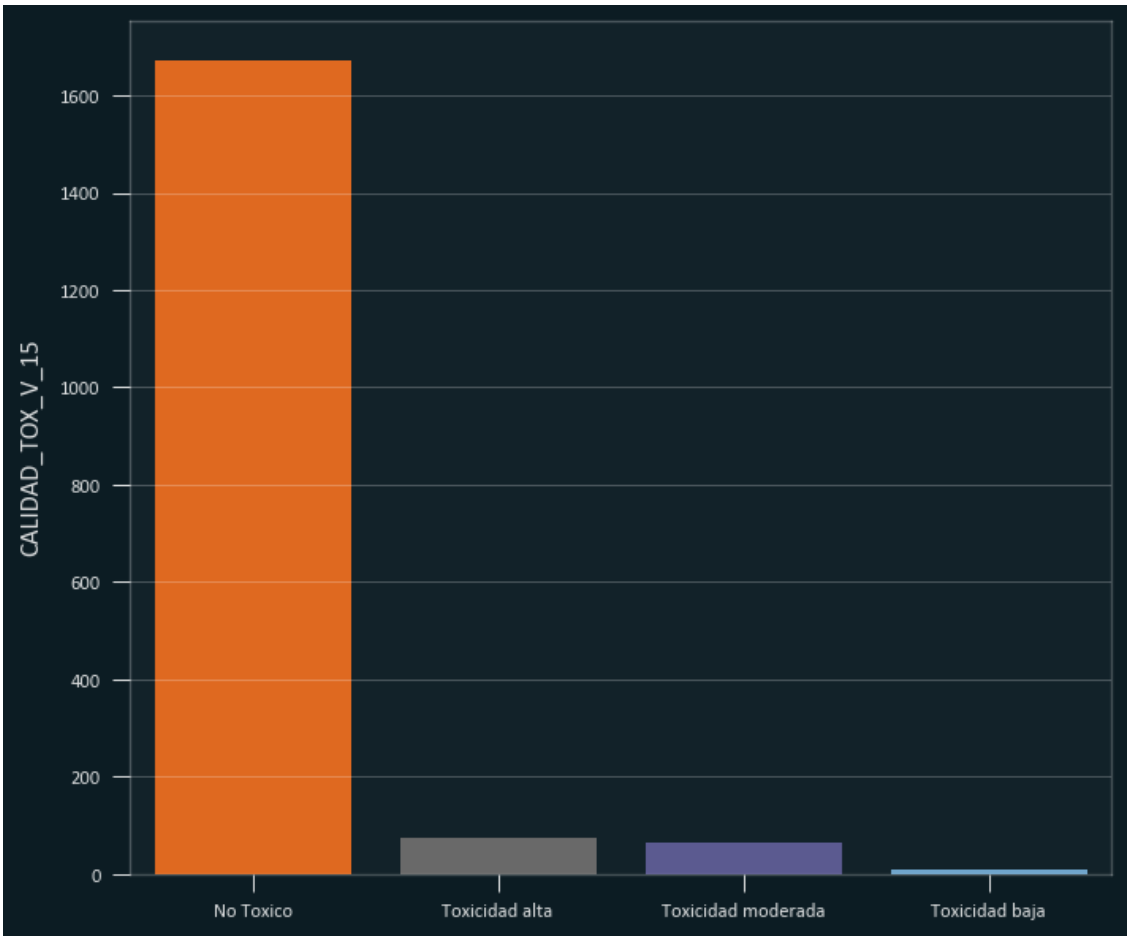
1.3.21 CALIDAD_TOX_D_48

```
[nan 'No Toxico' 'Toxicidad moderada' 'Toxicidad baja' 'Toxicidad alta']
```



1.3.22 CALIDAD_TOX_V_15

```
[nan 'No Toxico' 'Toxicidad moderada' 'Toxicidad alta' 'Toxicidad baja']
```



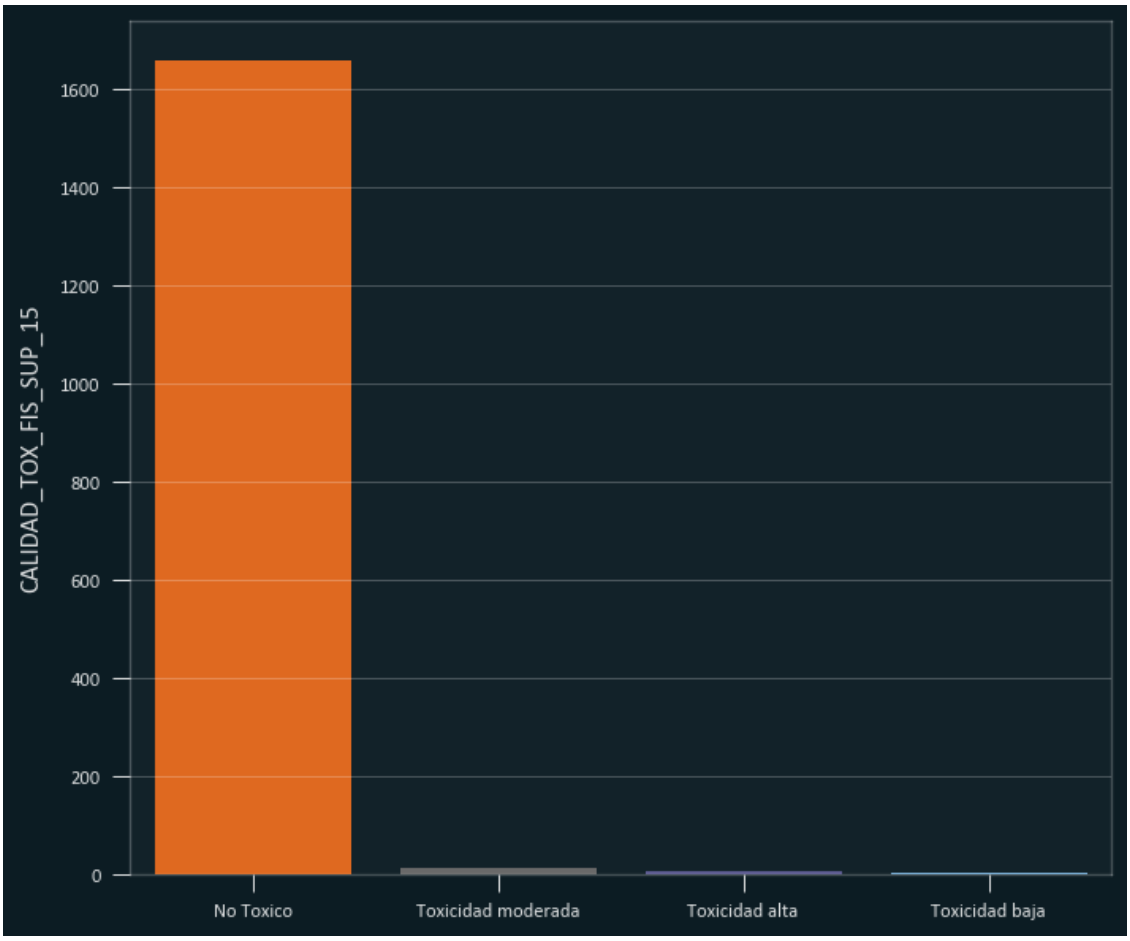
1.3.23 CALIDAD_TOX_D_48_FON

`[nan]`

Esta columna está vacía, por tanto no contiene ningún registro y se puede eliminar del conjunto de datos completo.

1.3.24 CALIDAD_TOX_FIS_SUP_15

`['No Toxico' nan 'Toxicidad moderada' 'Toxicidad baja' 'Toxicidad alta']`



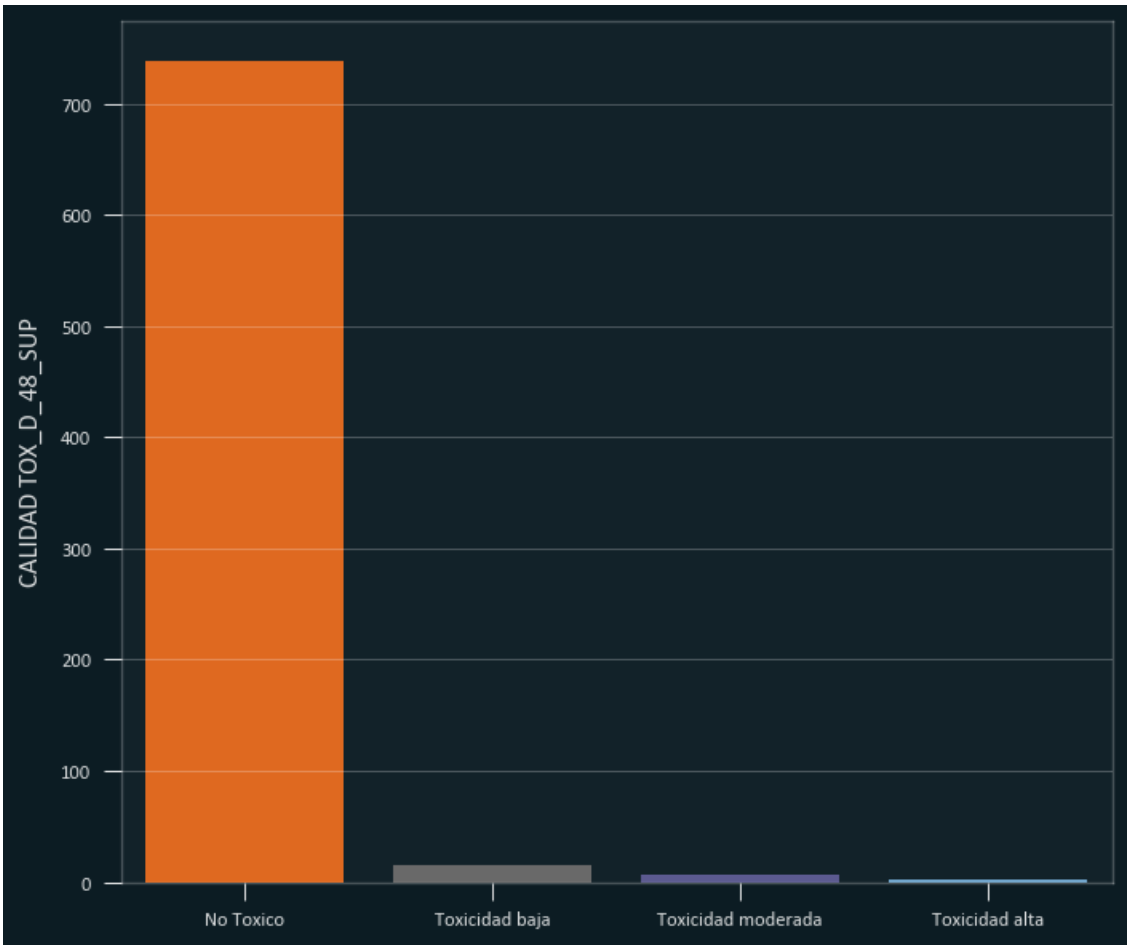
1.3.25 CALIDAD_TOX_FIS_FON_15

`[nan]`

Columna vacía, por tanto se elimina del conjunto de datos.

1.3.26 CALIDAD TOX_D_48_SUP

`['No Toxico' nan 'Toxicidad moderada' 'Toxicidad baja' 'Toxicidad alta']`



1.3.27 Contaminantes

```
['DQO,CF,' nan 'ENT_FEC,OD%F,' 'CF,' 'OD%L,' 'CF,OD%L,' 'CF,E_COLI,'
 'ENT_FEC,' 'OD%S,OD%M,OD%F,' 'OD%S,' 'DQO,' 'DBO,DQO,' 'SST,'
 'SST,CF,E_COLI,' 'DBO,DQO,CF,E_COLI,TOX_L,' 'DQO,CF,E_COLI,'
 'DBO,DQO,CF,E_COLI,OD%L,' 'DBO,DQO,CF,E_COLI,OD%L,TOX_L,'
 'DQO,CF,E_COLI,OD%L,' 'DBO,DQO,CF,E_COLI,' 'DQO,CF,OD%L,'
 'DQO,SST,CF,E_COLI,' 'SST,CF,' 'CF,OD%S,' 'DQO,CF,OD%S,' 'SST,ENT_FEC,'
 'ENT_FEC,OD%S,' 'DQO,OD%S,' 'CF,E_COLI,OD%L,' 'OD%F,' 'DQO,SST,CF,'
 'DQO,OD%L,' 'DBO,DQO,OD%L,' 'DBO,DQO,E_COLI,OD%L,' 'DBO,DQO,OD%L,TOX_L,'
 'DBO,DQO,SST,' 'DBO,DQO,SST,CF,E_COLI,TOX_L,' 'DQO,SST,' 'DQO,OD%M,OD%F,'
 'DQO,OD%F,' 'DBO,DQO,SST,CF,E_COLI,' 'DBO,DQO,OD%M,OD%F,'
 'CF,E_COLI,OD%S,OD%M,OD%F,' 'CF,E_COLI,OD%M,OD%F,' 'E_COLI,OD%M,OD%F,'
 'DQO,CF,E_COLI,OD%M,OD%F,' 'DBO,DQO,SST,E_COLI,OD%L,TOX_L,'
 'DBO,DQO,SST,CF,E_COLI,OD%L,TOX_L,' 'DBO,DQO,CF,'
 'DQO,SST,CF,E_COLI,OD%L,' 'DBO,DQO,CF,E_COLI,OD%S,OD%F,'
 'DBO,DQO,SST,CF,E_COLI,OD%L,' 'DBO,DQO,SST,E_COLI,' 'DQO,SST,OD%L,'
 'SST,CF,E_COLI,OD%L,' 'DQO,SST,CF,OD%L,' 'DBO,DQO,E_COLI,'
```

```
'DQO,CF,E_COLI,OD%S,OD%F,' 'DQO,E_COLI,' 'E_COLI,' 'DBO,E_COLI,'
'DBO,DQO,OD%S,OD%F,' 'OD%M,OD%F,' 'DBO,CF,E_COLI,' 'CF,E_COLI,OD%S,'
'DBO,DQO,OD%S,' 'DBO,DQO,SST,CF,' 'DQO,OD%S,OD%M,OD%F,'
'DBO,DQO,SST,CF,E_COLI,OD%S,' 'DBO,DQO,SST,OD%S,' 'DQO,CF,OD%F,'
'ENT_FEC,OD%S,OD%F,' 'DQO,SST,CF,E_COLI,OD%L,TOX_L,'
'CF,E_COLI,OD%S,OD%F,' 'DBO,DQO,SST,CF,OD%L,' 'DQO,SST,CF,E_COLI,TOX_L,'
'DQO,CF,E_COLI,OD%L,TOX_L,' 'DQO,TOX_S,' 'DBO,DQO,CF,OD%S,OD%F,'
'DQO,E_COLI,OD%S,OD%M,OD%F,' 'DQO,CF,E_COLI,OD%S,OD%M,OD%F,'
'DQO,CF,E_COLI,TOX_L,' 'DBO,DQO,SST,OD%L,TOX_L,' 'E_COLI,OD%L,'
'CF,E_COLI,OD%F,' 'SST,E_COLI,' 'DBO,DQO,SST,E_COLI,TOX_L,' 'TOX_S,'
'DQO,OD%S,OD%F,' 'OD%S,TOX_S,' 'OD%S,OD%F,' 'DQO,CF,E_COLI,OD%S,'
'CF,OD%F,' 'DQO,TOX_L,' 'DBO,DQO,CF,OD%L,TOX_L,' 'SST,OD%S,TOX_S,'
'DBO,DQO,CF,TOX_L,']
```

[68]: 1267

Imputamos un valor desconocido, ya que las categorías y los tipos de contaminación son muy variados.

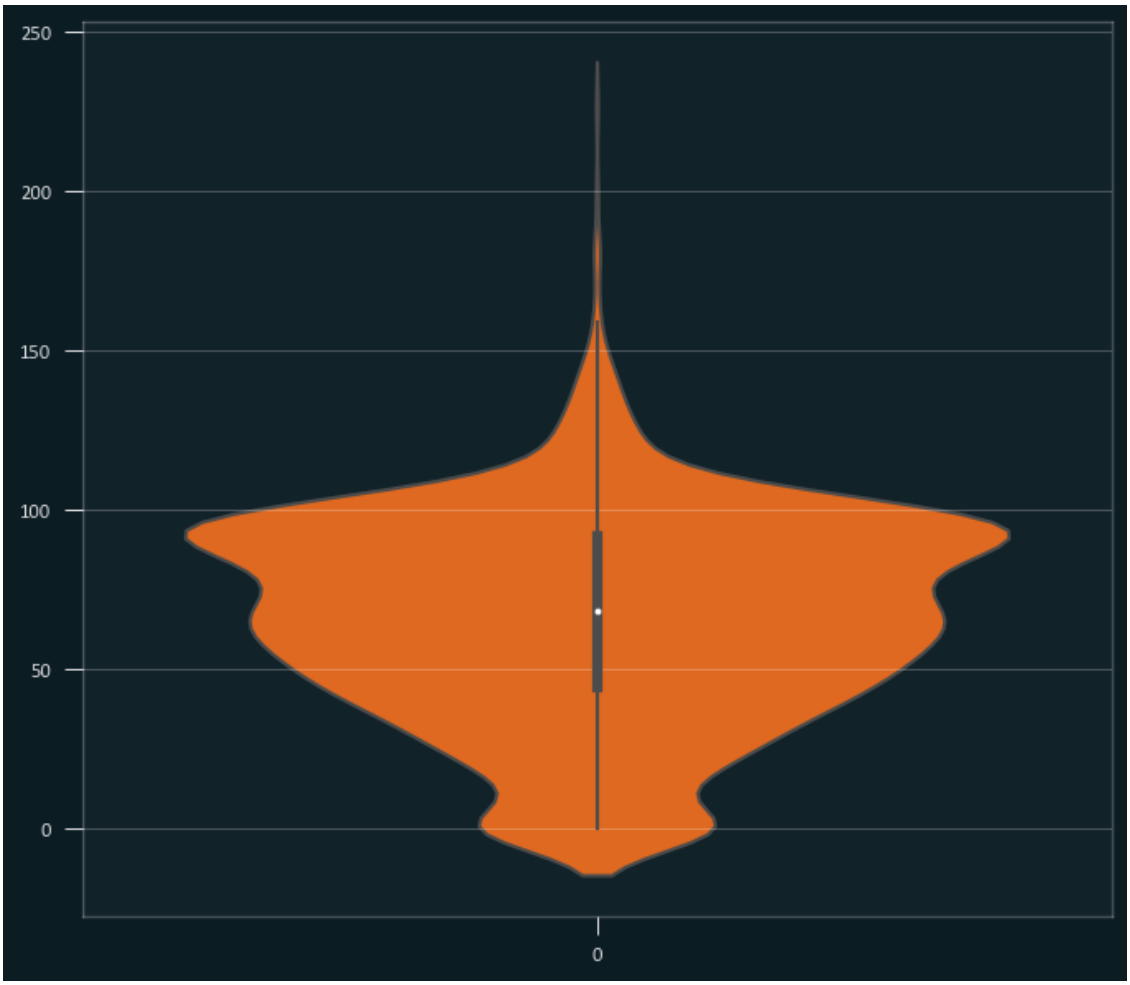
1.3.28 Numéricas

1.3.29 OD_PORC

[70]:
```
array([nan, '83.6', '137', '90', '25', '104.8', '95.1', '93.4', '85.4',
       '98.2', '86.2', '200', '143.07', '122.6', '111.8', '22.4', '69.1',
       '50.8', '84.1', '22', '58.2', '50.2', '45.3', '43.6', '49.2',
       '51.7', '60.2', '61.7', '57.1', '50.1', '49.7', '51', '54.7',
       '56.8', '55.6', '94.4', '33', '27', '<10', '120.2', '20.1', '53.9',
       '98', '97.1', '97.2', '96.4', '93.5', '101', '88.1', '102.7',
       '69.7', '94.1', '114.8', '128.4', '120.8', '145.4', '125.8', '86',
       '114.4', '60.3', '40.8', '64.6', '39.7', '54.6', '32.6', '96.3',
       '97.3', '92.5', '83.4', '96.1', '100.9', '82', '99', '100.2', '74',
       '94.6', '99.9', '96.9', '97.9', '102.3', '102', '103.8', '55.7',
       '92', '94', '100.5', '87.7', '70.7', '101.7', '119.8', '107',
       '100.8', '107.8', '104.6', '107.4', '101.8', '57.8', '106.9',
       '102.4', '101.4', '127.2', '109.9', '106.8', '97.6', '114',
       '158.8', '118.3', '100.6', '127.9', '98.6', '78.4', '40', '124.4',
       '83.9', '82.6', '80', '84.5', '140.7', '67.8', '103.4', '88.3',
       '96.6', '47.5', '56.1', '118.1', '226.1', '107.1', '109.8',
       '101.2', '92.8', '63.3', '76.2', '10.6', '18.8', '40.7', '11.8',
       '20.2', '17.8', '34.3', '39.4', '36.5', '37.4', '46.2', '22.2',
       '30.9', '48.3', '45.4', '45.1', '31.8', '33.5', '43.1', '37.9',
       '36.7', '90.4', '81.8', '99.6', '95.7', '43.7', '98.1', '88.2',
       '101.5', '16.4', '19.1', '32.8', '100.1', '95.3', '62.2', '78.6',
       '103.7', '15.6', '41.6', '82.7', '14.8', '38.22', '35.2', '18.6',
       '16', '73.5', '65.2', '21.3', '28.69', '11.3', '29.1', '18.2',
       '76.6', '26.1', '71.9', '26.4', '10.9', '81.1', '24', '33.1',
       '19.5', '18.3', '20.3', '75.5', '64.4', '134', '92.9', '42.8',
       '75.8', '53.4', '36.4', '68.9', '36', '26.6', '20', '65.7', '52.6',
```

'28.8', '22.6', '87.6', '27.8', '10.8', '12.7', '76.9', '71',
'80.01', '70.5', '71.6', '66.7', '68.3', '69.4', '68.8', '66.1',
'81.9', '66', '64.1', '68.1', '79.3', '67.4', '66.8', '66.3',
'70.6', '63.8', '79', '55.4', '50.6', '80.3', '69.2', '69.5',
'66.9', '70.3', '67.9', '63.4', '46.1', '68.4', '67.1', '68.5',
'69.3', '68.2', '66.4', '69.8', '60.4', '60.1', '51.4', '55.1',
'57.9', '52.4', '56.6', '56.3', '59.3', '59.4', '60.6', '52.8',
'59.9', '90.2', '93.3', '86.9', '79.5', '89.9', '72.1', '36.2',
'27.1', '17.5', '22.3', '31.5', '25.3', '78.5', '44.3', '73.6',
'62.8', '34.1', '32.3', '62.1', '37.2', '42.5', '44.1', '48.4',
'72.2', '67.3', '53.7', '69.9', '91.24', '24.6', '75', '38.4',
'99.2', '76.5', '63.5', '74.8', '50.4', '121.3', '108.1', '108.3',
'96.7', '58.7', '44.8', '49', '46.3', '60.8', '98.5', '52.1',
'64.7', '70', '80.7', '74.2', '94.3', '135.5', '51.6', '108.2',
'91.1', '70.1', '46.7', '59.1', '91.2', '108.8', '92.7', '95.5',
'92.2', '82.3', '96.2', '105.5', '91.6', '23.1', '49.6', '63',
'41', '37.1', '60', '48.5', '17.7', '21.5', '30.1', '58.39',
'16.7', '90.9', '88.9', '89.8', '88.6', '23.6', '179.5', '45.2',
'102.8', '79.27', '85.3', '85.8', '100', '130.3', '131.8', '115.2',
'66.5', '45', '100.3', '81.7', '84.2', '101.1', '74.1', '87.1',
'77.3', '72.8', '82.9', '79.4', '86.7', '93.6', '72.3', '82.4',
'91', '78.8', '98.4', '87.2', '89', '95.9', '95', '99.8', '93.7',
'94.2', '95.4', '97.7', '83.3', '27.5', '24.9', '35.8', '24.3',
'25.8', '21.2', '23.9', '32.5', '28', '44.6', '47.7', '63.7',
'27.9', '31.9', '65.9', '79.6', '53', '59.8', '34.8', '57.2',
'58.5', '51.5', '27.3', '23.3', '35.7', '29.4', '25.7', '33.3',
'43.8', '42.7', '49.9', '45.5', '41.9', '45.7', '41.5', '42.6',
'33.8', '36.6', '38.8', '42.4', '34.6', '36.3', '29.5', '29.7',
'30.7', '23.5', '35.1', '50.3', '53.8', '26.5', '47.3', '58',
'11.2', '12.3', '16.8', '137.1', '122.8', '109.2', '61.1', '83.8',
'77.5', '94.7', '48', '25.1', '87.8', '96.5', '38', '40.5', '40.1',
'56.5', '63.2', '64.5', '44.2', '56.7', '54.5', '56.4', '75.1',
'29.3', '95.8', '48.8', '40.4', '73.8', '83', '81.2', '72', '55',
'69', '104.5', '94.5', '58.6', '102.1', '38.5', '105.9', '86.5',
'104', '114.6', '101.9', '93', '49.4', '50', '78', '76', '97',
'74.3', '64.9', '65.6', '83.5', '92.3', '91.5', '90.3', '89.7',
'90.6', '17.3', '28.6', '41.4', '21.1', '42.3', '20.4', '36.1',
'55.9', '58.8', '48.9', '43.2', '35.3', '77', '81.6', '81.4',
'84.9', '42.9', '96.8', '99.4', '98.3', '88.8', '75.7', '95.2',
'48.7', '53.1', '67', '51.8', '11.5', '14.9', '14.4', '29.6',
'39.1', '17.2', '11.6', '19.3', '13.8', '13.6', '13.2', '14.2',
'28.2', '27.4', '103.2', '102.2', '141.1', '57.3', '74.9', '92.4',
'92.1', '80.4', '15.3', '44.4', '32.1', '17.6', '44.5', '13.1',
'23.4', '39.6', '89.1', '107.6', '87.5', '96', '102.9', '47.6',
'84.4', '38.6', '11.1', '41.2', '41.3', '28.3', '48.2', '75.3',
'86.3', '89.4', '86.8', '87.4', '97.5', '85.1', '89.6', '81.3',
'75.4', '80.1', '40.3', '49.3', '40.6', '47.8', '39.8', '31.6',

```
       '109', '129.6', '110.6', '106.1', '116.6', '135.9', '64.3', '52.5',
       '54.4', '54.9', '54.3', '61.4', '57.5', '53.2', '81.5', '68.7',
       '100.4', '77.2', '61.9', '58.4', '77.1', '59.2', '65.4', '83.7',
       '72.9', '71.2', '88.5', '60.5', '86.6', '62.4', '66.2', '80.2',
       '83.2', '46.8', '79.8', '71.1', '74.6', '85.6', '91.4', '80.8',
       '70.4', '59.7', '57', '62', '54.1', '71.4', '73.4', '64.8', '35',
       '32', '44', '102.5', '83.1', '88.4', '93.2', '66.6', '33.2',
       '61.6', '78.9', '77.7', '58.9', '69.6', '31.4', '87.9', '34',
       '26.9', '61.3', '65.1', '104.7', '13.9', '73.1', '29', '70.2',
       '23', '11.9', '31', '39', '68', '54', '26', '93.1', '102.6', '61',
       '42', '52', '59', '21.4', '70.8', '104.9', '77.9', '15', '51.3',
       '77.4', '82.8', '75.6', '12', '98.9', '101.6', '103.5', '46.4',
       '105.7', '46', '47', '56', '43', '108', '85.2', '81', '39.5',
       '50.9', '37', '74.5', '99.1', '64', '42.1', '12.8', '58.3', '46.5',
       '49.5', '89.2', '80.9', '82.5', '88', '61.2', '94.9', '98.7', '65',
       '85', '84', '93.9', '19.2', '98.8', '79.92', '30.2', '51.2',
       '28.4', '30.4', '25.9', '30.5', '56.2', '91.8', '87', '71.3',
       '63.9', '61.8', '86.1', '71.5', '67.5', '123', '85.5', '125',
       '113.4', '180.1', '165.8', '126', '97.8', '118.2', '138.8',
       '111.9', '106.4', '100.7', '103.3', '121', '73', '18.1', '62.3',
       '110.4', '72.7', '89.3', '22.7', '31.7', '112.4', '63.6', '19.6',
       '91.7', '105.6', '84.6', '109.4', '62.6', '67.2', '77.8', '82.1',
       '116.2', '115.6', '149.2', '62.7', '28.1', '131.1', '84.3',
       '110.3', '47.4', '71.7', '56.9', '45.8', '110.2', '94.8', '53.3',
       '60.9', '108.4', '141', '123.2', '24.8', '39.9', '41.8', '45.6',
       '49.65', '54.2', '73.2', '59.6', '65.3', '139.8', '115.7', '119.2',
       '137.8', '131.5', '182.6', '129.3', '140.3', '16.1', '130.1',
       '145.2', '75.2', '107.5', '130.2', '125.2', '146.7', '124.8',
       '127.7', '113.8', '150', '104.4', '121.8', '122.4', '157.3',
       '99.7', '117.2', '110.1', '127.4', '104.2', '109.7', '90.8',
       '124.3', '105.8', '88.7', '76.1', '10.14', '72.5', '91.3', '86.4'],
      dtype=object)

Percentage of unique value to total: 0.06121313299944352
Values distribution: count    1797.000000
mean       66.228158
std        32.594686
min         0.000000
25%        44.300000
50%        68.300000
75%        91.600000
max       226.100000
Name: OD_PORC, dtype: float64
```

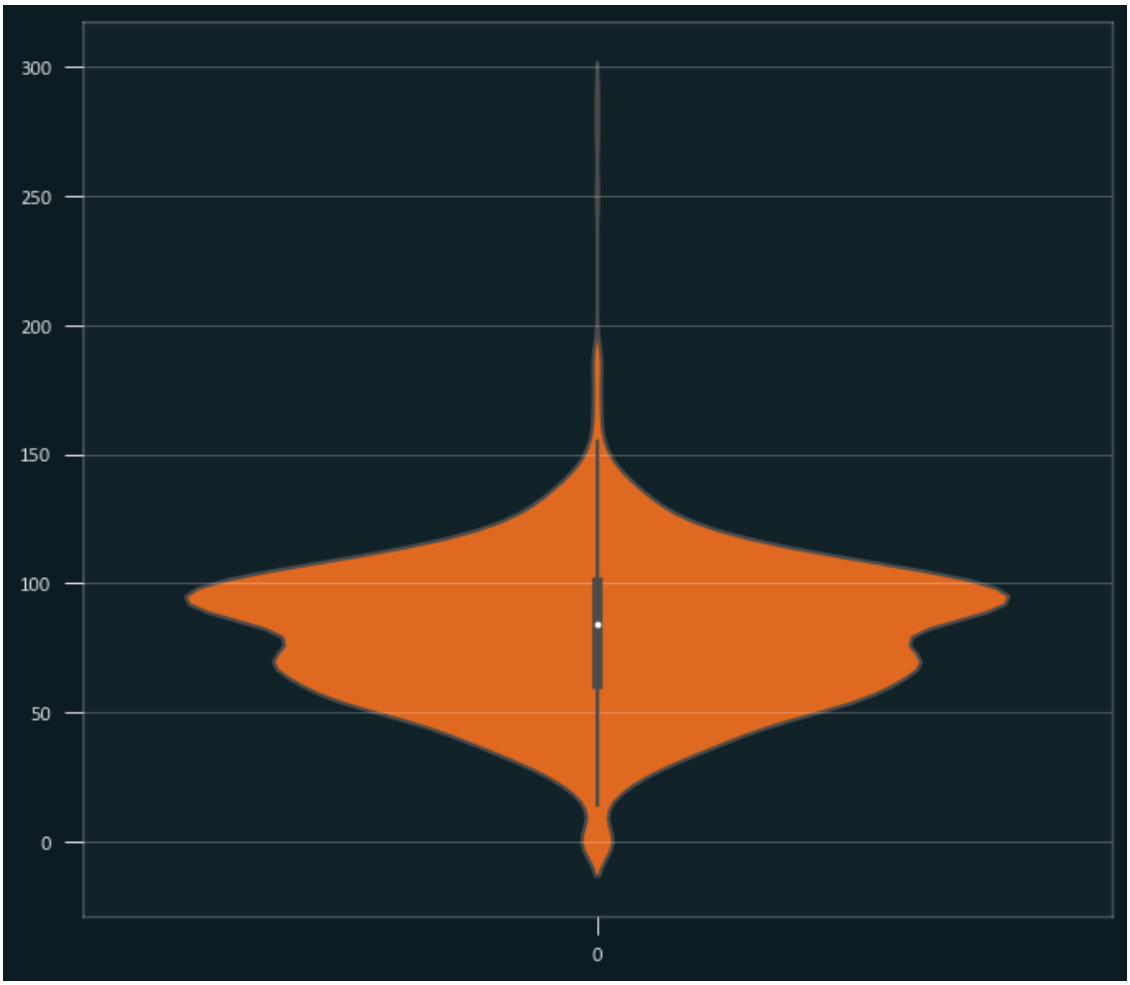


1.3.30 OD_PORC_SUP

[72]:
```
array(['46.8', '92', nan, '96.2', '90.1', '81.9', '89.2', '110.7',
       '106.3', '101.4', '102.4', '99.3', '118.4', '86.1', '87.2', '96.9',
       '86.7', '85.7', '76.4', '107.7', '86.5', '85.9', '90.4', '83.6',
       '93.6', '98.6', '114.9', '117.8', '130.6', '108.7', '71.3',
       '106.5', '111.3', '107.6', '108.9', '106', '101.7', '117.6',
       '93.5', '69.8', '101.9', '107.9', '107.4', '111.2', '114.2',
       '110.9', '115.2', '88.4', '90.9', '84.5', '88.5', '83.3', '82.3',
       '87.1', '96.5', '84', '80.4', '22.7', '63.1', '60.4', '56.3',
       '118.6', '70.5', '31.1', '61.2', '82.4', '83.7', '69.4', '61.3',
       '57.2', '59.3', '66.1', '63.2', '54.6', '52.5', '71.6', '47.9',
       '53.2', '58.3', '70.6', '120.1', '111.1', '73.3', '50.4', '83',
       '114', '108.8', '129.1', '119', '50.1', '99.1', '109.7', '105.8',
       '98.4', '103.9', '92.3', '92.1', '88.6', '93.1', '94.8', '105.9',
       '103.1', '68.7', '85', '100.9', '71.8', '38.5', '90.8', '142',
       '136.6', '52.7', '71.2', '93.2', '92.4', '113', '112.1', '<10',
```

'108.1', '30.8', '91.6', '92.2', '145.4', '74.8', '117.1', '109.3',
'146.1', '64.4', '187.7', '110.3', '119.4', '138.6', '189.3',
'97.6', '94.9', '108.2', '78.5', '76.1', '52.1', '97.8', '89.1',
'96.1', '104.5', '91', '96.7', '95.5', '100', '90.6', '98.7',
'103.4', '97.7', '132.9', '21.2', '37.1', '38.9', '33.4', '39.5',
'22.6', '33.2', '31.6', '124.1', '120', '112', '125.9', '78',
'112.6', '105.3', '89.7', '104.3', '99.6', '94.6', '39.8', '72.3',
'54.5', '60.6', '95.1', '35.8', '36.7', '112.2', '121.6', '101.6',
'111.6', '118.2', '125.1', '119.3', '105.1', '89.8', '97.9',
'67.4', '96.3', '95.7', '91.7', '103.5', '92.8', '104.8', '128.6',
'123.4', '87', '104.9', '91.2', '77.2', '46.3', '69', '68.9',
'71.5', '70.8', '70.4', '72.8', '70.2', '73.9', '68.4', '70.1',
'69.5', '69.7', '75.4', '77.4', '50.3', '56.6', '58.4', '60',
'53.1', '35.1', '81.2', '74.7', '61.1', '79.4', '61', '69.2',
'69.1', '72.1', '71.1', '70.3', '54.1', '56.2', '60.5', '63.6',
'52.8', '60.7', '52.4', '47.8', '54', '57.1', '63.3', '62.7',
'62.3', '63', '64', '56', '51', '58', '67.1', '68.5', '65.7',
'68.1', '69.9', '70.7', '87.6', '65.1', '42.2', '43', '39.6',
'40.2', '47.4', '26.6', '29', '32.1', '49.6', '49.9', '43.5',
'43.4', '69.6', '62', '72.4', '109', '137.3', '101', '123.9',
'23.3', '39.7', '42.4', '42.1', '51.7', '58.2', '51.6', '41.1',
'44.2', '84.9', '86.3', '33.9', '18.4', '39.4', '42.9', '108.4',
'85.4', '58.8', '40.6', '110.2', '95.2', '59.4', '79.5', '82.6',
'88.3', '41.3', '110.4', '29.2', '86.6', '100.7', '26.2', '26',
'52', '94.2', '95.6', '99.2', '101.8', '108.3', '111.7', '28.5',
'26.8', '27.1', '99', '104.2', '34.5', '53.3', '45.9', '107',
'126.3', '99.4', '125', '160.4', '91.4', '87.5', '96', '88.8',
'40.4', '100.6', '68.8', '66', '76.8', '86.4', '84.7', '102.3',
'93.4', '72.5', '73.5', '77.6', '96.4', '46.5', '274', '40.7',
'116.3', '50.8', '30.7', '105.5', '34.8', '30.9', '24.3', '66.5',
'107.1', '138.1', '177.3', '42.6', '99.9', '67', '27.7', '75.2',
'289', '28.4', '112.8', '109.4', '75.3', '137.8', '133.4', '26.7',
'77', '55.6', '146.4', '131.4', '46.1', '123', '87.7', '137.7',
'165.1', '89', '85.8', '88.1', '46.7', '57', '72.6', '31.2',
'22.4', '73.1', '101.1', '55.2', '95', '98.5', '100.1', '51.3',
'114.3', '105.6', '76.6', '122.7', '127', '122.1', '133.9',
'121.2', '123.6', '61.7', '65.5', '82.2', '78.7', '33', '36.6',
'42.5', '96.8', '60.8', '50.6', '50.2', '58.1', '51.5', '41.4',
'74.1', '37.8', '143.1', '32', '81.8', '250.9', '15.4', '111.5',
'104.1', '113.4', '102.1', '92.6', '30.2', '98.2', '82.5', '90.3',
'89.5', '98.8', '137.4', '103.2', '88.2', '115.3', '97.5', '76.7',
'87.3', '117', '79.6', '82.8', '76.2', '102', '59.9', '113.9',
'107.3', '103.6', '78.1', '115.8', '118.5', '131.1', '45.5',
'34.4', '129', '16', '49.3', '116.9', '64.1', '57.3', '48.4',
'67.3', '54.2', '44.1', '112.4', '64.8', '66.4', '68', '50', '72',
'84.1', '86', '92.9', '88.7', '77.5', '91.5', '97', '102.2',
'95.9', '116.5', '127.1', '93.3', '90.7', '82.1', '100.4', '97.2',

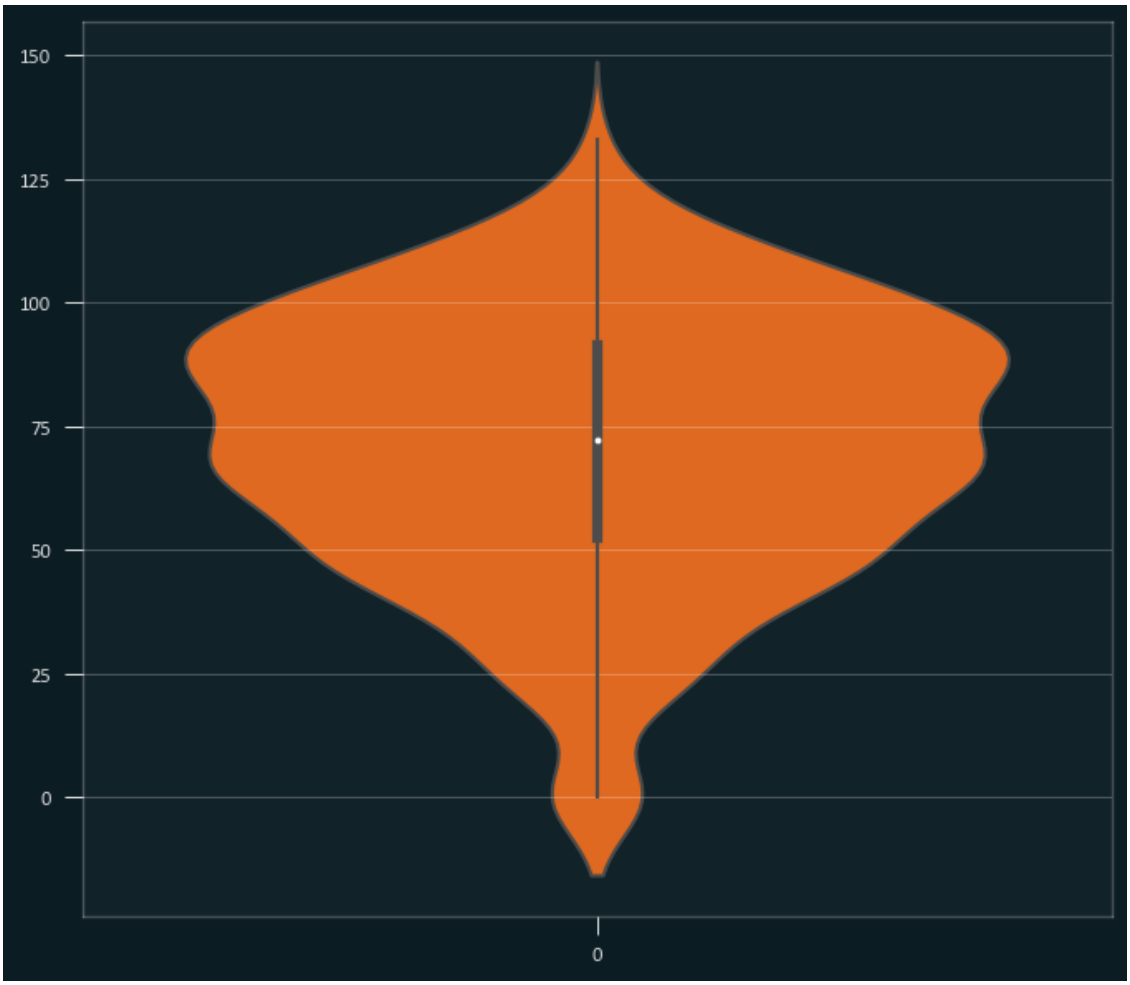
```
       '72.2', '64.2', '40.8', '45.4', '39.9', '41.6', '54.8', '44',
       '100.8', '95.8', '84.4', '119.2', '73.8', '37.2', '23.2', '80.5',
       '37.9', '106.2', '96.6', '123.1', '122.3', '122', '124.3', '112.9',
       '121.5', '45.2', '86.2', '97.3', '75.6', '103.3', '89.9', '78.4',
       '85.2', '80.8', '90.2', '79.1', '83.2', '75', '73.2', '88', '78.9',
       '55.7', '75.9', '75.1', '69.3', '90', '91.1', '72.7', '81.3',
       '71.9', '47.2', '63.7', '65.4', '62.1', '73.4', '77.9', '80.2',
       '78.6', '67.2', '80.1', '76.3', '80.6', '76', '54.3', '44.3',
       '80.9', '63.4', '86.8', '84.6', '66.7', '60.1', '52.6', '49.4',
       '49.7', '22.1', '93', '43.2', '135.4', '132.1', '138.9', '120.2',
       '79.9', '140.2', '28', '102.6', '34', '30', '47', '40', '70', '71',
       '91.8', '51.2', '73', '74', '62.6', '80.7', '102.7', '74.6',
       '17.9', '25.2', '94.5', '67.8', '97.1', '91.3', '68.2', '59',
       '53.7', '128.1', '126.7', '79.3', '134.3', '125.7', '106.9',
       '85.6', '48', '82.7', '42', '55', '49', '35', '39', '94.3', '94.1',
       '94.4', '65', '95.3', '83.4', '48.1', '53', '45', '54.7', '80.3',
       '73.6', '46', '51.8', '94', '98', '83.9', '89.6', '60.9', '48.3',
       '21.4', '110.1', '123.2', '126', '104.4', '52.2', '60.2', '36.9',
       '115', '37', '31.7', '106.6', '88.9', '100.2', '131', '136',
       '93.8', '125.8', '98.1', '106.4', '43.7', '101.3', '110', '123.5',
       '133', '103.8', '133.3', '98.9', '102.9', '107.2', '99.8', '13.9',
       '34.9', '66.6', '80', '91.9', '65.8', '111', '130', '100.5',
       '117.2', '181.6', '154.7', '185.6', '101.2', '102.5', '38.8',
       '100.3', '99.7', '87.4', '151.4', '103', '108.5', '169.9', '168.8',
       '127.9', '141.7', '130.5', '117.9', '139.2', '126.1', '51.9',
       '115.4', '105', '128.2', '99.5', '151.5', '144.9', '152.4', '86.9',
       '87.9', '89.4', '76.9', '118.9', '81.6', '84.2', '45.7', '64.7',
       '81.4', '66.9', '108', '89.3', '83.8', '140.5', '144.8', '71.4',
       '119.9', '118.7', '93.9', '81', '92.7', '61.6', '51.4', '64.9',
       '119.5', '35.6', '20.1', '38.4', '54.4', '97.4', '136.7', '118.3',
       '59.2', '56.4', '55.1', '58.6', '28.1', '65.9', '66.8', '53.4',
       '60.3', '55.4', '58.5', '56.8', '61.4', '57.6', '59.6', '50.5',
       '107.8', '114.1', '116.7', '105.2', '114.6', '109.2', '104.7',
       '34.3', '65.6', '62.5', '74.2', '67.5', '44.6', '57.4', '40.5',
       '53.5', '53.8', '53.6', '48.8', '116.2', '84.3'], dtype=object)

Percentage of unique value to total: 0.00926497838171711
Values distribution: count    1619.000000
mean       81.367326
std        28.809245
min         0.000000
25%        61.200000
50%        84.100000
75%        99.950000
max       289.000000
Name: OD_PORC_SUP, dtype: float64
```



1.3.31 OD_PORC_MED

[74]:
```
array([nan, '95.4', '95.9', '95.2', '82.4', '108.6', '106.7', '99.7',
       '102.4', '99.6', '114.5', '116.2', '105.5', '83.1', '93.2', '89.8',
       '95.5', '100.8', '133', '106', '109.7', '101.2', '101.1', '108',
       '103.2', '98.9', '108.9', '103', '107.8', '107.5', '111.3',
       '113.9', '116.4', '68.9', '40.3', '86.1', '89.1', '89.2', '92.4',
       '96.6', '94.8', '112.3', '111.6', '97.3', '77.4', '76', '46',
       '35.5', '39.3', '47.7', '40.7', '88.4', '98.4', '94.2', '47.8',
       '27.4', '22.7', '21.7', '19.1', '19.3', '19.7', '54.1', '90.9',
       '102.8', '55.6', '<10', '69.6', '49.1', '69.4', '87.4', '89.5',
       '42.8', '83.7', '33.8', '103.8', '69.3', '68.4', '69.7', '69.1',
       '68.6', '67.8', '68.2', '51', '56.4', '79.3', '71.8', '70.2',
       '59.6', '66', '77.1', '59.8', '60.2', '63', '64.9', '69.8', '68.1',
       '69.9', '36.4', '65.5', '65.4', '66.8', '61.3', '64.8', '67.3',
       '59.1', '64.2', '66.4', '29.9', '25', '31.7', '79.5', '74.2',
       '80.4', '60.3', '45.6', '43.6', '45.3', '46.4', '46.3', '80.6',
```

```
       '46.9', '46.7', '91.4', '93.1', '90.1', '107', '110.5', '103.6',
       '104.3', '48', '70.3', '91.2', '75.6', '64.7', '73.1', '90',
       '92.7', '83.9', '41.4', '37.4', '41.6', '39.1', '29.5', '46.5',
       '29.7', '22.9', '23.7', '37.8', '42.9', '47', '42.6', '20.4',
       '39.5', '98.5', '106.1', '48.6', '112.9', '81.8', '88.3', '87',
       '81.6', '82.2', '127.1', '114.9', '73.9', '83.4', '56.1', '93.8',
       '91.1', '76.8', '67.5', '46.2', '38.4', '38.6', '31.5', '98.1',
       '65.8', '63.8', '60', '71', '70.6', '72.4', '71.6', '74.6', '72.9',
       '25.1', '73', '24.8', '38.3', '39.6', '42.3', '54.6', '52.4',
       '23.1', '83.5', '45.7', '33.2', '67.4', '67.6', '71.1', '71.9',
       '81.1', '90.3', '84.1', '58.9', '55.1', '50.2', '51.3', '50.1',
       '88.5', '85.4', '86.8', '88.1', '63.4', '65.2', '64', '63.9',
       '59.9', '58.1', '48.3', '82.3', '84.2', '74.5', '81.5', '88.2',
       '90.6', '78.7', '78.5', '43.1', '43.3', '47.6', '47.2', '87.1',
       '88.7', '112.6', '96.4', '104.1', '60.5', '49.3', '61.8', '96.5',
       '60.7', '58', '51.7', '62.3', '107.7', '78', '59', '80', '60.6',
       '44', '52', '55', '100.1', '53', '49', '94', '61', '101', '117',
       '31.3', '33.5', '35.6', '35.2', '33.1', '97.1', '101.3', '102',
       '97.2', '96.8', '95.1', '83.3', '113.1', '72.1', '37.2', '91.6',
       '103.3', '99.2', '77.9', '99.3', '101.9', '121', '98.7', '90.2',
       '121.1', '118', '105.1', '97.7', '73.7', '84.4', '100.6', '57.5',
       '115', '83.8', '108.4', '21.3', '25.7', '23.2', '27', '19.9',
       '24.2', '26.9', '93.3', '96.2', '99.9', '50.6', '28.3', '109.8',
       '66.5', '60.9', '45.4', '93.7', '96.7', '87.2', '96.3', '68.5',
       '71.4', '75.8', '87.6', '81.3', '84.3', '84.6', '58.8', '96',
       '88.8', '85.8', '57.9', '57.3', '77.6', '75.1', '89.3', '64.4',
       '86.3', '82.9', '80.2', '100.7', '78.9', '65.3', '53.2', '80.1',
       '86.2', '88.9', '87.5', '79.6', '126.9', '124.3', '54.2', '49.4',
       '48.4', '44.9', '41.1', '53.4', '52.8', '44.6', '50.7', '47.9',
       '52.1', '63.2', '95.3', '83.6'], dtype=object)

Percentage of unique value to total: 0.028747433264887063
Values distribution: count    487.000000
mean      71.303491
std       27.086245
min        0.000000
25%       52.400000
50%       72.400000
75%       91.500000
max      133.000000
Name: OD_PORC_MED, dtype: float64
```

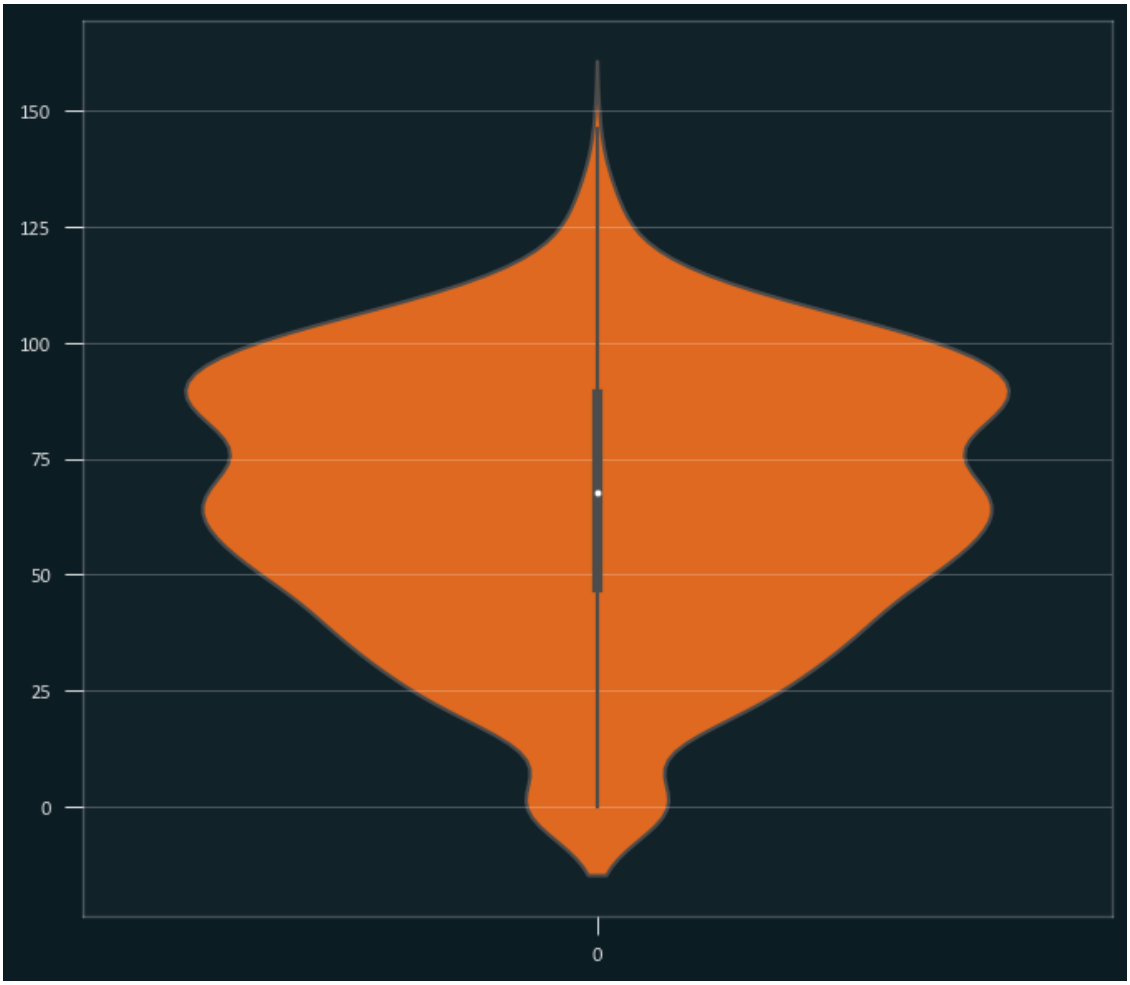


1.3.32 OD_PORC_FON

[76]:
```
array([nan, '92.2', '86.7', '95.5', '94.2', '79.2', '146', '101.6',
       '107.9', '104.9', '106.4', '103.2', '99.6', '104.2', '105.6',
       '114.2', '128.2', '118', '88.5', '105', '81.4', '111.9', '106.1',
       '99.5', '80.1', '102.3', '106', '99.2', '83.7', '84.2', '87.2',
       '92.9', '97.8', '131.4', '90.1', '112.2', '109.5', '92.4', '89.4',
       '92.8', '106.7', '98.6', '96.4', '117.3', '106.3', '122.6', '34.3',
       '89.7', '82', '102.9', '102.4', '107', '110.9', '101.4', '102.6',
       '113.2', '110.7', '89.6', '87.8', '82.1', '84', '80.4', '85.5',
       '70.5', '74.3', '36.6', '59.1', '39.8', '81.7', '67.7', '58.9',
       '59', '60.3', '68', '47.5', '69.8', '118.6', '35.1', '109', '83',
       '126.5', '85.7', '98.2', '98.4', '66.9', '78.4', '90.4', '93.1',
       '91.4', '95.1', '72.1', '35.5', '111.8', '102.8', '107.8', '101.3',
       '90.7', '91', '108.2', '85.6', '36', '60.7', '35.7', '26.4',
       '31.8', '31.7', '36.3', '73.8', '107.2', '19.2', '82.2', '86.2',
       '46', '21.8', '24.3', '16.5', '20.5', '15.4', '16.8', '14.9',
```

```
'20.2', '79.4', '26', '29.6', '27.3', '<10', '32.1', '17.3',
'112.6', '100.1', '18.3', '58.3', '42.3', '43.7', '42.9', '40',
'38.1', '40.9', '28', '56.2', '73.9', '94.9', '94.3', '90.3',
'70.2', '67.6', '66.7', '67.3', '67.4', '65.9', '66.4', '68.9',
'68.8', '65.2', '66.3', '67.9', '67.1', '45.1', '50', '50.1',
'68.5', '70.1', '47.4', '66.2', '52.4', '59.9', '69', '68.1',
'50.4', '51.1', '60', '64.9', '49.4', '68.6', '65.8', '63.8',
'47.8', '24.8', '55.6', '57.2', '33.7', '38.2', '46.3', '48.9',
'55.4', '49', '63.4', '27.5', '60.5', '65.4', '67', '51.7', '54.6',
'52.7', '28.3', '31.2', '39.1', '35', '67.2', '41.5', '22.2',
'42.7', '49.3', '20.7', '31', '38.4', '15.6', '25.6', '20.1',
'12.1', '87.5', '64.7', '59.2', '30.3', '20.3', '32.6', '23.4',
'20.4', '22.3', '20.8', '50.2', '87.4', '43', '35.2', '19.6',
'53.9', '46.5', '73.3', '75.8', '106.8', '109.3', '103.1', '102.7',
'44.1', '48.1', '52.6', '38', '71.7', '77.7', '66.1', '56.7',
'32.8', '48.6', '61.7', '55.9', '58.2', '52.5', '51.4', '79', '81',
'54.3', '33.5', '27.1', '35.6', '20.6', '35.4', '40.4', '45.2',
'95.3', '60.9', '23.7', '43.1', '99.9', '74.8', '40.5', '52.3',
'79.3', '66', '87.7', '58.1', '65.1', '71.1', '74.1', '80.7',
'95.7', '71', '80.2', '25.1', '61.5', '68.2', '22.6', '33', '73.2',
'84.8', '42.4', '53.4', '56', '40.8', '74.7', '64.3', '60.6',
'40.7', '94', '58.7', '39.6', '48.8', '42.1', '24.9', '37.9',
'33.4', '34.8', '32.5', '29.1', '71.6', '99.8', '97.1', '101.9',
'73.1', '43.3', '91.3', '91.6', '98.7', '106.2', '108.9', '96.8',
'90', '116.7', '116.3', '64.8', '58.5', '56.1', '58.4', '62.1',
'58', '61', '64', '75.3', '83.1', '100.5', '85', '93.4', '98',
'100.8', '101.2', '97.9', '57.5', '78.3', '49.2', '50.6', '55.5',
'49.5', '36.4', '30.6', '22.1', '21.6', '23.6', '19.5', '103.5',
'17.7', '40.3', '39.3', '22', '21.7', '34.9', '23.8', '21.4',
'70.9', '52.2', '48.4', '64.2', '59.6', '57.4', '77.3', '76.9',
'43.2', '28.4', '37.2', '23.5', '24.1', '88.8', '86.6', '88',
'80.9', '58.8', '56.3', '53.5', '76.4', '61.1', '66.5', '62.3',
'34.7', '83.4', '85.1', '84.5', '57.1', '76.8', '37.7', '82.4',
'38.9', '39.2', '41.4', '88.6', '87.1', '37.1', '102.2', '61.9',
'56.6', '48.5', '136.5', '14.1', '87.3', '96', '131.6', '62',
'70.6', '88.9', '55', '68.7', '49.7', '57', '61.3', '127.8',
'132.4', '71.2', '119.7', '62.7', '104.1', '39', '88.7', '72',
'52', '74', '53', '73', '63', '85.4', '95.4', '89.2', '94.5',
'95.8', '101.1', '48', '51', '66.6', '70', '90.2', '39.9', '33.9',
'110', '86.4', '78.6', '27.9', '29.7', '29.2', '30.5', '94.1',
'82.6', '97.3', '98.3', '41.6', '92.1', '94.7', '92', '81.3', '93',
'117.4', '120', '100.3', '90.9', '70.7', '97.7', '96.2', '86.3',
'77.1', '74.2', '77.2', '69.5', '61.6', '52.9', '41.7', '36.1',
'30.7', '82.5', '77.6', '64.6', '107.1', '78.1', '79.1', '100.7',
'99.4', '18.8', '21.1', '26.7', '25.3', '23.3', '93.5', '29.5',
'26.3', '78.9', '84.6', '51.6', '97', '81.6', '85.8', '91.8',
'85.3', '86.1', '84.9', '86.8', '83.5', '86.9', '94.8', '96.3',
```

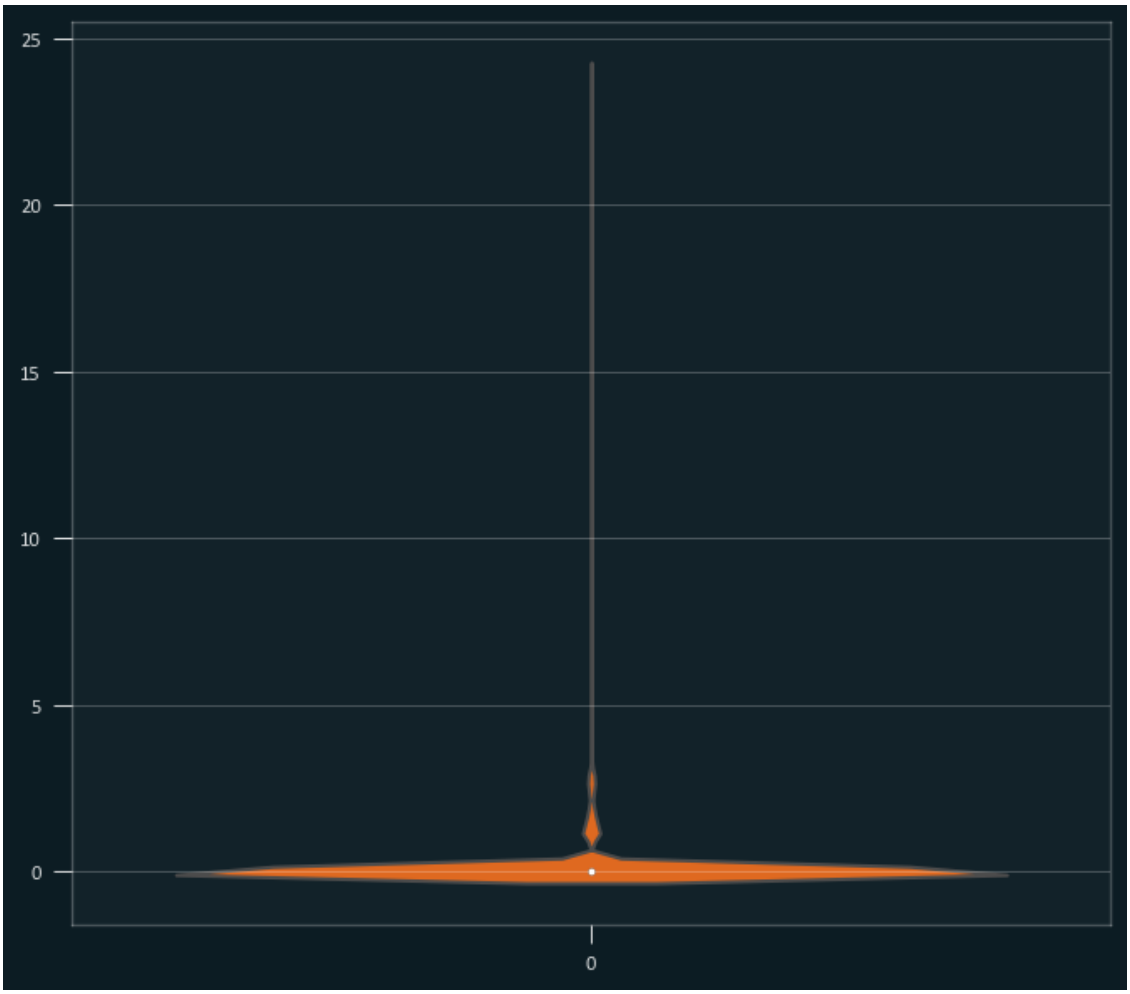
```
       '77.5', '86', '56.8', '128.4', '113.6', '112.8', '98.5', '74.9',
       '88.1', '78.5', '64.5', '72.6', '53.2', '82.7', '99.3', '34.6',
       '29.3', '76.3', '100.2', '97.6', '94.6', '48.7', '71.5', '46.7',
       '56.4', '48.2', '44.3', '42.5', '42', '49.6', '51.2', '50.3',
       '47.3', '45.3', '91.9', '81.9', '92.7', '89.9', '89.5', '98.9',
       '30', '31.6', '54.4', '73.6', '63.2', '72.7', '65.7', '73.5',
       '114.9', '63.3', '55.8', '69.2', '73.4', '93.8'], dtype=object)

Percentage of unique value to total: 0.03699788583509514
Values distribution: count    946.000000
mean      66.499260
std       28.996016
min        0.000000
25%       47.325000
50%       67.500000
75%       88.875000
max      146.000000
Name: OD_PORC_FON, dtype: float64
```



1.3.33 TOX_D_48_UT

[78]:
```
array([nan, '<1', '2.295', '2.699', '2.937', '1.696', '1.182', '1.184',
       '2.305', '3.546', '1.205', '1.881', '1.84', '1.387', '2.869',
       '2.759', '2.902', '2.584', '1.741', '1.658', '3.062', '1.639',
       '2.466', '2.783', '2.644', '1.919', '2.85', '1.272', '2.959',
       '2.317', '4.6', '1.318', '1.761', '2.625', '2.237', '1.34',
       '1.085', '1.04', '2.475', '1.645', '1.12', '1.35', '1.25', '4.474',
       '1.083', '1.08', '5.218', '1.533', '2.622', '2.332', '1.488',
       '2.717', '1.338', '1.148', '1.68', '1.337', '1.065', '1.374',
       '1.971', '1.191', '1.1', '2.73', '1.202', '10.627', '1.158',
       '1.095', '1.254', '1.047', '1.072', '1.211', '1.069', '1.55',
       '1.235', '1.174', '1.153', '6.63', '1.249', '1.003', '2.586',
       '1.644', '23.949', '1.036', '3.111', '1.586', '1.268', '1.392'],
      dtype=object)
```

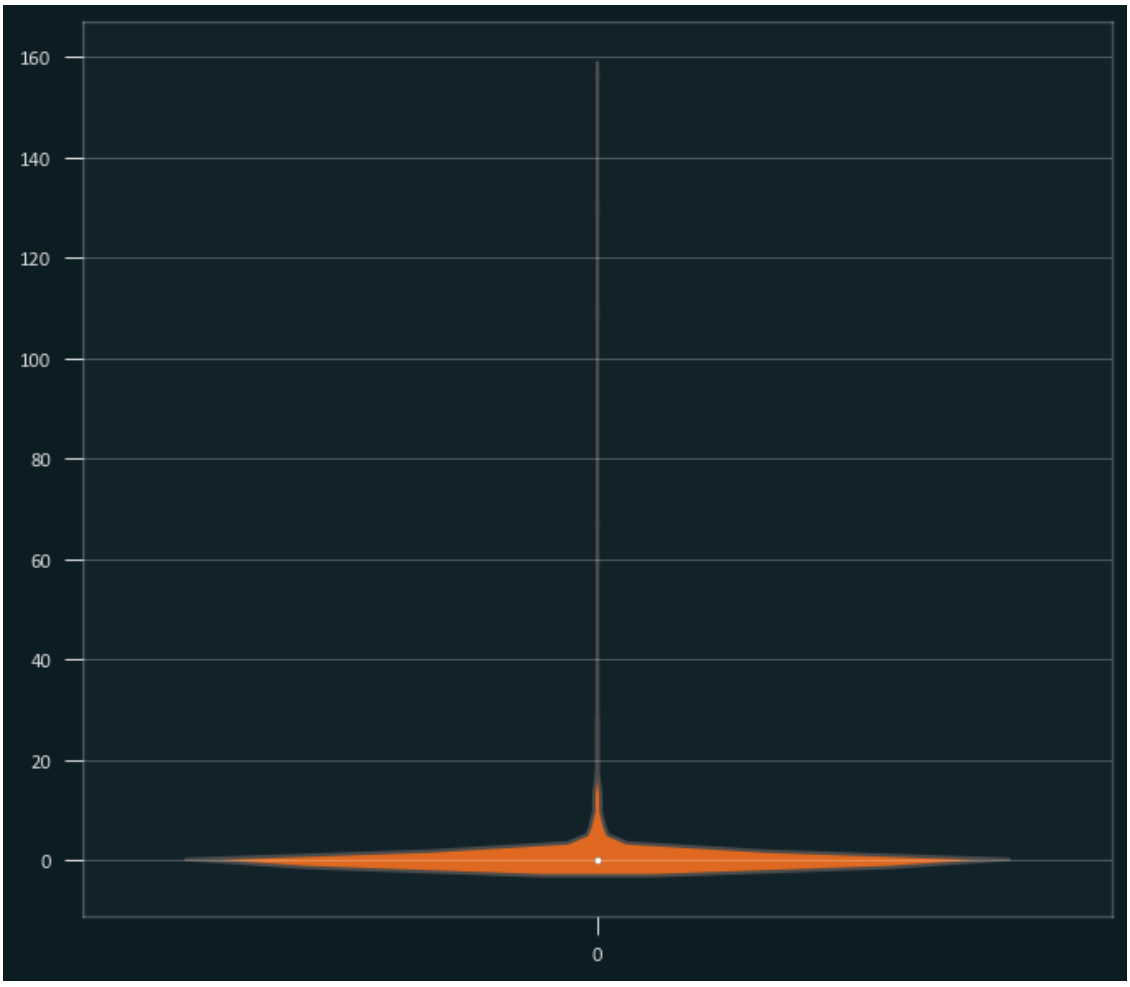
```
Percentage of unique value to total: 0.9520925110132159
Values distribution: count    1816.000000
mean        0.109529
std         0.767205
min         0.000000
25%         0.000000
50%         0.000000
75%         0.000000
max        23.949000
Name: TOX_D_48_UT, dtype: float64
```



1.3.34 TOX_V_15_UT

[80]:
```
array([nan, '<1', '1.651', '6.781', '1.602', '5.704', '1.859', '1.838',
       '4.701', '4.556', '5.339', '3.183', '4.223', '15.533', '8.271',
       '7.158', '6.649', '4.395', '8.866', '8.489', '2.146', '3.593',
       '3.095', '3.807', '3.225', '6.025', '2.612', '3.25', '20.317',
       '4.456', '59.844', '66.934', '156.25', '1.637', '2.36', '1.022',
       '3.817', '8.325', '3.626', '1.357', '28.209', '3.341', '21.968',
       '3.411', '4.973', '2.808', '1.764', '8.823', '3.004', '1.239',
       '6.729', '11.896', '11.85', '17.59', '12.56', '4.19', '5.8',
       '14.46', '9.12', '4.95', '24.65', '24.82', '1.73', '17.746',
       '11.75', '11.46', '14.11', '8.96', '22.04', '8.183', '8.46',
       '5.06', '6.964', '1.126', '13.708', '4.257', '3.654', '7.74',
       '1.804', '5.593', '7.283', '3.068', '3.667', '4.677', '4.733',
       '3.255', '3.126', '4.9', '7.7', '12.39', '1.566', '2.232', '1.93',
       '5.757', '11.07', '2.63', '2.52', '1.54', '6.84', '6.72', '38.8',
```

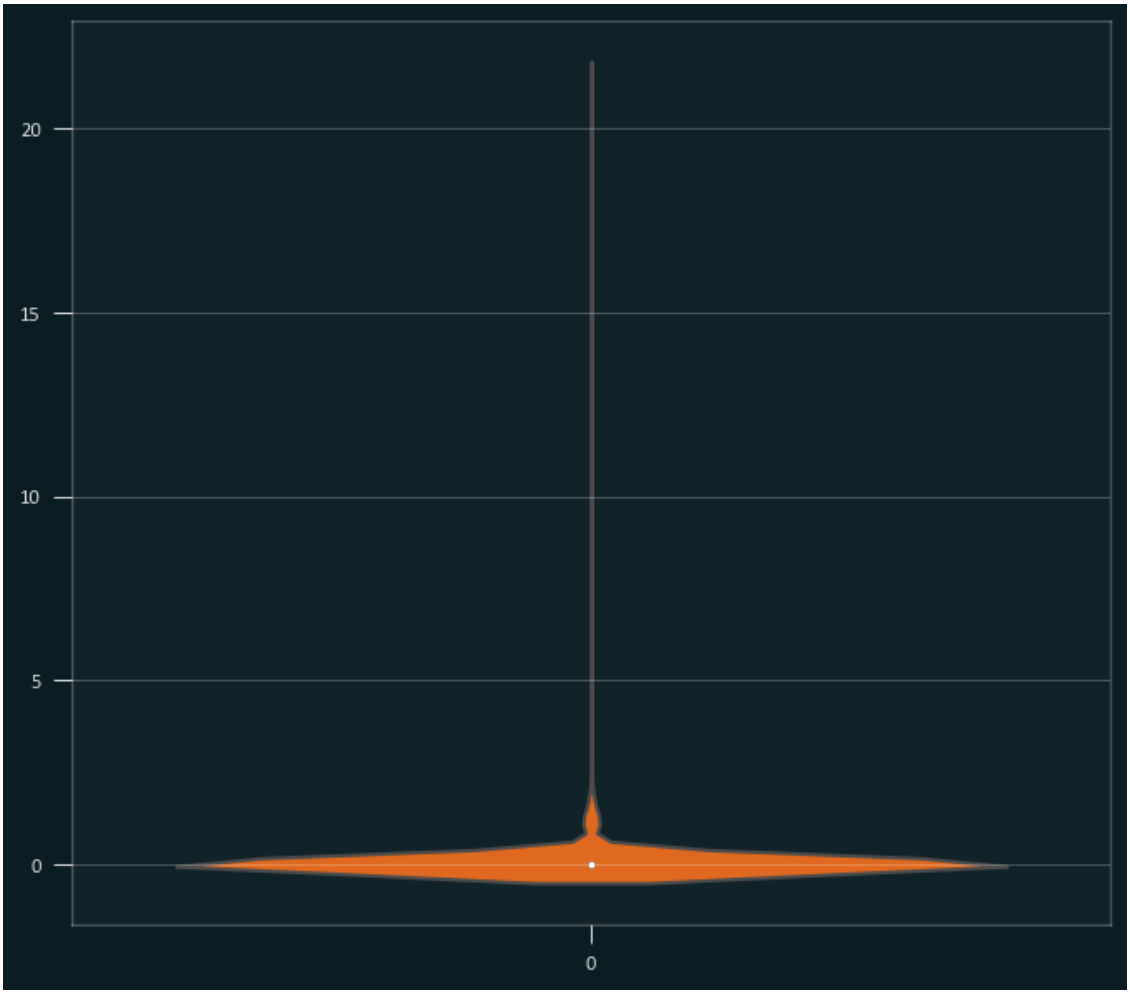
```
       '2.246', '1.874', '1.854', '2.739', '5.9', '4.378', '12.997',
       '3.336', '7.117', '10.797', '3.113', '2.931', '1.305', '2.803',
       '27.824', '1.639', '12.439', '35.398', '27.617', '14.93', '2.734',
       '109.29', '15.359', '4.047', '31.786', '1.641', '17.179', '1.153',
       '1.206', '1.01', '1.119', '1.791', '2.27', '2.546', '1.281',
       '1.131', '2.854', '13.774', '5.238', '14.824', '7.806', '12.718',
       '11.206', '13.011', '130.039', '5.534', '4.847', '8.163'],
      dtype=object)

Percentage of unique value to total: 0.9186366135239142
Values distribution: count    1819.000000
mean        0.877028
std         6.447175
min         0.000000
25%         0.000000
50%         0.000000
75%         0.000000
max       156.250000
Name: TOX_V_15_UT, dtype: float64
```



1.3.35 TOX_D_48_SUP_UT

[82]:
```
array(['<1', nan, '1.86', '2.97', '1.67', '1.13', '1.3', '1.04', '1.01',
       '1.64', '1.71', '1.11', '1.29', '1.12', '1.23', '1.14', '1.15',
       '21.32', '13.39', '2.39', '1.08', '1.05', '1.16', '1.58'],
      dtype=object)
```

```
Percentage of unique value to total: 0.968503937007874
Values distribution: count    762.000000
mean       0.086142
std        0.942977
min        0.000000
25%        0.000000
50%        0.000000
75%        0.000000
max       21.320000
Name: TOX_D_48_SUP_UT, dtype: float64
```



1.3.36 TOX_D_48_FON_UT

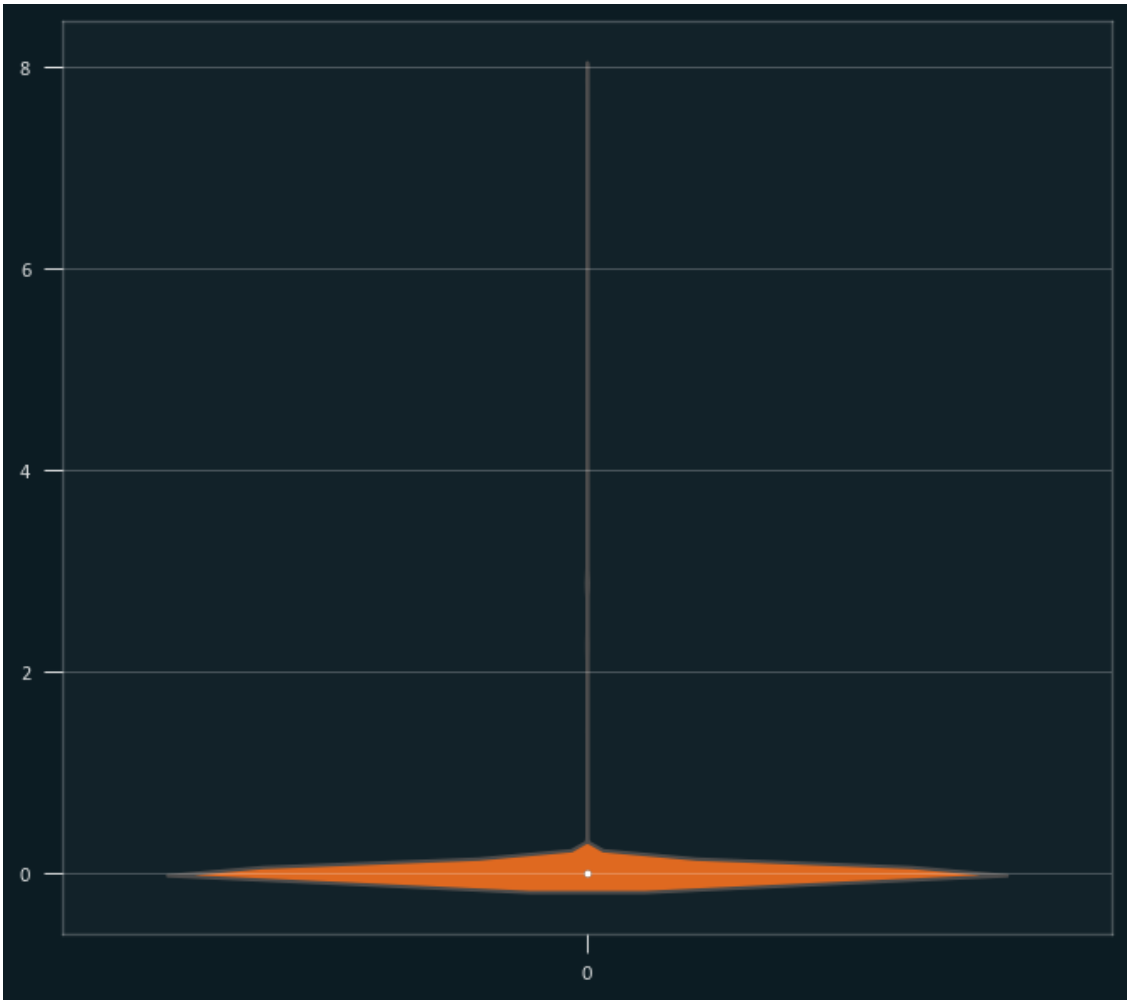
[84]:
```
array([nan])
```

La columna está totalmente vacía, por lo que se procede a eliminar del conjunto de datos.

1.3.37 TOX_FIS_SUP_15_UT

[86]:
```
array(['<1', nan, '2.87', '2.13', '1.09', '7.87', '2.89', '3.04', '5.67',
       '4.23', '7.86', '1.4', '1.46', '2.21', '2.81', '2.36', '5.57',
       '2.32'], dtype=object)
```

```
Percentage of unique value to total: 0.9898446833930705
Values distribution: count    1674.000000
mean        0.035048
std         0.402071
min         0.000000
25%         0.000000
50%         0.000000
75%         0.000000
max         7.870000
Name: TOX_FIS_SUP_15_UT, dtype: float64
```



[88]: Index(['TOX_FIS_FON_15_UT'], dtype='object')

1.3.38 TOX_FIS_FON_15_UT

[89]: array([nan])

La columna está vacía, por lo que se procede a eliminarla del conjunto de datos.

[91]: 178143

Una vez limpia la base de datos, podemos convertir a los tipos dee datos que corresponden a cada columna.

[92]:
```
CLAVE                  object
SITIO                  object
ORGANISMO_DE_CUENCA    object
ESTADO                 object
MUNICIPIO              object
```

```
CUENCA                     object
CUERPO DE AGUA             object
TIPO                       object
SUBTIPO                    object
LONGITUD                  float64
LATITUD                   float64
PERIODO                   float64
DBO_mg/L                  float64
CALIDAD_DBO                object
DQO_mg/L                  float64
CALIDAD_DQO                object
SST_mg/L                  float64
CALIDAD_SST                object
COLI_FEC_NMP_100mL        float64
CALIDAD_COLI_FEC           object
E_COLI_NMP_100mL          float64
CALIDAD_E_COLI             object
ENTEROC_NMP_100mL         float64
CALIDAD_ENTEROC            object
OD_PORC                   float64
CALIDAD_OD_PORC            object
OD_PORC_SUP               float64
CALIDAD_OD_PORC_SUP        object
OD_PORC_MED               float64
CALIDAD_OD_PORC_MED        object
OD_PORC_FON               float64
CALIDAD_OD_PORC_FON        object
TOX_D_48_UT               float64
CALIDAD_TOX_D_48           object
TOX_V_15_UT               float64
CALIDAD_TOX_V_15           object
TOX_D_48_SUP_UT           float64
CALIDAD TOX_D_48_SUP       object
TOX_FIS_SUP_15_UT         float64
CALIDAD_TOX_FIS_SUP_15     object
SEMAFORO                   object
CONTAMINANTES              object
CUMPLE_CON_DBO             object
CUMPLE_CON_DQO             object
CUMPLE_CON_SST             object
CUMPLE_CON_CF              object
CUMPLE_CON_E_COLI          object
CUMPLE_CON_ENTEROC         object
CUMPLE_CON_OD              object
CUMPLE_CON_TOX             object
GRUPO                      object
dtype: object
```

1.3.39 Codificación de los datos categóricos

```
{'Amarillo': 0, 'Rojo': 1, 'Verde': 2}
```

[94]:
```
<bound method Series.unique of 0       1
1       2
2       2
3       2
4       2
       ..
3488    0
3489    1
3490    1
3491    0
3492    0
Name: SEMAFORO, Length: 3493, dtype: int32>
```

1.4 Exploración de los Datos

1.4.1 Medidas de Tendencias Central

- Promedio (mean)
- Mediana (50%) ### Medidas de Dispersión
- Desviación Estándard (std)
- Máximo (max)
- Mínimo (min) ### Medidas de Posición No Centrales
- Cuartiles (25%, 50%, 75%)
- Outliers (Boxplot)

[95]:

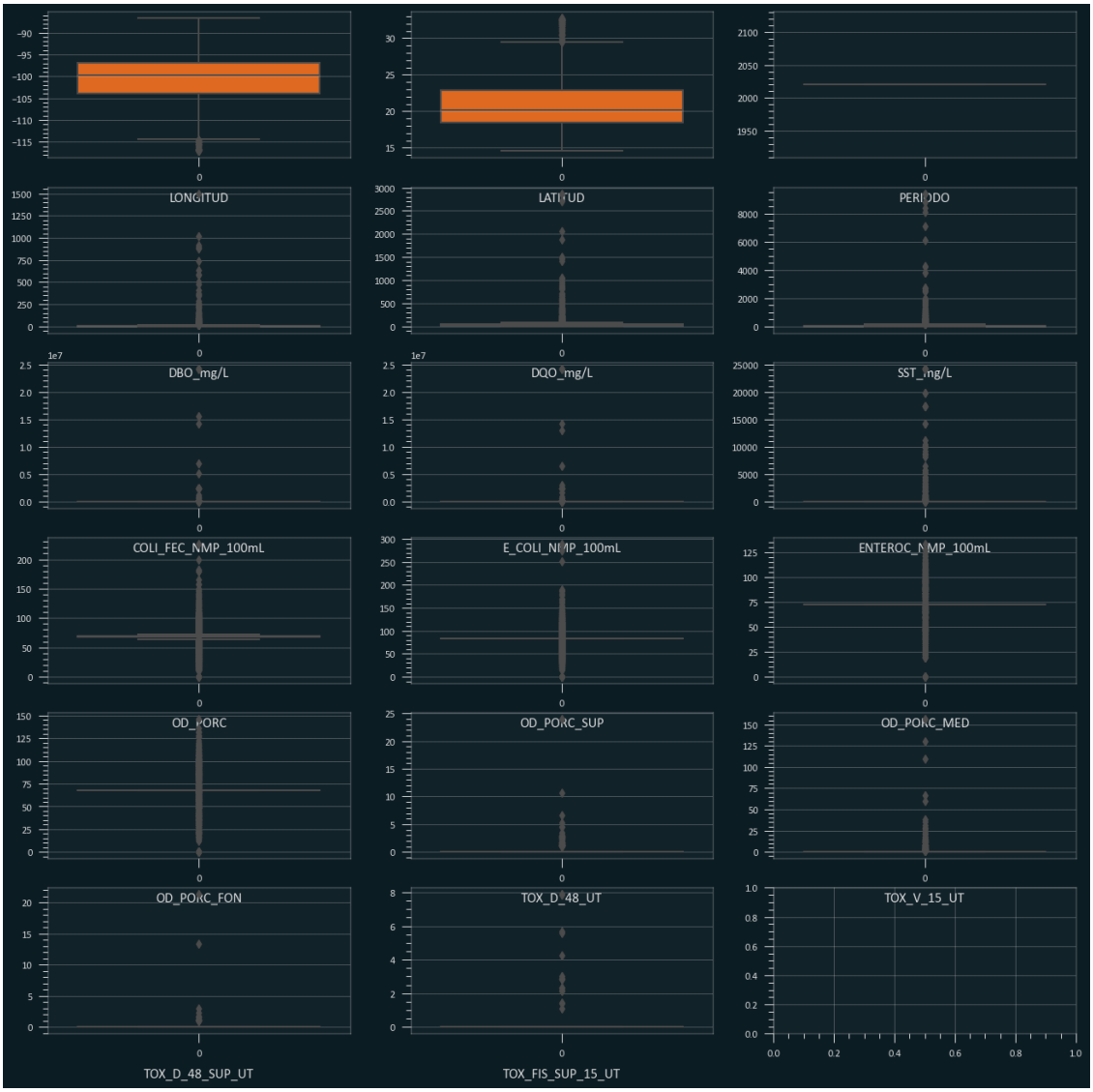
	LONGITUD	LATITUD	PERIODO	DBO_mg/L	DQO_mg/L	\
count	3493.000000	3493.000000	3493.0	3493.000000	3493.000000	
mean	-100.359969	21.046992	2020.0	12.463374	52.987848	
std	6.122773	3.893696	0.0	56.487492	130.424856	
min	-117.124030	14.534910	2020.0	0.000000	0.000000	
25%	-103.882310	18.396070	2020.0	0.000000	16.740000	
50%	-99.795530	20.148980	2020.0	2.630000	27.010000	
75%	-96.860230	22.828930	2020.0	6.760000	42.620000	
max	-86.732150	32.706500	2020.0	1500.000000	2871.250000	

	SST_mg/L	COLI_FEC_NMP_100mL	E_COLI_NMP_100mL	ENTEROC_NMP_100mL	\
count	3493.000000	3.493000e+03	3.493000e+03	3493.000000	
mean	99.539672	7.135831e+04	5.875611e+04	280.845605	
std	442.161444	1.005750e+06	9.045168e+05	2240.832128	
min	0.000000	0.000000e+00	0.000000e+00	0.000000	
25%	0.000000	8.550000e+02	9.000000e+01	0.000000	
50%	24.300000	2.400000e+03	4.240000e+02	0.000000	
75%	57.000000	1.100000e+04	2.400000e+03	0.000000	
max	9430.000000	2.419600e+07	2.419600e+07	24196.000000	

```
            OD_PORC  OD_PORC_SUP  OD_PORC_MED  OD_PORC_FON  TOX_D_48_UT  \
count   3493.000000  3493.000000  3493.000000  3493.000000  3493.000000
mean      67.234125    82.833410    72.247123    67.228972     0.056944
std       23.398534    19.657592    10.111986    15.090577     0.555813
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%       67.400000    84.100000    72.400000    67.500000     0.000000
50%       68.300000    84.100000    72.400000    67.500000     0.000000
75%       69.400000    84.100000    72.400000    67.500000     0.000000
max      226.100000   289.000000   133.000000   146.000000    23.949000

        TOX_V_15_UT  TOX_D_48_SUP_UT  TOX_FIS_SUP_15_UT    SEMAFORO
count   3493.000000      3493.000000        3493.000000  3493.00000
mean       0.456717         0.018792           0.016796     1.03779
std        4.672482         0.441642           0.278850     0.82851
min        0.000000         0.000000           0.000000     0.00000
25%        0.000000         0.000000           0.000000     0.00000
50%        0.000000         0.000000           0.000000     1.00000
75%        0.000000         0.000000           0.000000     2.00000
max      156.250000        21.320000           7.870000     2.00000
```

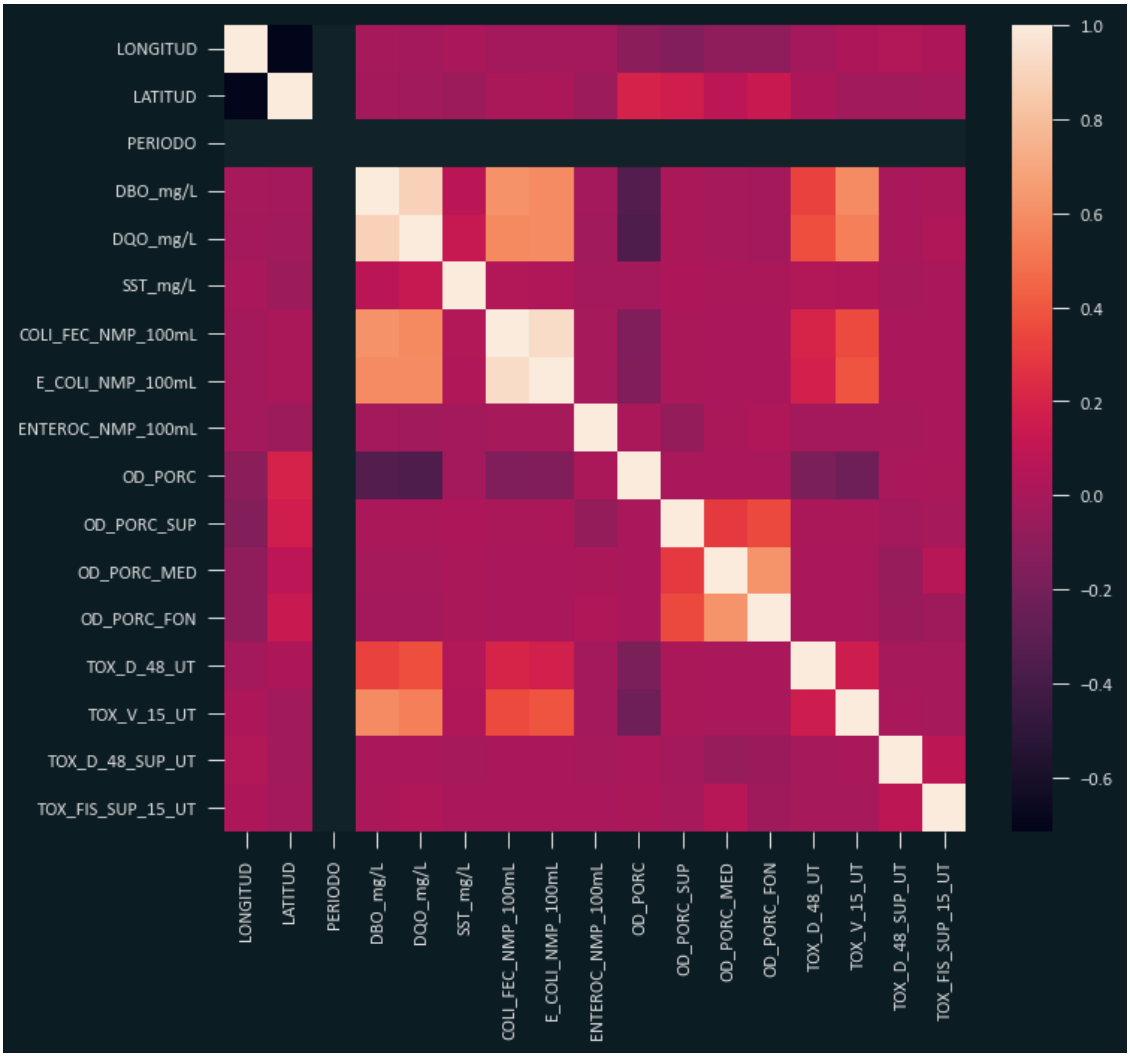
1.4.2 Boxplot

[96]: (17,)



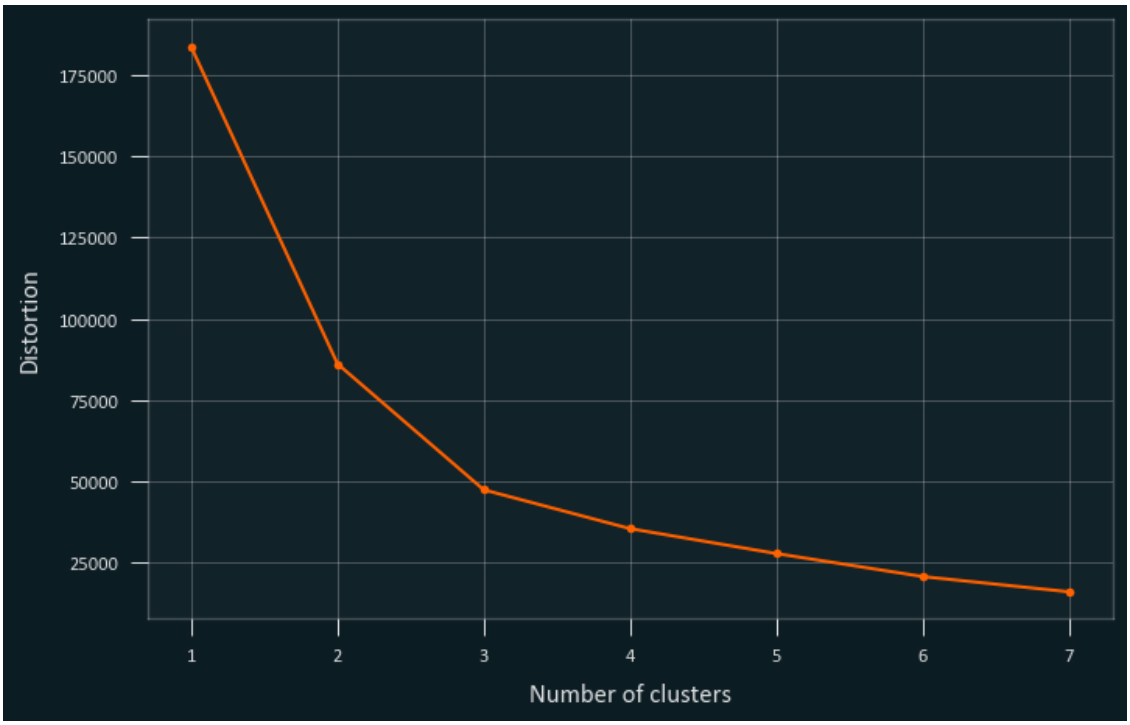
1.5 Matriz de correlación

[98]: <AxesSubplot:>


LONGITUD
LATITUD
PERIODO
DBO_mg/L
DQO_mg/L
SST_mg/L
COLI_FEC_NMP_100mL
E_COLI_NMP_100mL
ENTEROC_NMP_100mL
OD_PORC
OD_PORC_SUP
OD_PORC_MED
OD_PORC_FON
TOX_D_48_UT
TOX_V_15_UT
TOX_D_48_SUP_UT
TOX_FIS_SUP_15_UT
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

1.6 Análisis para encontrar si existe una relación entre la calidad del agua y su ubicación geográfica a través de K-means

1.6.1 Método de Curva de Codo ('Elbow Curve Method')



Como podemos observar en la gráfic, empleando el método del codo el punto de inflexión se genera aproximadamente cuando k=3

1.6.2 Agrupamiento de latitudes y longitudes con K means

[102]:
```
array([2, 1, 1, …, 2, 2, 2])
```

[103]:

	LONGITUD	LATITUD	SEMAFORO	Cluster
0	-102.33911	22.24730	Rojo	2
1	-109.84290	22.90473	Verde	1
2	-109.86442	22.89880	Verde	1
3	-109.88604	22.89609	Verde	1
4	-109.89657	22.87694	Verde	1

1.6.3 Resultados de agrupamiento de latitudes y longitudes

```
[[ -93.47131781   17.89790126]
 [-110.64834892   27.69513598]
 [-101.05200614   20.84569401]]
```

1.7 Resultados de agrupamiento de latitudes y longitudes con K-means en el mapa de México.

1.7.1 Funciones para la visualización de los Resultados

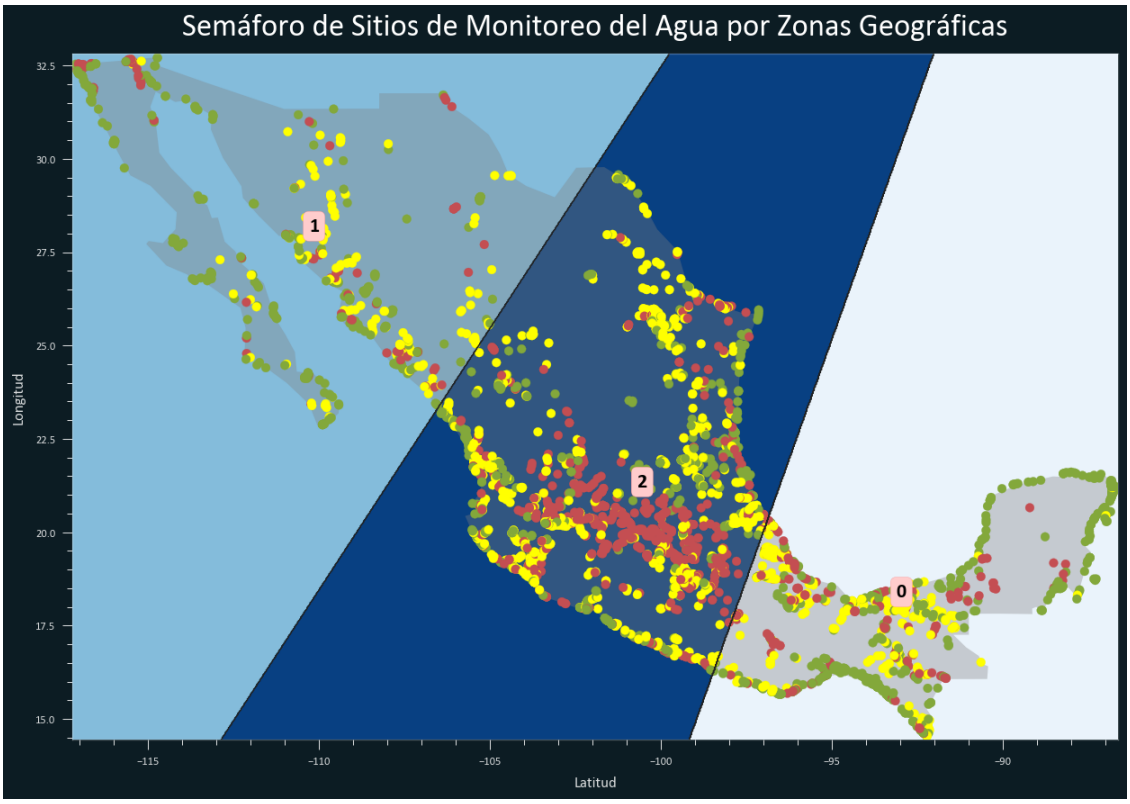
1.7.2 Visualización de Mapa de México

1.7.3 Visualización de Agrupamientos por K-means

Agrupamiento de latitudes y longitudes con K means en el mapa de México

Empleando el valor que nos dio el método del codo de k=3 podemos ver que la división quedaría aproximadamente por regiones geográficas Norte, Centro, Sur

Relación entre la calidad del agua y su ubicación geográfica a través de K- means



Podemos observar que el tener estas agrupaciones (clústers) nos ayuda a tener una mejor comprensión visual de cómo se distribuyen los semáforos por zona geográfica y cuáles podrían predominar en cada zona.

Contabilización de los semáforos por región (clúster: 0 = sur, 1 = norte, 2 = centro)

[115]:
Cluster	Amarillo	Rojo	Verde
0	300	202	484
1	146	93	288
2	689	796	495

Vamos a procedec a hacer una segunda iteración para ver como cambia la visualización de los semáforos al tener más clústers

1.8 Segunda Iteración aumentando el valor de K

Agrupamiento de latitudes y longitudes con K means

[116]:
```
array([ 1,  9,  9, …, 10, 10, 10])
```

[117]:
	LONGITUD	LATITUD	SEMAFORO	Cluster
0	-102.33911	22.24730	Rojo	1
1	-109.84290	22.90473	Verde	9
2	-109.86442	22.89880	Verde	9

```
3 -109.88604   22.89609    Verde        9
4 -109.89657   22.87694    Verde        9
```

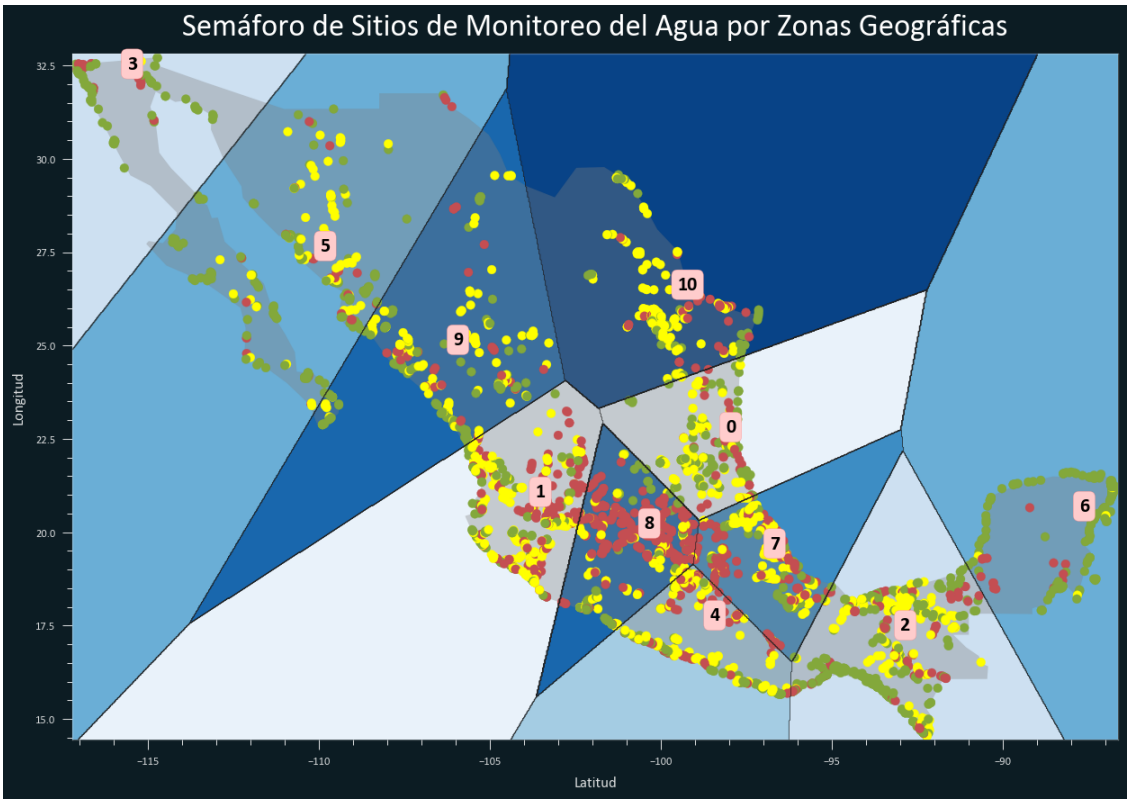
Resultados de agrupamiento de latitudes y longitudes

```
[[ -98.46585058   22.31270277]
 [-104.03198299   20.5815282 ]
 [ -93.3567396    17.00776996]
 [-115.9685       32.04154587]
 [ -98.9325036    17.27167597]
 [-110.32301463   27.16573165]
 [ -88.12757743   20.19476814]
 [ -97.18249064   19.19588754]
 [-100.84510617   19.73664197]
 [-106.441198     24.652672  ]
 [ -99.72717873   26.12937338]]
```

Visualización de Agrupamientos por K-means

Visualización del agrupamiento de latitudes y longitudes con K means en el mapa de México

Relación entre la calidad del agua y su ubicación geográfica a través de K- means



Contabilización de los semáforos por región

[121]:

Cluster	SEMAFORO	Amarillo	Rojo	Verde
0		84	46	112
1		210	158	132
2		170	90	263
3		4	41	76
4		133	96	96
5		68	31	143
6		4	15	148
7		142	193	56
8		89	344	79
9		110	37	119
10		121	40	43

A comparación de la primera iteración podemos ver de forma más fácil de como están distribuidos los semáforos por regiones más específicas, no se ve una relación directa pero comparando por cantidades pero si se vuelve más sencilla la visualización para su enfoque en una determinada zona geográfica.