Practica-5-Outliers

October 18, 2024

1 Detección de Atípicos

Cosijoeza Melchor Nolasco

[ ]:
```
import pandas as pd
import numpy as np
```

[ ]:
```
data = pd.read_csv('train.csv')
```

[ ]:
```
data.columns
```

[ ]:
```
Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath',
       'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual',
       'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType',
       'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual',
       'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC',
       'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType',
       'SaleCondition', 'SalePrice'],
      dtype='object')
```

[ ]:
```
data.head()
```

[ ]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	\
0	1	60	RL	65.0	8450	Pave	NaN	Reg	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	

```
  LandContour Utilities  … PoolArea PoolQC Fence MiscFeature MiscVal MoSold  \
0         Lvl    AllPub  …        0    NaN   NaN         NaN       0      2
1         Lvl    AllPub  …        0    NaN   NaN         NaN       0      5
2         Lvl    AllPub  …        0    NaN   NaN         NaN       0      9
3         Lvl    AllPub  …        0    NaN   NaN         NaN       0      2
4         Lvl    AllPub  …        0    NaN   NaN         NaN       0     12

  YrSold  SaleType  SaleCondition  SalePrice
0   2008        WD         Normal     208500
1   2007        WD         Normal     181500
2   2008        WD         Normal     223500
3   2006        WD        Abnorml     140000
4   2008        WD         Normal     250000

[5 rows x 81 columns]
```

[ ]:
```
out = []
def Zscore_outlier(df,umbral):
  mean = np.mean(df)
  standarDesviation = np.std(df)
  for i in df:
      z = (i - mean) / standarDesviation
      if np.abs(z) > umbral:
        out.append(i)
  print("Outliers: ",out)
Zscore_outlier(data['LotArea'],umbral=3)
```

```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200]
```

[ ]:
```
Zscore_outlier(data['SalePrice'])
```

```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 438780, 501837, 475000, 437154, 426000, 555000,
440000, 430000, 446261, 451950, 423000, 755000, 538000, 485000, 582933, 611657,
556581, 424870, 625000, 745000, 465000, 466500]
```

[ ]:
```
data.dtypes
```

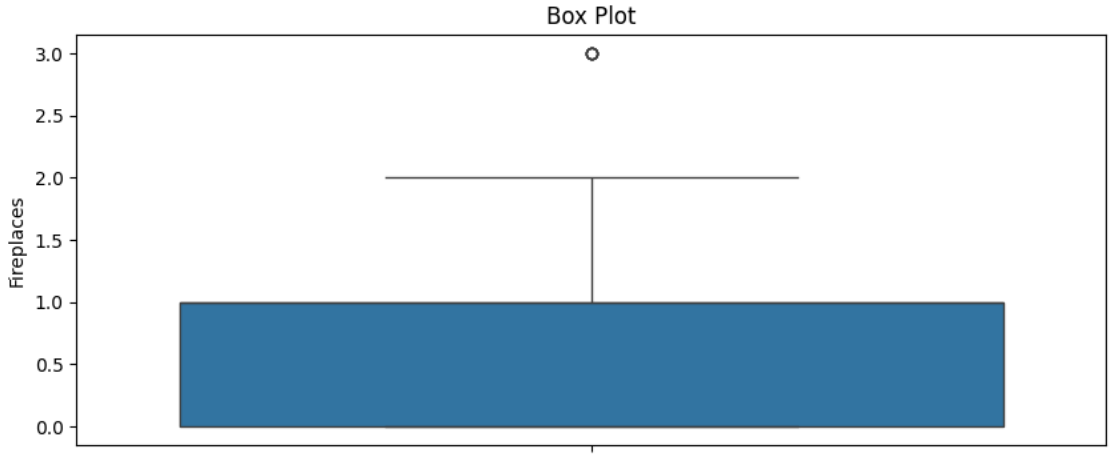
[ ]:
```
Id               int64
MSSubClass       int64
MSZoning        object
LotFrontage    float64
LotArea          int64
                 …
MoSold           int64
```

```
YrSold             int64
SaleType          object
SaleCondition     object
SalePrice          int64
Length: 81, dtype: object
```

[ ]:
```
from matplotlib import pyplot as plt
import seaborn as sns
def box_plots(df):
  plt.figure(figsize=(10,4))
  plt.title("Box Plot")
  sns.boxplot(df)
  plt.show()
```

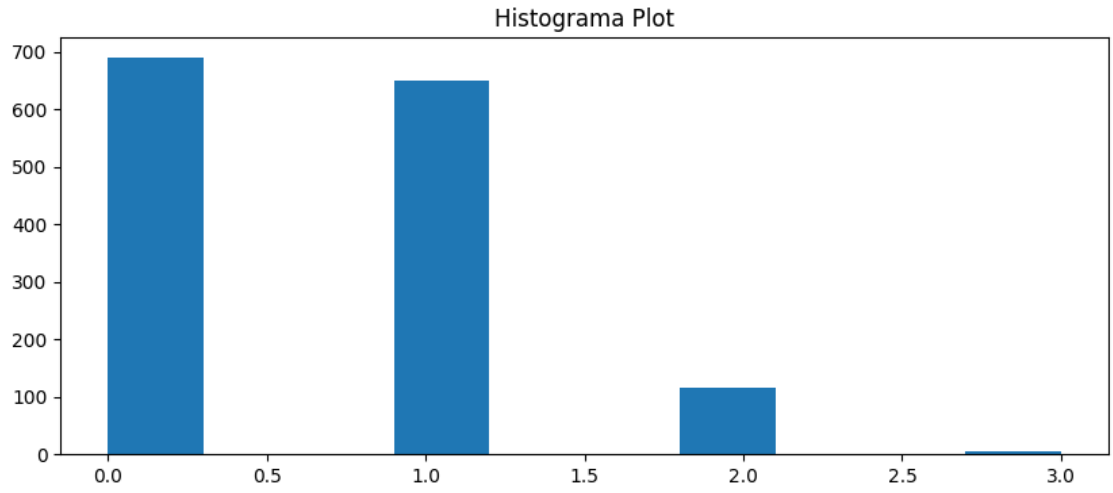
1.1 Ejemplo - Fireplaces

[ ]:
```
box_plots(data['Fireplaces'])
```



[ ]:
```
def hists_plot(df):
  plt.figure(figsize=(10,4))
  plt.hist(df)
  plt.title("Histograma Plot")
  plt.show()
```

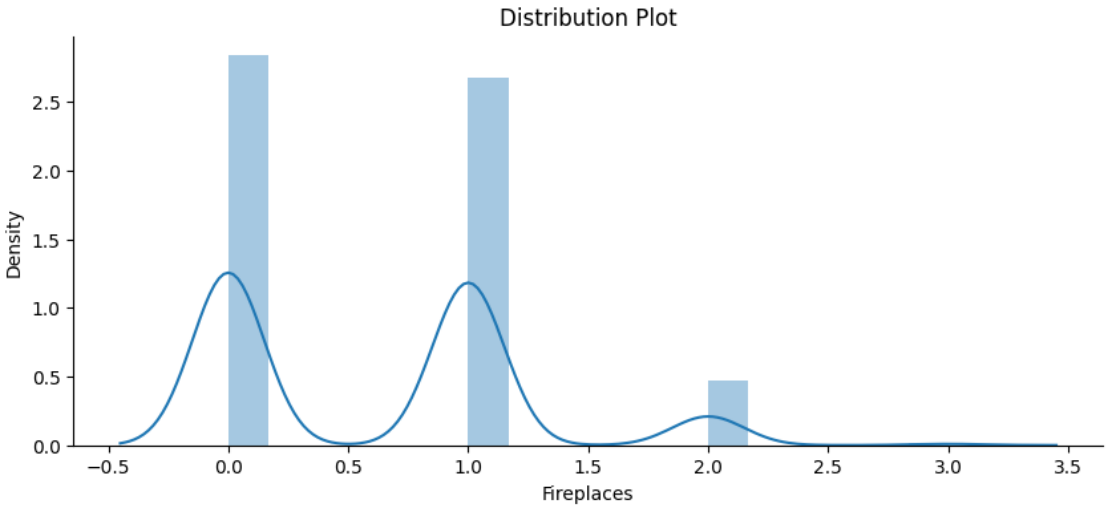
[ ]:
```
hists_plot(data['Fireplaces'])
```



[ ]:
```
def dist_plot(df):
  plt.figure(figsize=(10,4))
  sns.distplot(df)
  plt.title("Distribution Plot")
  sns.despine()
  plt.show()
```

[ ]:
```
dist_plot(data['Fireplaces'])
```

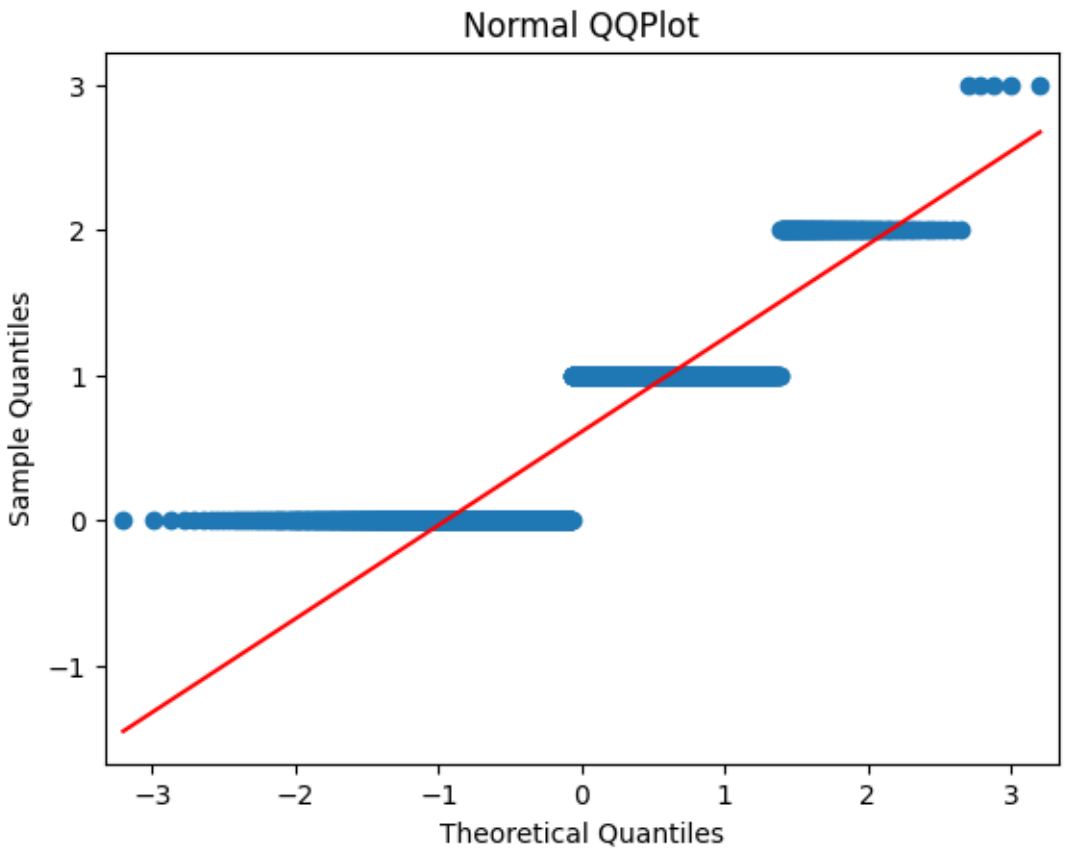
```
<ipython-input-17-70e5f0238ccb>:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df)
```



[ ]:
```
from statsmodels.graphics.gofplots import qqplot
def qq_plots(df):
  plt.figure(figsize=(10,4))
  qqplot(df,line='s')
  plt.title("Normal QQPlot")
  plt.show()
```

[ ]:
```
qq_plots(data['Fireplaces'])
```

```
<Figure size 1000x400 with 0 Axes>
```



1.2 1.- LotArea

Sus datos atípicos:

[ ]:
```
Zscore_outlier(data['LotArea'],umbral=3)
```

```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 1880, 1880,
1880, 1880, 1875, 1872, 2003, 1915, 2004, 1931, 1939, 2005, 2006, 1929, 2004,
2005, 1930, 2002, 2007, 2007, 1927, 1920, 2007, 2005, 2004, 2005, 2003, 2006,
1920, 1934, 2004, 2006, 2004, 1920, 2006, 1921, 2004, 2003, 1915, 1910, 2007,
2005, 2009, 1915, 1921, 1910, 1920, 2009, 1931, 2003, 1885, 1919, 2007, 2006,
2005, 1939, 1935, 1930, 2005, 2004, 1931, 2004, 1936, 2007, 1923, 1924, 2009,
2004, 2005, 2003, 2005, 1926, 1940, 2007, 2004, 1941, 2008, 1923, 1920, 2003,
1908, 1892, 1916, 1932, 2004, 2007, 1918, 1912, 2004, 2003, 1924, 2004, 1925,
2009, 2009, 2004, 1925, 1939, 2005, 2006, 2002, 2003, 2005, 2002, 2004, 2007,
2008, 1900, 1910, 2003, 1940, 2006, 2004, 2003, 2006, 2007, 1929, 1925, 1939,
1940, 2006, 1925, 2003, 2006, 2006, 2007, 2008, 2006, 1915, 2006, 1912, 1880,
2004, 1920, 1940, 2003, 1939, 1925, 2004, 2006, 2006, 2004, 1941, 1916, 1920,
2003, 2004, 2005, 2002, 2002, 2005, 1939, 2003, 2005, 2007, 1941, 1928, 1940,
```

1940, 2003, 1920, 2003, 1922, 2004, 2010, 1924, 2006, 2006, 1928, 2004, 1910,
2007, 1900, 1941, 1940, 1920, 2006, 2005, 1940, 1936, 1915, 2006, 2007, 2009,
1927, 2007, 1918, 1940, 2007, 1920, 1920, 1926, 1913, 1920, 2008, 1930, 2006,
1937, 1930, 2008, 1916, 1925, 2009, 1936, 1934, 2004, 2006, 2005, 2006, 2006,
2007, 1937, 2004, 2003, 1915, 1900, 1941, 2006, 1938, 1920, 1925, 2005, 2009,
1928, 2005, 1926, 2009, 1926, 1900, 2007, 2005, 2008, 1920, 1920, 2004, 1910,
2006, 2006, 1923, 2003, 2006, 2003, 1922, 1920, 2003, 1940, 1918, 1915, 2005,
2004, 2009, 1928, 2007, 2008, 1893, 1935, 2005, 1918, 1930, 2004, 2008, 2003,
2005, 1910, 2006, 2005, 1937, 2004, 2002, 1934, 2007, 2002, 2007, 2003, 1914,
1880, 2006, 1914, 1936, 1910, 2006, 2003, 2009, 2007, 1940, 1906, 1931, 2005,
1922, 2005, 1925, 1900, 1924, 2008, 1932, 2002, 2007, 2004, 2007, 2005, 2004,
1921, 1936, 1921, 2004, 2002, 2006, 1900, 2004, 1930, 2006, 2007, 1935, 1900,
1890, 2004, 2007, 2007, 1925, 2003, 1914, 2005, 2004, 1910, 1880, 1910, 2003,
2005, 2007, 1924, 2009, 2008, 1940, 2004, 2003, 2006, 2005, 1914, 1915, 2004,
2005, 2007, 2008, 1937, 1916, 2005, 2008, 2008, 1923, 1898, 2004, 1918, 2002,
2009, 2003, 2003, 1940, 2006, 2007, 1924, 2005, 2005, 2003, 1925, 1904, 1915,
1908, 2003, 2003, 1941, 1918, 2007, 2006, 1941, 2007, 2004, 2005, 1912, 1936,
2009, 2003, 2006, 2006, 1936, 2005, 1925, 2009, 1900, 2005, 2003, 2007, 2006,
2004, 1926, 2003, 2005, 2006, 1940, 1890, 2002, 2002, 2003, 2007, 2002, 2005,
1940, 1910, 2003, 2007, 1910, 1923, 2006, 2007, 2002, 1910, 2009, 2006, 1882,
2005, 2006, 1922, 2006, 1920, 2006, 2005, 2004, 1926, 1923, 1910, 2005, 2005,
2006, 1930, 2005, 2007, 1941, 1916, 1920, 2002, 1938, 2007, 2005, 2005, 2007,
2007, 2002, 2005, 2008, 1932, 1935, 1900, 1925, 2006, 1940, 1936, 2004, 2002,
2005, 2004, 2005, 1925, 2006, 1914, 1936, 1920, 2006, 2004, 2007, 2002, 1920,
2005, 2007, 2004, 2004, 1928, 1880, 1926, 1875, 1920, 2006, 1941, 1928, 1941,
1926, 1920, 1930, 2007, 2007, 2009, 2008, 1935, 2006, 1932, 1926, 1921, 2008,
1920, 1924, 1900, 2002, 1925, 2005, 2006, 1916, 1925, 2003, 2006, 1941, 2009,
2004, 2007, 2008, 1911, 1914, 2003, 2004, 2005, 2006, 2003, 2007, 2006, 1929,
2005, 1917, 2003, 2003, 1931, 1922, 2005, 1919, 1910, 2008, 1935, 1910, 2002,
1920, 2002, 1919, 1939, 2006, 1892, 2005, 2008, 2005, 2004, 2006, 2005, 2005,
2008, 2006, 1940, 2006, 1922, 1931, 1920, 2006, 1938, 1938, 1941, 2002, 2003,
2002, 1928, 2006, 1920, 2006, 1872, 1937, 2004, 1921, 2005, 1920, 2006, 2005,
2003, 2003, 1920, 2005, 2007, 1930, 1941, 2006, 1914, 1920, 1918, 1939, 1922,
1916, 2006, 1941, 1905, 2006, 2005, 1920, 1925, 1929, 2004, 2006, 2007, 1915,
2004, 1910, 2005, 1923, 2007, 1885, 2003, 1940, 2005, 1927, 2008, 1922, 2004,
2008, 1916, 2004, 1910, 2008, 2005, 2006, 2004, 1941, 1880, 1880, 1880, 1880,
1875, 1872, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 21000, 21453, 31770, 22950, 25419, 159000, 39104, 215245,
164660, 53107, 34650, 22420, 21750, 70761, 53227, 40094, 32668, 21872, 21780,
25095, 46589, 20896, 21535, 26178, 115149, 21695, 53504, 21384, 28698, 45600,
25286, 27650, 24090, 25000, 21286, 21750, 29959, 23257, 35760, 35133, 32463,
24682, 23595, 36500, 63887, 20781, 25339, 57200, 20544, 21930, 26142, 50271,
31770, 159000, 39104, 215245, 164660, 53107, 34650, 70761, 53227, 40094, 32668,
46589, 115149, 53504, 45600, 35760, 35133, 32463, 36500, 63887, 57200, 159000,
215245, 164660, 53107, 70761, 53227, 115149, 53504, 63887, 57200, 50271, 159000,

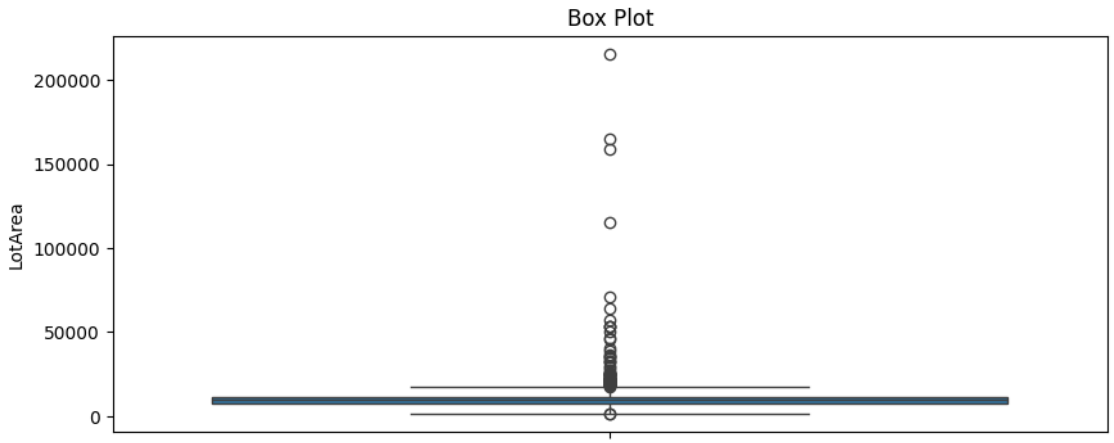
```
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200]
```

[ ]:
```
Zscore_outlier(data['LotArea'],umbral=400)
```

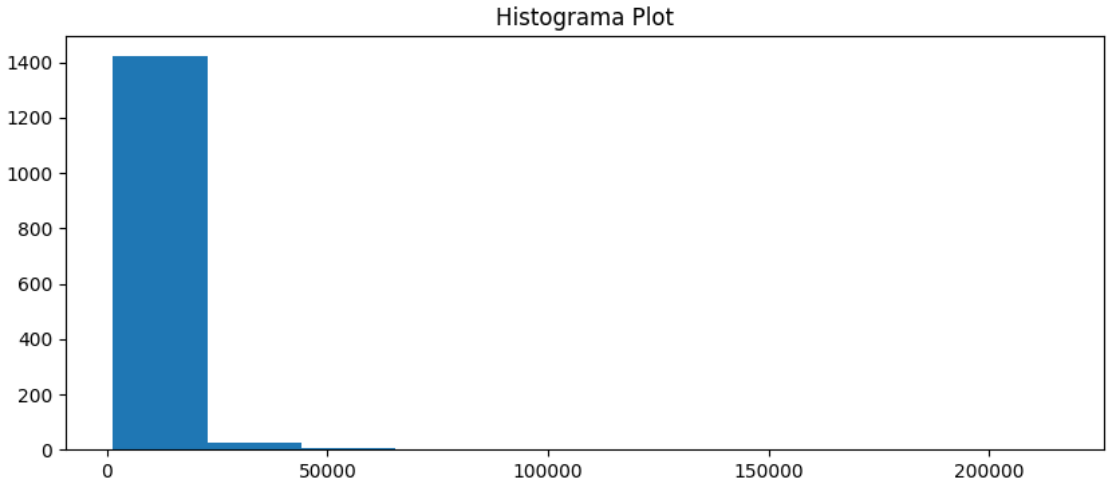
```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 1880, 1880,
1880, 1880, 1875, 1872, 2003, 1915, 2004, 1931, 1939, 2005, 2006, 1929, 2004,
2005, 1930, 2002, 2007, 2007, 1927, 1920, 2007, 2005, 2004, 2005, 2003, 2006,
1920, 1934, 2004, 2006, 2004, 1920, 2006, 1921, 2004, 2003, 1915, 1910, 2007,
2005, 2009, 1915, 1921, 1910, 1920, 2009, 1931, 2003, 1885, 1919, 2007, 2006,
2005, 1939, 1935, 1930, 2005, 2004, 1931, 2004, 1936, 2007, 1923, 1924, 2009,
2004, 2005, 2003, 2005, 1926, 1940, 2007, 2004, 1941, 2008, 1923, 1920, 2003,
1908, 1892, 1916, 1932, 2004, 2007, 1918, 1912, 2004, 2003, 1924, 2004, 1925,
2009, 2009, 2004, 1925, 1939, 2005, 2006, 2002, 2003, 2005, 2002, 2004, 2007,
2008, 1900, 1910, 2003, 1940, 2006, 2004, 2003, 2006, 2007, 1929, 1925, 1939,
1940, 2006, 1925, 2003, 2006, 2006, 2007, 2008, 2006, 1915, 2006, 1912, 1880,
2004, 1920, 1940, 2003, 1939, 1925, 2004, 2006, 2006, 2004, 1941, 1916, 1920,
2003, 2004, 2005, 2002, 2002, 2005, 1939, 2003, 2005, 2007, 1941, 1928, 1940,
1940, 2003, 1920, 2003, 1922, 2004, 2010, 1924, 2006, 2006, 1928, 2004, 1910,
2007, 1900, 1941, 1940, 1920, 2006, 2005, 1940, 1936, 1915, 2006, 2007, 2009,
1927, 2007, 1918, 1940, 2007, 1920, 1920, 1926, 1913, 1920, 2008, 1930, 2006,
1937, 1930, 2008, 1916, 1925, 2009, 1936, 1934, 2004, 2006, 2005, 2006, 2006,
2007, 1937, 2004, 2003, 1915, 1900, 1941, 2006, 1938, 1920, 1925, 2005, 2009,
1928, 2005, 1926, 2009, 1926, 1900, 2007, 2005, 2008, 1920, 1920, 2004, 1910,
2006, 2006, 1923, 2003, 2006, 2003, 1922, 1920, 2003, 1940, 1918, 1915, 2005,
2004, 2009, 1928, 2007, 2008, 1893, 1935, 2005, 1918, 1930, 2004, 2008, 2003,
2005, 1910, 2006, 2005, 1937, 2004, 2002, 1934, 2007, 2002, 2007, 2003, 1914,
1880, 2006, 1914, 1936, 1910, 2006, 2003, 2009, 2007, 1940, 1906, 1931, 2005,
1922, 2005, 1925, 1900, 1924, 2008, 1932, 2002, 2007, 2004, 2007, 2005, 2004,
1921, 1936, 1921, 2004, 2002, 2006, 1900, 2004, 1930, 2006, 2007, 1935, 1900,
1890, 2004, 2007, 2007, 1925, 2003, 1914, 2005, 2004, 1910, 1880, 1910, 2003,
2005, 2007, 1924, 2009, 2008, 1940, 2004, 2003, 2006, 2005, 1914, 1915, 2004,
2005, 2007, 2008, 1937, 1916, 2005, 2008, 2008, 1923, 1898, 2004, 1918, 2002,
2009, 2003, 2003, 1940, 2006, 2007, 1924, 2005, 2005, 2003, 1925, 1904, 1915,
1908, 2003, 2003, 1941, 1918, 2007, 2006, 1941, 2007, 2004, 2005, 1912, 1936,
2009, 2003, 2006, 2006, 1936, 2005, 1925, 2009, 1900, 2005, 2003, 2007, 2006,
2004, 1926, 2003, 2005, 2006, 1940, 1890, 2002, 2002, 2003, 2007, 2002, 2005,
1940, 1910, 2003, 2007, 1910, 1923, 2006, 2007, 2002, 1910, 2009, 2006, 1882,
2005, 2006, 1922, 2006, 1920, 2006, 2005, 2004, 1926, 1923, 1910, 2005, 2005,
2006, 1930, 2005, 2007, 1941, 1916, 1920, 2002, 1938, 2007, 2005, 2005, 2007,
2007, 2002, 2005, 2008, 1932, 1935, 1900, 1925, 2006, 1940, 1936, 2004, 2002,
2005, 2004, 2005, 1925, 2006, 1914, 1936, 1920, 2006, 2004, 2007, 2002, 1920,
2005, 2007, 2004, 2004, 1928, 1880, 1926, 1875, 1920, 2006, 1941, 1928, 1941,
1926, 1920, 1930, 2007, 2007, 2009, 2008, 1935, 2006, 1932, 1926, 1921, 2008,
1920, 1924, 1900, 2002, 1925, 2005, 2006, 1916, 1925, 2003, 2006, 1941, 2009,
```

```
2004, 2007, 2008, 1911, 1914, 2003, 2004, 2005, 2006, 2003, 2007, 2006, 1929,
2005, 1917, 2003, 2003, 1931, 1922, 2005, 1919, 1910, 2008, 1935, 1910, 2002,
1920, 2002, 1919, 1939, 2006, 1892, 2005, 2008, 2005, 2004, 2006, 2005, 2005,
2008, 2006, 1940, 2006, 1922, 1931, 1920, 2006, 1938, 1938, 1941, 2002, 2003,
2002, 1928, 2006, 1920, 2006, 1872, 1937, 2004, 1921, 2005, 1920, 2006, 2005,
2003, 2003, 1920, 2005, 2007, 1930, 1941, 2006, 1914, 1920, 1918, 1939, 1922,
1916, 2006, 1941, 1905, 2006, 2005, 1920, 1925, 1929, 2004, 2006, 2007, 1915,
2004, 1910, 2005, 1923, 2007, 1885, 2003, 1940, 2005, 1927, 2008, 1922, 2004,
2008, 1916, 2004, 1910, 2008, 2005, 2006, 2004, 1941, 1880, 1880, 1880, 1880,
1875, 1872, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 21000, 21453, 31770, 22950, 25419, 159000, 39104, 215245,
164660, 53107, 34650, 22420, 21750, 70761, 53227, 40094, 32668, 21872, 21780,
25095, 46589, 20896, 21535, 26178, 115149, 21695, 53504, 21384, 28698, 45600,
25286, 27650, 24090, 25000, 21286, 21750, 29959, 23257, 35760, 35133, 32463,
24682, 23595, 36500, 63887, 20781, 25339, 57200, 20544, 21930, 26142, 50271,
31770, 159000, 39104, 215245, 164660, 53107, 34650, 70761, 53227, 40094, 32668,
46589, 115149, 53504, 45600, 35760, 35133, 32463, 36500, 63887, 57200, 159000,
215245, 164660, 53107, 70761, 53227, 115149, 53504, 63887, 57200]
```

[ ]:
```
box_plots(data['LotArea'])
```

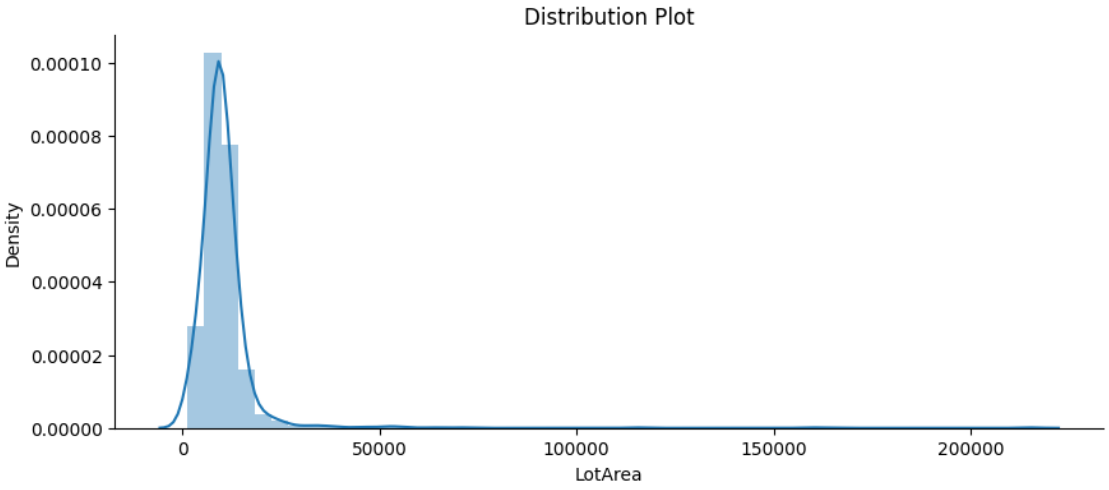


[ ]:
```
hists_plot(data['LotArea'])
```



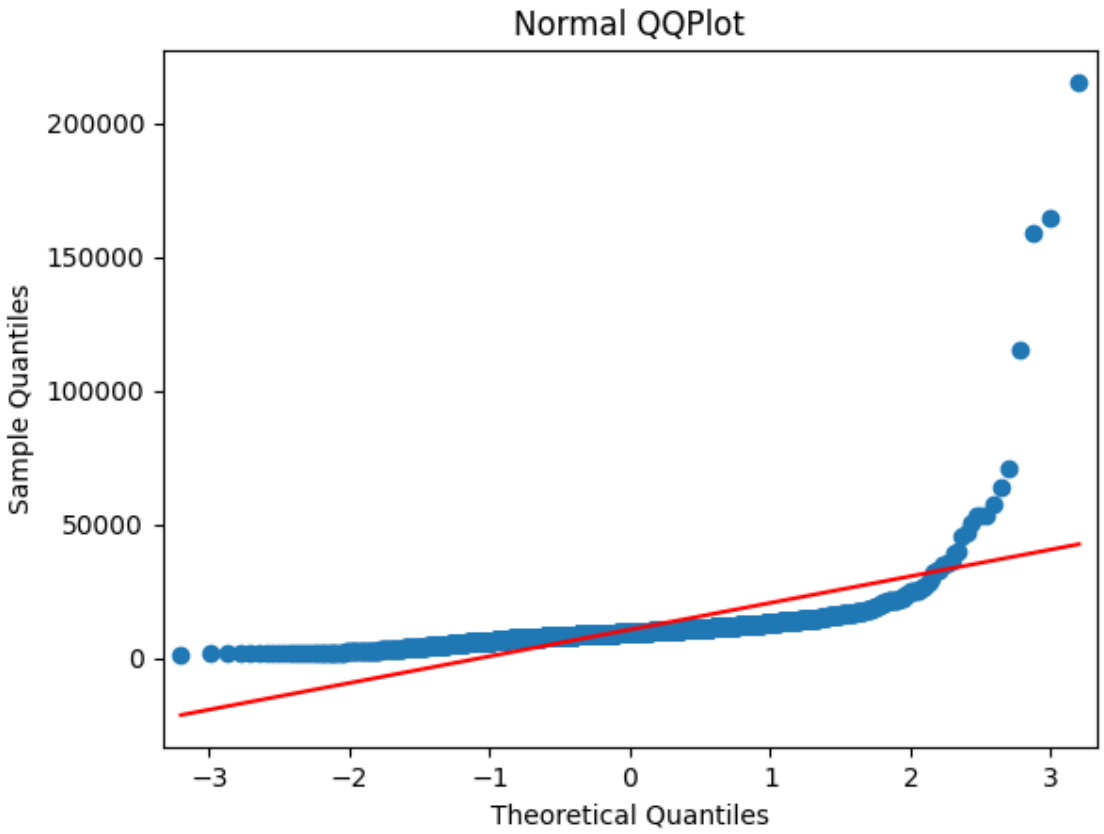
[ ]:
```
dist_plot(data['LotArea'])
```

```
<ipython-input-17-70e5f0238ccb>:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df)
```



[ ]:
```
qq_plots(data['LotArea'])
```

```
<Figure size 1000x400 with 0 Axes>
```



Los datos atípicos están de 50000 en adelante

1.3 2.-YearBuilt

Sus datos atípicos:

[ ]:
```
Zscore_outlier(data['YearBuilt'],umbral=3)
```

```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 1880, 1880,
1880, 1880, 1875, 1872, 2003, 1915, 2004, 1931, 1939, 2005, 2006, 1929, 2004,
2005, 1930, 2002, 2007, 2007, 1927, 1920, 2007, 2005, 2004, 2005, 2003, 2006,
1920, 1934, 2004, 2006, 2004, 1920, 2006, 1921, 2004, 2003, 1915, 1910, 2007,
2005, 2009, 1915, 1921, 1910, 1920, 2009, 1931, 2003, 1885, 1919, 2007, 2006,
2005, 1939, 1935, 1930, 2005, 2004, 1931, 2004, 1936, 2007, 1923, 1924, 2009,
2004, 2005, 2003, 2005, 1926, 1940, 2007, 2004, 1941, 2008, 1923, 1920, 2003,
1908, 1892, 1916, 1932, 2004, 2007, 1918, 1912, 2004, 2003, 1924, 2004, 1925,
2009, 2009, 2004, 1925, 1939, 2005, 2006, 2002, 2003, 2005, 2002, 2004, 2007,
```

2008, 1900, 1910, 2003, 1940, 2006, 2004, 2003, 2006, 2007, 1929, 1925, 1939,
1940, 2006, 1925, 2003, 2006, 2006, 2007, 2008, 2006, 1915, 2006, 1912, 1880,
2004, 1920, 1940, 2003, 1939, 1925, 2004, 2006, 2006, 2004, 1941, 1916, 1920,
2003, 2004, 2005, 2002, 2002, 2005, 1939, 2003, 2005, 2007, 1941, 1928, 1940,
1940, 2003, 1920, 2003, 1922, 2004, 2010, 1924, 2006, 2006, 1928, 2004, 1910,
2007, 1900, 1941, 1940, 1920, 2006, 2005, 1940, 1936, 1915, 2006, 2007, 2009,
1927, 2007, 1918, 1940, 2007, 1920, 1920, 1926, 1913, 1920, 2008, 1930, 2006,
1937, 1930, 2008, 1916, 1925, 2009, 1936, 1934, 2004, 2006, 2005, 2006, 2006,
2007, 1937, 2004, 2003, 1915, 1900, 1941, 2006, 1938, 1920, 1925, 2005, 2009,
1928, 2005, 1926, 2009, 1926, 1900, 2007, 2005, 2008, 1920, 1920, 2004, 1910,
2006, 2006, 1923, 2003, 2006, 2003, 1922, 1920, 2003, 1940, 1918, 1915, 2005,
2004, 2009, 1928, 2007, 2008, 1893, 1935, 2005, 1918, 1930, 2004, 2008, 2003,
2005, 1910, 2006, 2005, 1937, 2004, 2002, 1934, 2007, 2002, 2007, 2003, 1914,
1880, 2006, 1914, 1936, 1910, 2006, 2003, 2009, 2007, 1940, 1906, 1931, 2005,
1922, 2005, 1925, 1900, 1924, 2008, 1932, 2002, 2007, 2004, 2007, 2005, 2004,
1921, 1936, 1921, 2004, 2002, 2006, 1900, 2004, 1930, 2006, 2007, 1935, 1900,
1890, 2004, 2007, 2007, 1925, 2003, 1914, 2005, 2004, 1910, 1880, 1910, 2003,
2005, 2007, 1924, 2009, 2008, 1940, 2004, 2003, 2006, 2005, 1914, 1915, 2004,
2005, 2007, 2008, 1937, 1916, 2005, 2008, 2008, 1923, 1898, 2004, 1918, 2002,
2009, 2003, 2003, 1940, 2006, 2007, 1924, 2005, 2005, 2003, 1925, 1904, 1915,
1908, 2003, 2003, 1941, 1918, 2007, 2006, 1941, 2007, 2004, 2005, 1912, 1936,
2009, 2003, 2006, 2006, 1936, 2005, 1925, 2009, 1900, 2005, 2003, 2007, 2006,
2004, 1926, 2003, 2005, 2006, 1940, 1890, 2002, 2002, 2003, 2007, 2002, 2005,
1940, 1910, 2003, 2007, 1910, 1923, 2006, 2007, 2002, 1910, 2009, 2006, 1882,
2005, 2006, 1922, 2006, 1920, 2006, 2005, 2004, 1926, 1923, 1910, 2005, 2005,
2006, 1930, 2005, 2007, 1941, 1916, 1920, 2002, 1938, 2007, 2005, 2005, 2007,
2007, 2002, 2005, 2008, 1932, 1935, 1900, 1925, 2006, 1940, 1936, 2004, 2002,
2005, 2004, 2005, 1925, 2006, 1914, 1936, 1920, 2006, 2004, 2007, 2002, 1920,
2005, 2007, 2004, 2004, 1928, 1880, 1926, 1875, 1920, 2006, 1941, 1928, 1941,
1926, 1920, 1930, 2007, 2007, 2009, 2008, 1935, 2006, 1932, 1926, 1921, 2008,
1920, 1924, 1900, 2002, 1925, 2005, 2006, 1916, 1925, 2003, 2006, 1941, 2009,
2004, 2007, 2008, 1911, 1914, 2003, 2004, 2005, 2006, 2003, 2007, 2006, 1929,
2005, 1917, 2003, 2003, 1931, 1922, 2005, 1919, 1910, 2008, 1935, 1910, 2002,
1920, 2002, 1919, 1939, 2006, 1892, 2005, 2008, 2005, 2004, 2006, 2005, 2005,
2008, 2006, 1940, 2006, 1922, 1931, 1920, 2006, 1938, 1938, 1941, 2002, 2003,
2002, 1928, 2006, 1920, 2006, 1872, 1937, 2004, 1921, 2005, 1920, 2006, 2005,
2003, 2003, 1920, 2005, 2007, 1930, 1941, 2006, 1914, 1920, 1918, 1939, 1922,
1916, 2006, 1941, 1905, 2006, 2005, 1920, 1925, 1929, 2004, 2006, 2007, 1915,
2004, 1910, 2005, 1923, 2007, 1885, 2003, 1940, 2005, 1927, 2008, 1922, 2004,
2008, 1916, 2004, 1910, 2008, 2005, 2006, 2004, 1941, 1880, 1880, 1880, 1880,
1875, 1872, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 21000, 21453, 31770, 22950, 25419, 159000, 39104, 215245,
164660, 53107, 34650, 22420, 21750, 70761, 53227, 40094, 32668, 21872, 21780,
25095, 46589, 20896, 21535, 26178, 115149, 21695, 53504, 21384, 28698, 45600,
25286, 27650, 24090, 25000, 21286, 21750, 29959, 23257, 35760, 35133, 32463,

```
24682, 23595, 36500, 63887, 20781, 25339, 57200, 20544, 21930, 26142, 50271,
31770, 159000, 39104, 215245, 164660, 53107, 34650, 70761, 53227, 40094, 32668,
46589, 115149, 53504, 45600, 35760, 35133, 32463, 36500, 63887, 57200, 159000,
215245, 164660, 53107, 70761, 53227, 115149, 53504, 63887, 57200, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872]
```

[ ]:
```
Zscore_outlier(data['YearBuilt'],umbral=1950)
```

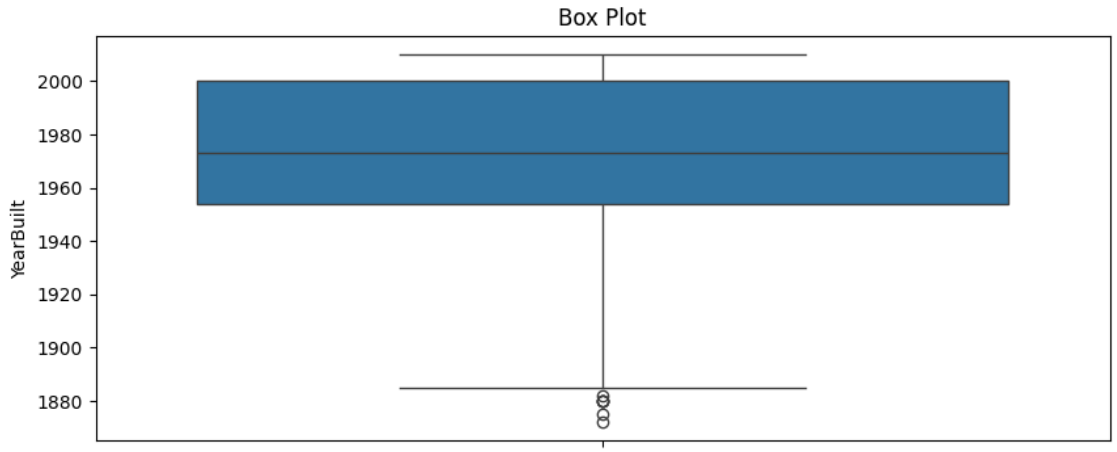
```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 1880, 1880,
1880, 1880, 1875, 1872, 2003, 1915, 2004, 1931, 1939, 2005, 2006, 1929, 2004,
2005, 1930, 2002, 2007, 2007, 1927, 1920, 2007, 2005, 2004, 2005, 2003, 2006,
1920, 1934, 2004, 2006, 2004, 1920, 2006, 1921, 2004, 2003, 1915, 1910, 2007,
2005, 2009, 1915, 1921, 1910, 1920, 2009, 1931, 2003, 1885, 1919, 2007, 2006,
2005, 1939, 1935, 1930, 2005, 2004, 1931, 2004, 1936, 2007, 1923, 1924, 2009,
2004, 2005, 2003, 2005, 1926, 1940, 2007, 2004, 1941, 2008, 1923, 1920, 2003,
1908, 1892, 1916, 1932, 2004, 2007, 1918, 1912, 2004, 2003, 1924, 2004, 1925,
2009, 2009, 2004, 1925, 1939, 2005, 2006, 2002, 2003, 2005, 2002, 2004, 2007,
2008, 1900, 1910, 2003, 1940, 2006, 2004, 2003, 2006, 2007, 1929, 1925, 1939,
1940, 2006, 1925, 2003, 2006, 2006, 2007, 2008, 2006, 1915, 2006, 1912, 1880,
2004, 1920, 1940, 2003, 1939, 1925, 2004, 2006, 2006, 2004, 1941, 1916, 1920,
2003, 2004, 2005, 2002, 2002, 2005, 1939, 2003, 2005, 2007, 1941, 1928, 1940,
1940, 2003, 1920, 2003, 1922, 2004, 2010, 1924, 2006, 2006, 1928, 2004, 1910,
2007, 1900, 1941, 1940, 1920, 2006, 2005, 1940, 1936, 1915, 2006, 2007, 2009,
1927, 2007, 1918, 1940, 2007, 1920, 1920, 1926, 1913, 1920, 2008, 1930, 2006,
1937, 1930, 2008, 1916, 1925, 2009, 1936, 1934, 2004, 2006, 2005, 2006, 2006,
2007, 1937, 2004, 2003, 1915, 1900, 1941, 2006, 1938, 1920, 1925, 2005, 2009,
1928, 2005, 1926, 2009, 1926, 1900, 2007, 2005, 2008, 1920, 1920, 2004, 1910,
2006, 2006, 1923, 2003, 2006, 2003, 1922, 1920, 2003, 1940, 1918, 1915, 2005,
2004, 2009, 1928, 2007, 2008, 1893, 1935, 2005, 1918, 1930, 2004, 2008, 2003,
2005, 1910, 2006, 2005, 1937, 2004, 2002, 1934, 2007, 2002, 2007, 2003, 1914,
1880, 2006, 1914, 1936, 1910, 2006, 2003, 2009, 2007, 1940, 1906, 1931, 2005,
1922, 2005, 1925, 1900, 1924, 2008, 1932, 2002, 2007, 2004, 2007, 2005, 2004,
1921, 1936, 1921, 2004, 2002, 2006, 1900, 2004, 1930, 2006, 2007, 1935, 1900,
1890, 2004, 2007, 2007, 1925, 2003, 1914, 2005, 2004, 1910, 1880, 1910, 2003,
2005, 2007, 1924, 2009, 2008, 1940, 2004, 2003, 2006, 2005, 1914, 1915, 2004,
2005, 2007, 2008, 1937, 1916, 2005, 2008, 2008, 1923, 1898, 2004, 1918, 2002,
2009, 2003, 2003, 1940, 2006, 2007, 1924, 2005, 2005, 2003, 1925, 1904, 1915,
1908, 2003, 2003, 1941, 1918, 2007, 2006, 1941, 2007, 2004, 2005, 1912, 1936,
2009, 2003, 2006, 2006, 1936, 2005, 1925, 2009, 1900, 2005, 2003, 2007, 2006,
2004, 1926, 2003, 2005, 2006, 1940, 1890, 2002, 2002, 2003, 2007, 2002, 2005,
1940, 1910, 2003, 2007, 1910, 1923, 2006, 2007, 2002, 1910, 2009, 2006, 1882,
2005, 2006, 1922, 2006, 1920, 2006, 2005, 2004, 1926, 1923, 1910, 2005, 2005,
2006, 1930, 2005, 2007, 1941, 1916, 1920, 2002, 1938, 2007, 2005, 2005, 2007,
2007, 2002, 2005, 2008, 1932, 1935, 1900, 1925, 2006, 1940, 1936, 2004, 2002,
```

```
2005, 2004, 2005, 1925, 2006, 1914, 1936, 1920, 2006, 2004, 2007, 2002, 1920,
2005, 2007, 2004, 2004, 1928, 1880, 1926, 1875, 1920, 2006, 1941, 1928, 1941,
1926, 1920, 1930, 2007, 2007, 2009, 2008, 1935, 2006, 1932, 1926, 1921, 2008,
1920, 1924, 1900, 2002, 1925, 2005, 2006, 1916, 1925, 2003, 2006, 1941, 2009,
2004, 2007, 2008, 1911, 1914, 2003, 2004, 2005, 2006, 2003, 2007, 2006, 1929,
2005, 1917, 2003, 2003, 1931, 1922, 2005, 1919, 1910, 2008, 1935, 1910, 2002,
1920, 2002, 1919, 1939, 2006, 1892, 2005, 2008, 2005, 2004, 2006, 2005, 2005,
2008, 2006, 1940, 2006, 1922, 1931, 1920, 2006, 1938, 1938, 1941, 2002, 2003,
2002, 1928, 2006, 1920, 2006, 1872, 1937, 2004, 1921, 2005, 1920, 2006, 2005,
2003, 2003, 1920, 2005, 2007, 1930, 1941, 2006, 1914, 1920, 1918, 1939, 1922,
1916, 2006, 1941, 1905, 2006, 2005, 1920, 1925, 1929, 2004, 2006, 2007, 1915,
2004, 1910, 2005, 1923, 2007, 1885, 2003, 1940, 2005, 1927, 2008, 1922, 2004,
2008, 1916, 2004, 1910, 2008, 2005, 2006, 2004, 1941, 1880, 1880, 1880, 1880,
1875, 1872]
```

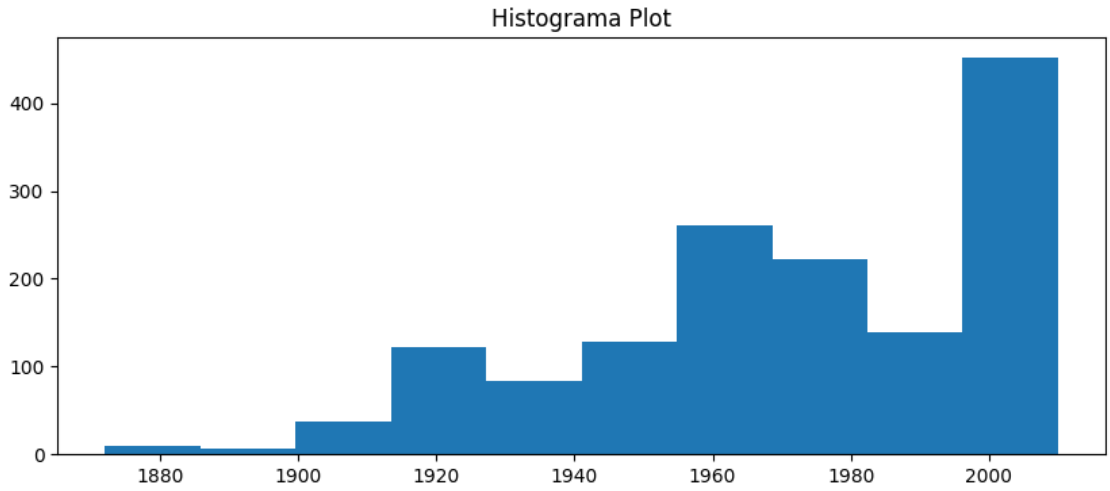
[ ]:
```
data['YearBuilt'].dtypes
```

[ ]:
```
dtype('int64')
```

[ ]:
```
box_plots(data['YearBuilt'])
```

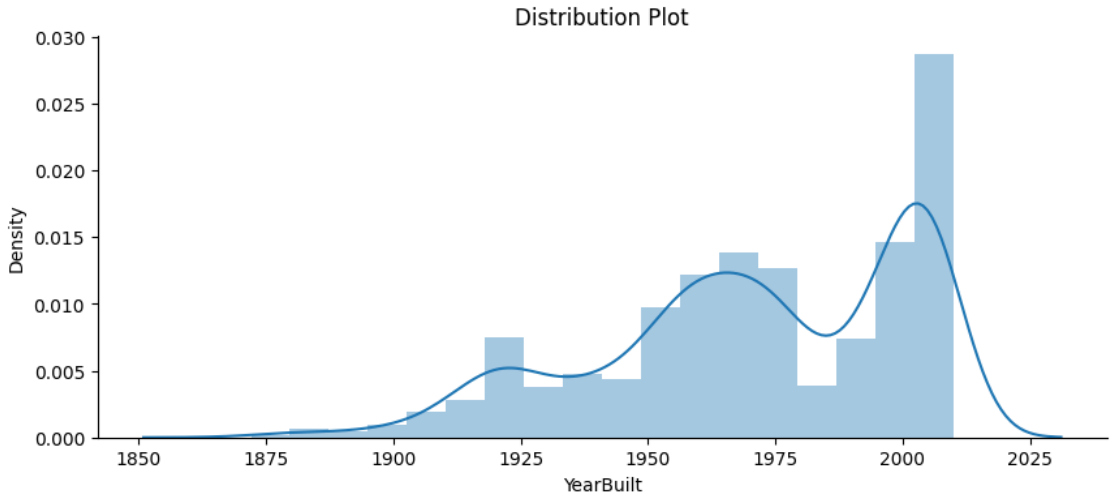


[ ]:
```
hists_plot(data['YearBuilt'])
```



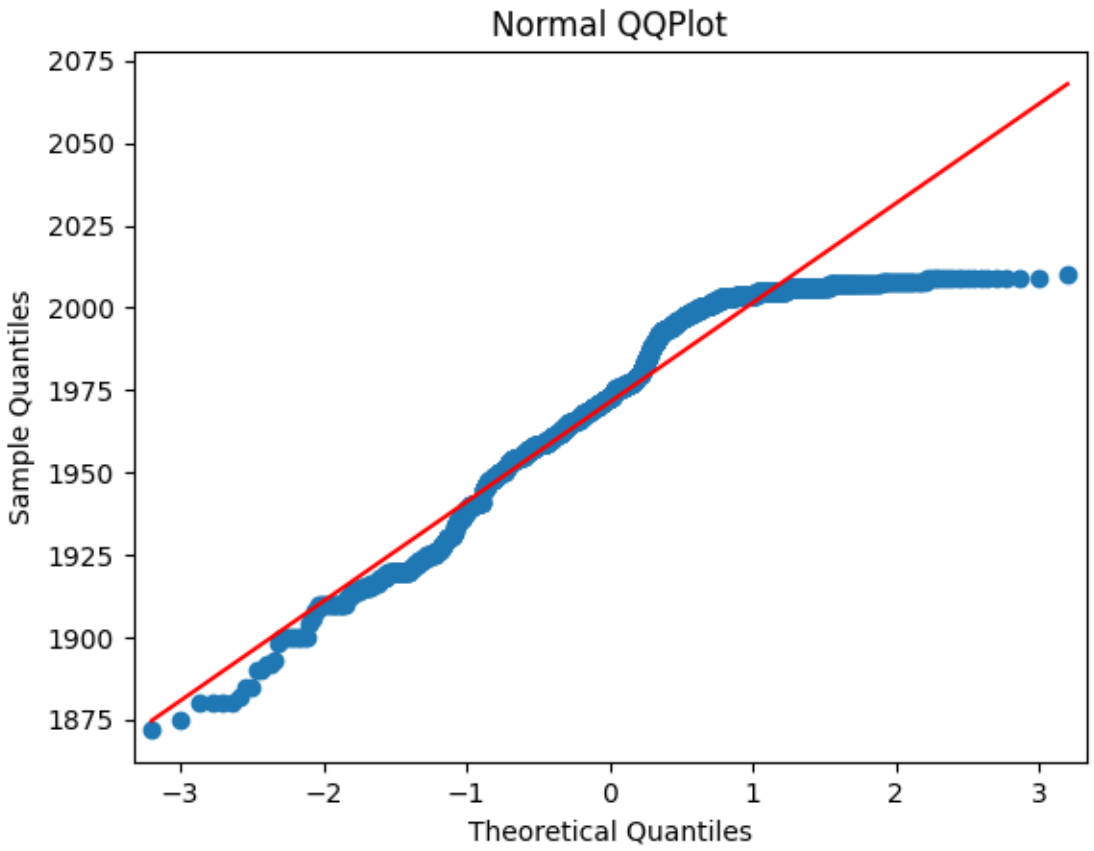
[ ]:
```
dist_plot(data['YearBuilt'])
```

```
<ipython-input-17-70e5f0238ccb>:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df)
```



[ ]:
```
qq_plots(data['YearBuilt'])
```

```
<Figure size 1000x400 with 0 Axes>
```



Los datos atípicos se encuentran por debajo del año 1900

1.4 3.-YrSold

Sus datos atípicos:

[ ]:
```
Zscore_outlier(data['YrSold'],umbral=3)
```

```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 1880, 1880,
1880, 1880, 1875, 1872, 2003, 1915, 2004, 1931, 1939, 2005, 2006, 1929, 2004,
2005, 1930, 2002, 2007, 2007, 1927, 1920, 2007, 2005, 2004, 2005, 2003, 2006,
1920, 1934, 2004, 2006, 2004, 1920, 2006, 1921, 2004, 2003, 1915, 1910, 2007,
2005, 2009, 1915, 1921, 1910, 1920, 2009, 1931, 2003, 1885, 1919, 2007, 2006,
2005, 1939, 1935, 1930, 2005, 2004, 1931, 2004, 1936, 2007, 1923, 1924, 2009,
2004, 2005, 2003, 2005, 1926, 1940, 2007, 2004, 1941, 2008, 1923, 1920, 2003,
1908, 1892, 1916, 1932, 2004, 2007, 1918, 1912, 2004, 2003, 1924, 2004, 1925,
```

2009, 2009, 2004, 1925, 1939, 2005, 2006, 2002, 2003, 2005, 2002, 2004, 2007,
2008, 1900, 1910, 2003, 1940, 2006, 2004, 2003, 2006, 2007, 1929, 1925, 1939,
1940, 2006, 1925, 2003, 2006, 2006, 2007, 2008, 2006, 1915, 2006, 1912, 1880,
2004, 1920, 1940, 2003, 1939, 1925, 2004, 2006, 2006, 2004, 1941, 1916, 1920,
2003, 2004, 2005, 2002, 2002, 2005, 1939, 2003, 2005, 2007, 1941, 1928, 1940,
1940, 2003, 1920, 2003, 1922, 2004, 2010, 1924, 2006, 2006, 1928, 2004, 1910,
2007, 1900, 1941, 1940, 1920, 2006, 2005, 1940, 1936, 1915, 2006, 2007, 2009,
1927, 2007, 1918, 1940, 2007, 1920, 1920, 1926, 1913, 1920, 2008, 1930, 2006,
1937, 1930, 2008, 1916, 1925, 2009, 1936, 1934, 2004, 2006, 2005, 2006, 2006,
2007, 1937, 2004, 2003, 1915, 1900, 1941, 2006, 1938, 1920, 1925, 2005, 2009,
1928, 2005, 1926, 2009, 1926, 1900, 2007, 2005, 2008, 1920, 1920, 2004, 1910,
2006, 2006, 1923, 2003, 2006, 2003, 1922, 1920, 2003, 1940, 1918, 1915, 2005,
2004, 2009, 1928, 2007, 2008, 1893, 1935, 2005, 1918, 1930, 2004, 2008, 2003,
2005, 1910, 2006, 2005, 1937, 2004, 2002, 1934, 2007, 2002, 2007, 2003, 1914,
1880, 2006, 1914, 1936, 1910, 2006, 2003, 2009, 2007, 1940, 1906, 1931, 2005,
1922, 2005, 1925, 1900, 1924, 2008, 1932, 2002, 2007, 2004, 2007, 2005, 2004,
1921, 1936, 1921, 2004, 2002, 2006, 1900, 2004, 1930, 2006, 2007, 1935, 1900,
1890, 2004, 2007, 2007, 1925, 2003, 1914, 2005, 2004, 1910, 1880, 1910, 2003,
2005, 2007, 1924, 2009, 2008, 1940, 2004, 2003, 2006, 2005, 1914, 1915, 2004,
2005, 2007, 2008, 1937, 1916, 2005, 2008, 2008, 1923, 1898, 2004, 1918, 2002,
2009, 2003, 2003, 1940, 2006, 2007, 1924, 2005, 2005, 2003, 1925, 1904, 1915,
1908, 2003, 2003, 1941, 1918, 2007, 2006, 1941, 2007, 2004, 2005, 1912, 1936,
2009, 2003, 2006, 2006, 1936, 2005, 1925, 2009, 1900, 2005, 2003, 2007, 2006,
2004, 1926, 2003, 2005, 2006, 1940, 1890, 2002, 2002, 2003, 2007, 2002, 2005,
1940, 1910, 2003, 2007, 1910, 1923, 2006, 2007, 2002, 1910, 2009, 2006, 1882,
2005, 2006, 1922, 2006, 1920, 2006, 2005, 2004, 1926, 1923, 1910, 2005, 2005,
2006, 1930, 2005, 2007, 1941, 1916, 1920, 2002, 1938, 2007, 2005, 2005, 2007,
2007, 2002, 2005, 2008, 1932, 1935, 1900, 1925, 2006, 1940, 1936, 2004, 2002,
2005, 2004, 2005, 1925, 2006, 1914, 1936, 1920, 2006, 2004, 2007, 2002, 1920,
2005, 2007, 2004, 2004, 1928, 1880, 1926, 1875, 1920, 2006, 1941, 1928, 1941,
1926, 1920, 1930, 2007, 2007, 2009, 2008, 1935, 2006, 1932, 1926, 1921, 2008,
1920, 1924, 1900, 2002, 1925, 2005, 2006, 1916, 1925, 2003, 2006, 1941, 2009,
2004, 2007, 2008, 1911, 1914, 2003, 2004, 2005, 2006, 2003, 2007, 2006, 1929,
2005, 1917, 2003, 2003, 1931, 1922, 2005, 1919, 1910, 2008, 1935, 1910, 2002,
1920, 2002, 1919, 1939, 2006, 1892, 2005, 2008, 2005, 2004, 2006, 2005, 2005,
2008, 2006, 1940, 2006, 1922, 1931, 1920, 2006, 1938, 1938, 1941, 2002, 2003,
2002, 1928, 2006, 1920, 2006, 1872, 1937, 2004, 1921, 2005, 1920, 2006, 2005,
2003, 2003, 1920, 2005, 2007, 1930, 1941, 2006, 1914, 1920, 1918, 1939, 1922,
1916, 2006, 1941, 1905, 2006, 2005, 1920, 1925, 1929, 2004, 2006, 2007, 1915,
2004, 1910, 2005, 1923, 2007, 1885, 2003, 1940, 2005, 1927, 2008, 1922, 2004,
2008, 1916, 2004, 1910, 2008, 2005, 2006, 2004, 1941, 1880, 1880, 1880, 1880,
1875, 1872, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 21000, 21453, 31770, 22950, 25419, 159000, 39104, 215245,
164660, 53107, 34650, 22420, 21750, 70761, 53227, 40094, 32668, 21872, 21780,
25095, 46589, 20896, 21535, 26178, 115149, 21695, 53504, 21384, 28698, 45600,

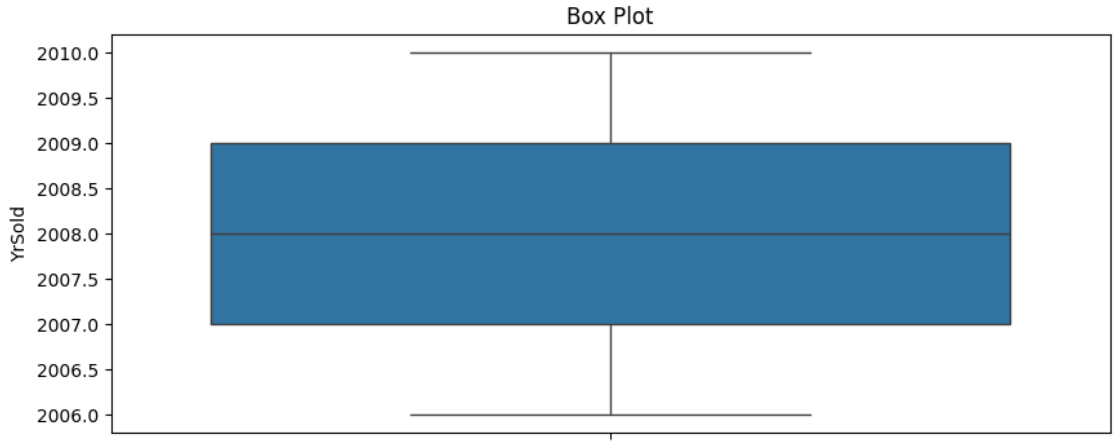
```
25286, 27650, 24090, 25000, 21286, 21750, 29959, 23257, 35760, 35133, 32463,
24682, 23595, 36500, 63887, 20781, 25339, 57200, 20544, 21930, 26142, 50271,
31770, 159000, 39104, 215245, 164660, 53107, 34650, 70761, 53227, 40094, 32668,
46589, 115149, 53504, 45600, 35760, 35133, 32463, 36500, 63887, 57200, 159000,
215245, 164660, 53107, 70761, 53227, 115149, 53504, 63887, 57200, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872]
```

[ ]:
```
Zscore_outlier(data['YrSold'],umbral=1950)
```

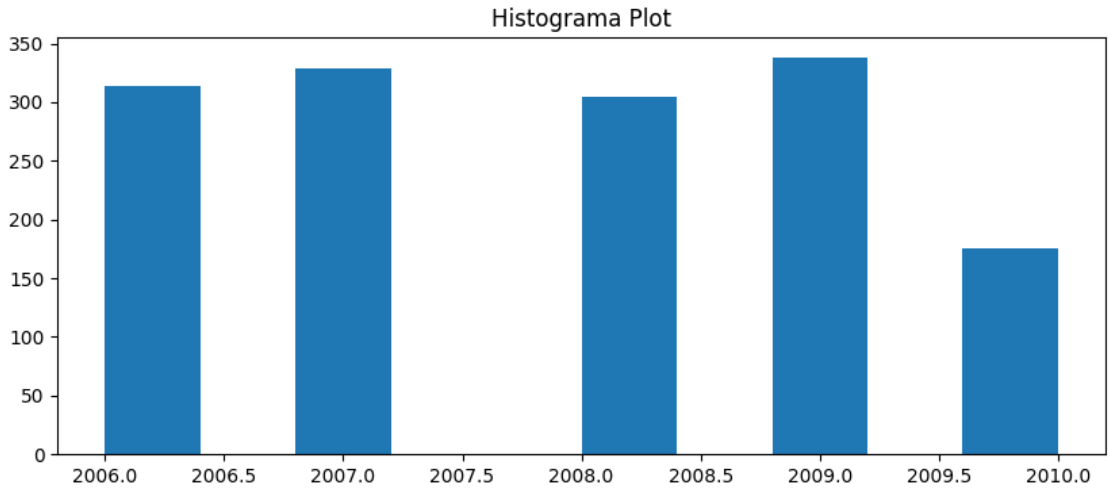
```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 1880, 1880,
1880, 1880, 1875, 1872, 2003, 1915, 2004, 1931, 1939, 2005, 2006, 1929, 2004,
2005, 1930, 2002, 2007, 2007, 1927, 1920, 2007, 2005, 2004, 2005, 2003, 2006,
1920, 1934, 2004, 2006, 2004, 1920, 2006, 1921, 2004, 2003, 1915, 1910, 2007,
2005, 2009, 1915, 1921, 1910, 1920, 2009, 1931, 2003, 1885, 1919, 2007, 2006,
2005, 1939, 1935, 1930, 2005, 2004, 1931, 2004, 1936, 2007, 1923, 1924, 2009,
2004, 2005, 2003, 2005, 1926, 1940, 2007, 2004, 1941, 2008, 1923, 1920, 2003,
1908, 1892, 1916, 1932, 2004, 2007, 1918, 1912, 2004, 2003, 1924, 2004, 1925,
2009, 2009, 2004, 1925, 1939, 2005, 2006, 2002, 2003, 2005, 2002, 2004, 2007,
2008, 1900, 1910, 2003, 1940, 2006, 2004, 2003, 2006, 2007, 1929, 1925, 1939,
1940, 2006, 1925, 2003, 2006, 2006, 2007, 2008, 2006, 1915, 2006, 1912, 1880,
2004, 1920, 1940, 2003, 1939, 1925, 2004, 2006, 2006, 2004, 1941, 1916, 1920,
2003, 2004, 2005, 2002, 2002, 2005, 1939, 2003, 2005, 2007, 1941, 1928, 1940,
1940, 2003, 1920, 2003, 1922, 2004, 2010, 1924, 2006, 2006, 1928, 2004, 1910,
2007, 1900, 1941, 1940, 1920, 2006, 2005, 1940, 1936, 1915, 2006, 2007, 2009,
1927, 2007, 1918, 1940, 2007, 1920, 1920, 1926, 1913, 1920, 2008, 1930, 2006,
1937, 1930, 2008, 1916, 1925, 2009, 1936, 1934, 2004, 2006, 2005, 2006, 2006,
2007, 1937, 2004, 2003, 1915, 1900, 1941, 2006, 1938, 1920, 1925, 2005, 2009,
1928, 2005, 1926, 2009, 1926, 1900, 2007, 2005, 2008, 1920, 1920, 2004, 1910,
2006, 2006, 1923, 2003, 2006, 2003, 1922, 1920, 2003, 1940, 1918, 1915, 2005,
2004, 2009, 1928, 2007, 2008, 1893, 1935, 2005, 1918, 1930, 2004, 2008, 2003,
2005, 1910, 2006, 2005, 1937, 2004, 2002, 1934, 2007, 2002, 2007, 2003, 1914,
1880, 2006, 1914, 1936, 1910, 2006, 2003, 2009, 2007, 1940, 1906, 1931, 2005,
1922, 2005, 1925, 1900, 1924, 2008, 1932, 2002, 2007, 2004, 2007, 2005, 2004,
1921, 1936, 1921, 2004, 2002, 2006, 1900, 2004, 1930, 2006, 2007, 1935, 1900,
1890, 2004, 2007, 2007, 1925, 2003, 1914, 2005, 2004, 1910, 1880, 1910, 2003,
2005, 2007, 1924, 2009, 2008, 1940, 2004, 2003, 2006, 2005, 1914, 1915, 2004,
2005, 2007, 2008, 1937, 1916, 2005, 2008, 2008, 1923, 1898, 2004, 1918, 2002,
2009, 2003, 2003, 1940, 2006, 2007, 1924, 2005, 2005, 2003, 1925, 1904, 1915,
1908, 2003, 2003, 1941, 1918, 2007, 2006, 1941, 2007, 2004, 2005, 1912, 1936,
2009, 2003, 2006, 2006, 1936, 2005, 1925, 2009, 1900, 2005, 2003, 2007, 2006,
2004, 1926, 2003, 2005, 2006, 1940, 1890, 2002, 2002, 2003, 2007, 2002, 2005,
1940, 1910, 2003, 2007, 1910, 1923, 2006, 2007, 2002, 1910, 2009, 2006, 1882,
2005, 2006, 1922, 2006, 1920, 2006, 2005, 2004, 1926, 1923, 1910, 2005, 2005,
2006, 1930, 2005, 2007, 1941, 1916, 1920, 2002, 1938, 2007, 2005, 2005, 2007,
```

```
2007, 2002, 2005, 2008, 1932, 1935, 1900, 1925, 2006, 1940, 1936, 2004, 2002,
2005, 2004, 2005, 1925, 2006, 1914, 1936, 1920, 2006, 2004, 2007, 2002, 1920,
2005, 2007, 2004, 2004, 1928, 1880, 1926, 1875, 1920, 2006, 1941, 1928, 1941,
1926, 1920, 1930, 2007, 2007, 2009, 2008, 1935, 2006, 1932, 1926, 1921, 2008,
1920, 1924, 1900, 2002, 1925, 2005, 2006, 1916, 1925, 2003, 2006, 1941, 2009,
2004, 2007, 2008, 1911, 1914, 2003, 2004, 2005, 2006, 2003, 2007, 2006, 1929,
2005, 1917, 2003, 2003, 1931, 1922, 2005, 1919, 1910, 2008, 1935, 1910, 2002,
1920, 2002, 1919, 1939, 2006, 1892, 2005, 2008, 2005, 2004, 2006, 2005, 2005,
2008, 2006, 1940, 2006, 1922, 1931, 1920, 2006, 1938, 1938, 1941, 2002, 2003,
2002, 1928, 2006, 1920, 2006, 1872, 1937, 2004, 1921, 2005, 1920, 2006, 2005,
2003, 2003, 1920, 2005, 2007, 1930, 1941, 2006, 1914, 1920, 1918, 1939, 1922,
1916, 2006, 1941, 1905, 2006, 2005, 1920, 1925, 1929, 2004, 2006, 2007, 1915,
2004, 1910, 2005, 1923, 2007, 1885, 2003, 1940, 2005, 1927, 2008, 1922, 2004,
2008, 1916, 2004, 1910, 2008, 2005, 2006, 2004, 1941, 1880, 1880, 1880, 1880,
1875, 1872, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200]
```

[ ]:
```
box_plots(data['YrSold'])
```

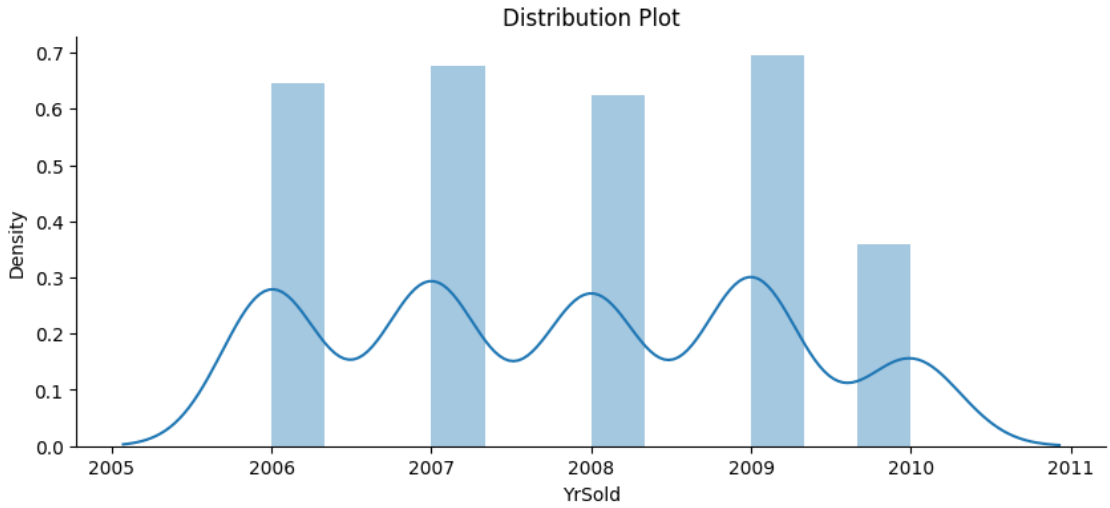


[ ]:
```
hists_plot(data['YrSold'])
```



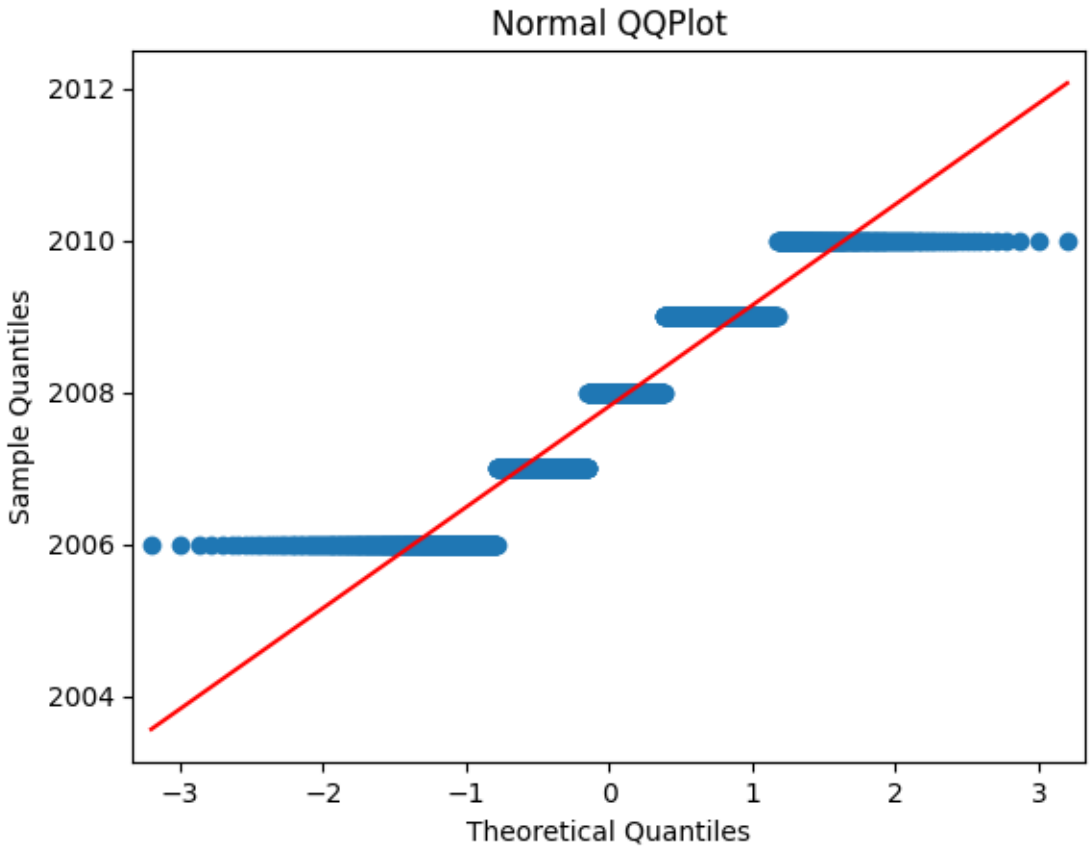
[ ]:
```
dist_plot(data['YrSold'])
```

```
<ipython-input-17-70e5f0238ccb>:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df)
```



[ ]:
```
qq_plots(data['YrSold'])
```

```
<Figure size 1000x400 with 0 Axes>
```



Agregué este ejercicio porque llamo mi atención al ver que no se detectan como tal atípicos en las gráficas o no son facil de observar si es que los hay.

1.5 4.-LotArea

Sus datos atípicos:

[ ]:
```
Zscore_outlier(data['LotArea'],umbral=3)
```

```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 1880, 1880,
1880, 1880, 1875, 1872, 2003, 1915, 2004, 1931, 1939, 2005, 2006, 1929, 2004,
2005, 1930, 2002, 2007, 2007, 1927, 1920, 2007, 2005, 2004, 2005, 2003, 2006,
1920, 1934, 2004, 2006, 2004, 1920, 2006, 1921, 2004, 2003, 1915, 1910, 2007,
2005, 2009, 1915, 1921, 1910, 1920, 2009, 1931, 2003, 1885, 1919, 2007, 2006,
2005, 1939, 1935, 1930, 2005, 2004, 1931, 2004, 1936, 2007, 1923, 1924, 2009,
```

2004, 2005, 2003, 2005, 1926, 1940, 2007, 2004, 1941, 2008, 1923, 1920, 2003,
1908, 1892, 1916, 1932, 2004, 2007, 1918, 1912, 2004, 2003, 1924, 2004, 1925,
2009, 2009, 2004, 1925, 1939, 2005, 2006, 2002, 2003, 2005, 2002, 2004, 2007,
2008, 1900, 1910, 2003, 1940, 2006, 2004, 2003, 2006, 2007, 1929, 1925, 1939,
1940, 2006, 1925, 2003, 2006, 2006, 2007, 2008, 2006, 1915, 2006, 1912, 1880,
2004, 1920, 1940, 2003, 1939, 1925, 2004, 2006, 2006, 2004, 1941, 1916, 1920,
2003, 2004, 2005, 2002, 2002, 2005, 1939, 2003, 2005, 2007, 1941, 1928, 1940,
1940, 2003, 1920, 2003, 1922, 2004, 2010, 1924, 2006, 2006, 1928, 2004, 1910,
2007, 1900, 1941, 1940, 1920, 2006, 2005, 1940, 1936, 1915, 2006, 2007, 2009,
1927, 2007, 1918, 1940, 2007, 1920, 1920, 1926, 1913, 1920, 2008, 1930, 2006,
1937, 1930, 2008, 1916, 1925, 2009, 1936, 1934, 2004, 2006, 2005, 2006, 2006,
2007, 1937, 2004, 2003, 1915, 1900, 1941, 2006, 1938, 1920, 1925, 2005, 2009,
1928, 2005, 1926, 2009, 1926, 1900, 2007, 2005, 2008, 1920, 1920, 2004, 1910,
2006, 2006, 1923, 2003, 2006, 2003, 1922, 1920, 2003, 1940, 1918, 1915, 2005,
2004, 2009, 1928, 2007, 2008, 1893, 1935, 2005, 1918, 1930, 2004, 2008, 2003,
2005, 1910, 2006, 2005, 1937, 2004, 2002, 1934, 2007, 2002, 2007, 2003, 1914,
1880, 2006, 1914, 1936, 1910, 2006, 2003, 2009, 2007, 1940, 1906, 1931, 2005,
1922, 2005, 1925, 1900, 1924, 2008, 1932, 2002, 2007, 2004, 2007, 2005, 2004,
1921, 1936, 1921, 2004, 2002, 2006, 1900, 2004, 1930, 2006, 2007, 1935, 1900,
1890, 2004, 2007, 2007, 1925, 2003, 1914, 2005, 2004, 1910, 1880, 1910, 2003,
2005, 2007, 1924, 2009, 2008, 1940, 2004, 2003, 2006, 2005, 1914, 1915, 2004,
2005, 2007, 2008, 1937, 1916, 2005, 2008, 2008, 1923, 1898, 2004, 1918, 2002,
2009, 2003, 2003, 1940, 2006, 2007, 1924, 2005, 2005, 2003, 1925, 1904, 1915,
1908, 2003, 2003, 1941, 1918, 2007, 2006, 1941, 2007, 2004, 2005, 1912, 1936,
2009, 2003, 2006, 2006, 1936, 2005, 1925, 2009, 1900, 2005, 2003, 2007, 2006,
2004, 1926, 2003, 2005, 2006, 1940, 1890, 2002, 2002, 2003, 2007, 2002, 2005,
1940, 1910, 2003, 2007, 1910, 1923, 2006, 2007, 2002, 1910, 2009, 2006, 1882,
2005, 2006, 1922, 2006, 1920, 2006, 2005, 2004, 1926, 1923, 1910, 2005, 2005,
2006, 1930, 2005, 2007, 1941, 1916, 1920, 2002, 1938, 2007, 2005, 2005, 2007,
2007, 2002, 2005, 2008, 1932, 1935, 1900, 1925, 2006, 1940, 1936, 2004, 2002,
2005, 2004, 2005, 1925, 2006, 1914, 1936, 1920, 2006, 2004, 2007, 2002, 1920,
2005, 2007, 2004, 2004, 1928, 1880, 1926, 1875, 1920, 2006, 1941, 1928, 1941,
1926, 1920, 1930, 2007, 2007, 2009, 2008, 1935, 2006, 1932, 1926, 1921, 2008,
1920, 1924, 1900, 2002, 1925, 2005, 2006, 1916, 1925, 2003, 2006, 1941, 2009,
2004, 2007, 2008, 1911, 1914, 2003, 2004, 2005, 2006, 2003, 2007, 2006, 1929,
2005, 1917, 2003, 2003, 1931, 1922, 2005, 1919, 1910, 2008, 1935, 1910, 2002,
1920, 2002, 1919, 1939, 2006, 1892, 2005, 2008, 2005, 2004, 2006, 2005, 2005,
2008, 2006, 1940, 2006, 1922, 1931, 1920, 2006, 1938, 1938, 1941, 2002, 2003,
2002, 1928, 2006, 1920, 2006, 1872, 1937, 2004, 1921, 2005, 1920, 2006, 2005,
2003, 2003, 1920, 2005, 2007, 1930, 1941, 2006, 1914, 1920, 1918, 1939, 1922,
1916, 2006, 1941, 1905, 2006, 2005, 1920, 1925, 1929, 2004, 2006, 2007, 1915,
2004, 1910, 2005, 1923, 2007, 1885, 2003, 1940, 2005, 1927, 2008, 1922, 2004,
2008, 1916, 2004, 1910, 2008, 2005, 2006, 2004, 1941, 1880, 1880, 1880, 1880,
1875, 1872, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 21000, 21453, 31770, 22950, 25419, 159000, 39104, 215245,

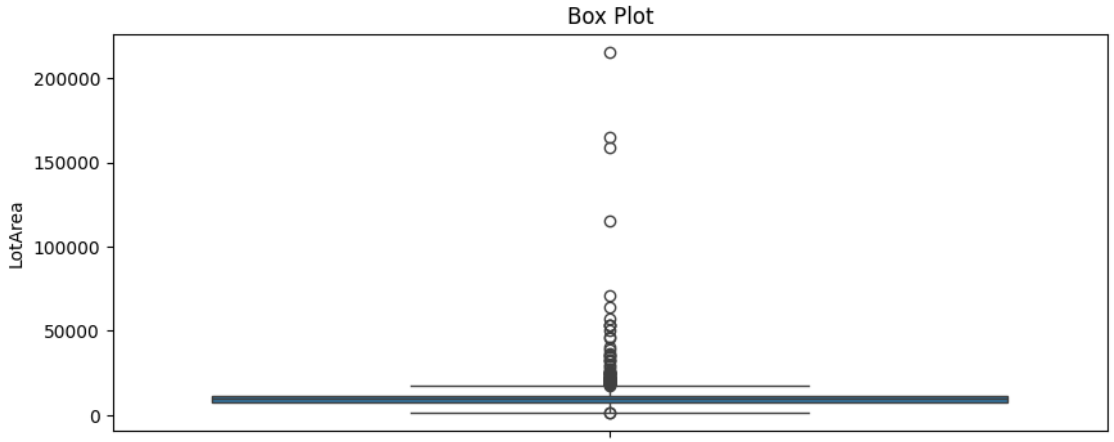
```
164660, 53107, 34650, 22420, 21750, 70761, 53227, 40094, 32668, 21872, 21780,
25095, 46589, 20896, 21535, 26178, 115149, 21695, 53504, 21384, 28698, 45600,
25286, 27650, 24090, 25000, 21286, 21750, 29959, 23257, 35760, 35133, 32463,
24682, 23595, 36500, 63887, 20781, 25339, 57200, 20544, 21930, 26142, 50271,
31770, 159000, 39104, 215245, 164660, 53107, 34650, 70761, 53227, 40094, 32668,
46589, 115149, 53504, 45600, 35760, 35133, 32463, 36500, 63887, 57200, 159000,
215245, 164660, 53107, 70761, 53227, 115149, 53504, 63887, 57200, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200]
```

[ ]:
```
Zscore_outlier(data['LotArea'],umbral=150000)
```

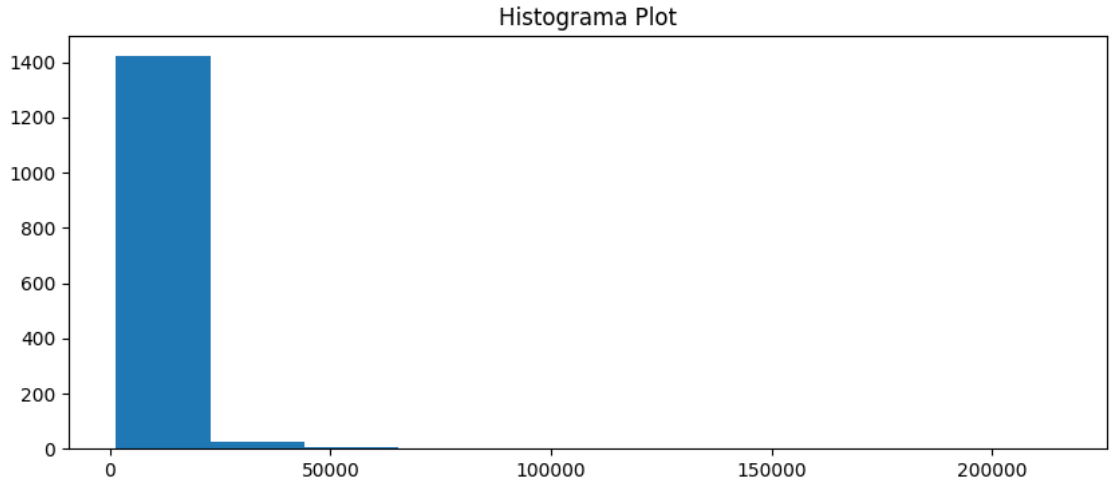
```
Outliers:  [50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 1880, 1880,
1880, 1880, 1875, 1872, 2003, 1915, 2004, 1931, 1939, 2005, 2006, 1929, 2004,
2005, 1930, 2002, 2007, 2007, 1927, 1920, 2007, 2005, 2004, 2005, 2003, 2006,
1920, 1934, 2004, 2006, 2004, 1920, 2006, 1921, 2004, 2003, 1915, 1910, 2007,
2005, 2009, 1915, 1921, 1910, 1920, 2009, 1931, 2003, 1885, 1919, 2007, 2006,
2005, 1939, 1935, 1930, 2005, 2004, 1931, 2004, 1936, 2007, 1923, 1924, 2009,
2004, 2005, 2003, 2005, 1926, 1940, 2007, 2004, 1941, 2008, 1923, 1920, 2003,
1908, 1892, 1916, 1932, 2004, 2007, 1918, 1912, 2004, 2003, 1924, 2004, 1925,
2009, 2009, 2004, 1925, 1939, 2005, 2006, 2002, 2003, 2005, 2002, 2004, 2007,
2008, 1900, 1910, 2003, 1940, 2006, 2004, 2003, 2006, 2007, 1929, 1925, 1939,
1940, 2006, 1925, 2003, 2006, 2006, 2007, 2008, 2006, 1915, 2006, 1912, 1880,
2004, 1920, 1940, 2003, 1939, 1925, 2004, 2006, 2006, 2004, 1941, 1916, 1920,
2003, 2004, 2005, 2002, 2002, 2005, 1939, 2003, 2005, 2007, 1941, 1928, 1940,
1940, 2003, 1920, 2003, 1922, 2004, 2010, 1924, 2006, 2006, 1928, 2004, 1910,
2007, 1900, 1941, 1940, 1920, 2006, 2005, 1940, 1936, 1915, 2006, 2007, 2009,
1927, 2007, 1918, 1940, 2007, 1920, 1920, 1926, 1913, 1920, 2008, 1930, 2006,
1937, 1930, 2008, 1916, 1925, 2009, 1936, 1934, 2004, 2006, 2005, 2006, 2006,
2007, 1937, 2004, 2003, 1915, 1900, 1941, 2006, 1938, 1920, 1925, 2005, 2009,
1928, 2005, 1926, 2009, 1926, 1900, 2007, 2005, 2008, 1920, 1920, 2004, 1910,
2006, 2006, 1923, 2003, 2006, 2003, 1922, 1920, 2003, 1940, 1918, 1915, 2005,
2004, 2009, 1928, 2007, 2008, 1893, 1935, 2005, 1918, 1930, 2004, 2008, 2003,
2005, 1910, 2006, 2005, 1937, 2004, 2002, 1934, 2007, 2002, 2007, 2003, 1914,
1880, 2006, 1914, 1936, 1910, 2006, 2003, 2009, 2007, 1940, 1906, 1931, 2005,
1922, 2005, 1925, 1900, 1924, 2008, 1932, 2002, 2007, 2004, 2007, 2005, 2004,
1921, 1936, 1921, 2004, 2002, 2006, 1900, 2004, 1930, 2006, 2007, 1935, 1900,
1890, 2004, 2007, 2007, 1925, 2003, 1914, 2005, 2004, 1910, 1880, 1910, 2003,
2005, 2007, 1924, 2009, 2008, 1940, 2004, 2003, 2006, 2005, 1914, 1915, 2004,
2005, 2007, 2008, 1937, 1916, 2005, 2008, 2008, 1923, 1898, 2004, 1918, 2002,
2009, 2003, 2003, 1940, 2006, 2007, 1924, 2005, 2005, 2003, 1925, 1904, 1915,
1908, 2003, 2003, 1941, 1918, 2007, 2006, 1941, 2007, 2004, 2005, 1912, 1936,
```

```
2009, 2003, 2006, 2006, 1936, 2005, 1925, 2009, 1900, 2005, 2003, 2007, 2006,
2004, 1926, 2003, 2005, 2006, 1940, 1890, 2002, 2002, 2003, 2007, 2002, 2005,
1940, 1910, 2003, 2007, 1910, 1923, 2006, 2007, 2002, 1910, 2009, 2006, 1882,
2005, 2006, 1922, 2006, 1920, 2006, 2005, 2004, 1926, 1923, 1910, 2005, 2005,
2006, 1930, 2005, 2007, 1941, 1916, 1920, 2002, 1938, 2007, 2005, 2005, 2007,
2007, 2002, 2005, 2008, 1932, 1935, 1900, 1925, 2006, 1940, 1936, 2004, 2002,
2005, 2004, 2005, 1925, 2006, 1914, 1936, 1920, 2006, 2004, 2007, 2002, 1920,
2005, 2007, 2004, 2004, 1928, 1880, 1926, 1875, 1920, 2006, 1941, 1928, 1941,
1926, 1920, 1930, 2007, 2007, 2009, 2008, 1935, 2006, 1932, 1926, 1921, 2008,
1920, 1924, 1900, 2002, 1925, 2005, 2006, 1916, 1925, 2003, 2006, 1941, 2009,
2004, 2007, 2008, 1911, 1914, 2003, 2004, 2005, 2006, 2003, 2007, 2006, 1929,
2005, 1917, 2003, 2003, 1931, 1922, 2005, 1919, 1910, 2008, 1935, 1910, 2002,
1920, 2002, 1919, 1939, 2006, 1892, 2005, 2008, 2005, 2004, 2006, 2005, 2005,
2008, 2006, 1940, 2006, 1922, 1931, 1920, 2006, 1938, 1938, 1941, 2002, 2003,
2002, 1928, 2006, 1920, 2006, 1872, 1937, 2004, 1921, 2005, 1920, 2006, 2005,
2003, 2003, 1920, 2005, 2007, 1930, 1941, 2006, 1914, 1920, 1918, 1939, 1922,
1916, 2006, 1941, 1905, 2006, 2005, 1920, 1925, 1929, 2004, 2006, 2007, 1915,
2004, 1910, 2005, 1923, 2007, 1885, 2003, 1940, 2005, 1927, 2008, 1922, 2004,
2008, 1916, 2004, 1910, 2008, 2005, 2006, 2004, 1941, 1880, 1880, 1880, 1880,
1875, 1872, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 21000, 21453, 31770, 22950, 25419, 159000, 39104, 215245,
164660, 53107, 34650, 22420, 21750, 70761, 53227, 40094, 32668, 21872, 21780,
25095, 46589, 20896, 21535, 26178, 115149, 21695, 53504, 21384, 28698, 45600,
25286, 27650, 24090, 25000, 21286, 21750, 29959, 23257, 35760, 35133, 32463,
24682, 23595, 36500, 63887, 20781, 25339, 57200, 20544, 21930, 26142, 50271,
31770, 159000, 39104, 215245, 164660, 53107, 34650, 70761, 53227, 40094, 32668,
46589, 115149, 53504, 45600, 35760, 35133, 32463, 36500, 63887, 57200, 159000,
215245, 164660, 53107, 70761, 53227, 115149, 53504, 63887, 57200, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200, 50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149,
53504, 45600, 63887, 57200, 1880, 1880, 1880, 1880, 1875, 1872, 50271, 159000,
215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600, 63887, 57200,
50271, 159000, 215245, 164660, 53107, 70761, 53227, 46589, 115149, 53504, 45600,
63887, 57200]
```

[ ]:
```
box_plots(data['LotArea'])
```

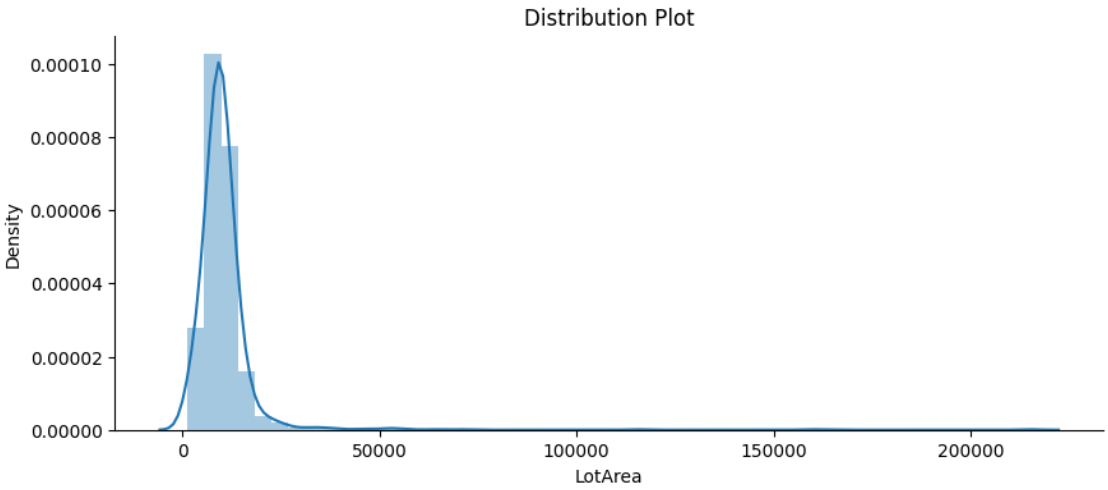


[ ]:
```
hists_plot(data['LotArea'])
```



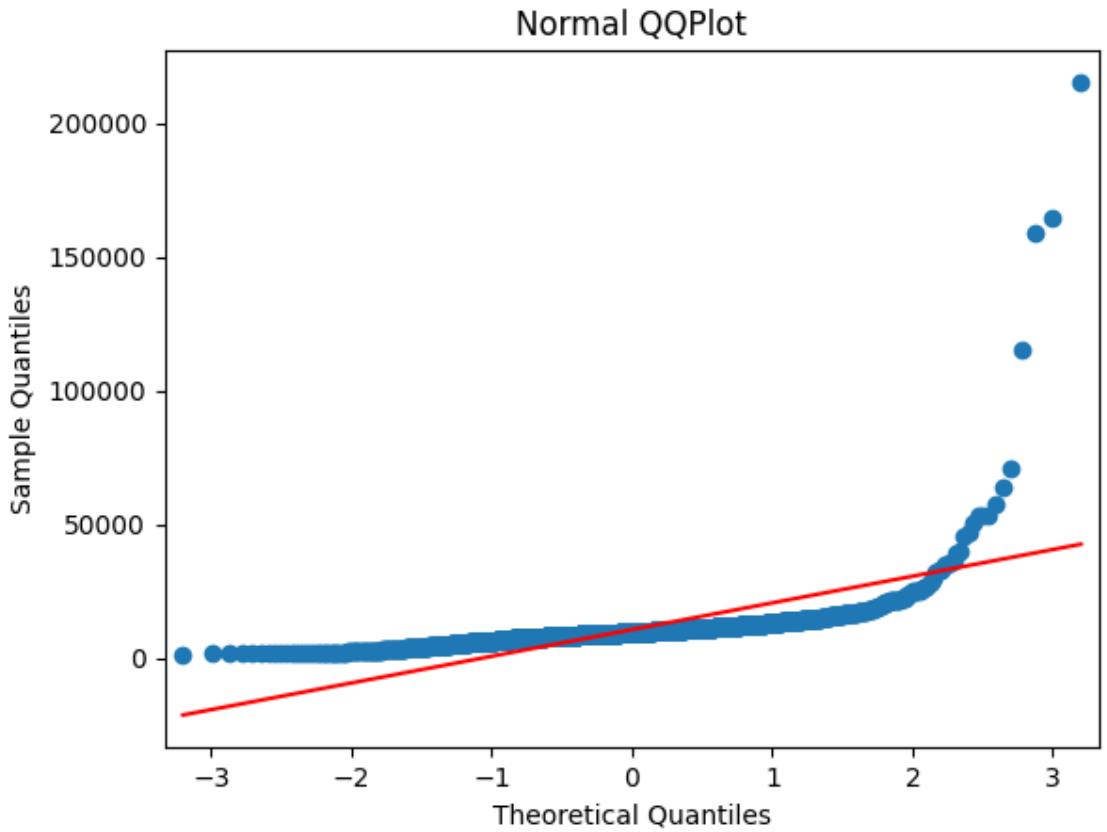
[ ]:
```
dist_plot(data['LotArea'])
```

```
<ipython-input-17-70e5f0238ccb>:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df)
```



[ ]:
```
qq_plots(data['LotArea'])
```

```
<Figure size 1000x400 with 0 Axes>
```



Los datos atípicos se encuentran arriba de los 2500 o 3000