Tarea1_Arevalo_Arancibia

May 5, 2025

1 Tarea 1

1.1 Importando Librerías

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as smf
import sklearn
import scipy
from scipy.stats import nbinom
import seaborn as sns
from statsmodels.iolib.summary2 import summary_col

import warnings
warnings.filterwarnings("ignore")

%matplotlib inline
```

1.2 1. Cargar la base de datos en el ambiente. Identifique los tipos de datos que se encuentran en la base, realice estadisticas descriptivas sobre las variables importantes (Hint: Revisar la distribuciones, datos faltantes, outliers, etc.) y limpie las variables cuando sea necesario.

[2]:
```
df = pd.read_csv('../data/machine_failure_data.csv')
df
```

[2]:
```
             Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
0       12/1/2008         3      13.4      22.9      0.6          NaN
1       12/2/2008         3       7.4      25.1      0.0          NaN
2       12/3/2008         3      12.9      25.7      0.0          NaN
3       12/4/2008         3       9.2      28.0      0.0          NaN
4       12/5/2008         3      17.5      32.3      1.0          NaN
…              …       …       …        …        …          …
142188  6/20/2017        42       3.5      21.8      0.0          NaN
142189  6/21/2017        42       2.8      23.4      0.0          NaN
142190  6/22/2017        42       3.6      25.3      0.0          NaN
```

```
142191  6/23/2017        42       5.4       26.9       0.0           NaN
142192  6/24/2017        42       7.8       27.0       0.0           NaN

        Electricity Parameter1_Dir  Parameter1_Speed Parameter2_9am  …  \
0               NaN              W              44.0              W  …
1               NaN            WNW              44.0            NNW  …
2               NaN            WSW              46.0              W  …
3               NaN             NE              24.0             SE  …
4               NaN              W              41.0            ENE  …
…               …              …                …              …  …
142188          NaN              E              31.0            ESE  …
142189          NaN              E              31.0             SE  …
142190          NaN            NNW              22.0             SE  …
142191          NaN              N              37.0             SE  …
142192          NaN             SE              28.0            SSE  …

       Parameter3_3pm  Parameter4_9am  Parameter4_3pm  Parameter5_9am  \
0                24.0            71.0            22.0          1007.7
1                22.0            44.0            25.0          1010.6
2                26.0            38.0            30.0          1007.6
3                 9.0            45.0            16.0          1017.6
4                20.0            82.0            33.0          1010.8
…                 …              …               …              …
142188           13.0            59.0            27.0          1024.7
142189           11.0            51.0            24.0          1024.6
142190            9.0            56.0            21.0          1023.5
142191            9.0            53.0            24.0          1021.0
142192            7.0            51.0            24.0          1019.4

        Parameter5_3pm  Parameter6_9am  Parameter6_3pm  Parameter7_9am  \
0               1007.1             8.0             NaN            16.9
1               1007.8             NaN             NaN            17.2
2               1008.7             NaN             2.0            21.0
3               1012.8             NaN             NaN            18.1
4               1006.0             7.0             8.0            17.8
…                 …              …               …              …
142188          1021.2             NaN             NaN             9.4
142189          1020.3             NaN             NaN            10.1
142190          1019.1             NaN             NaN            10.9
142191          1016.8             NaN             NaN            12.5
142192          1016.5             3.0             2.0            15.1

        Parameter7_3pm  Failure_today
0                 21.8             No
1                 24.3             No
2                 23.2             No
3                 26.5             No
```

```
4                 29.7            No
…                …            …
142188            20.9            No
142189            22.4            No
142190            24.5            No
142191            26.1            No
142192            26.0            No

[142193 rows x 22 columns]
```

[3]:
```
df.describe()
```

[3]:
```
            Location        Min_Temp        Max_Temp         Leakage  \
count  142193.000000  141556.000000  141871.000000  140787.000000
mean       24.740655      12.186400      23.226784       2.349974
std        14.237503       6.403283       7.117618       8.465173
min         1.000000      -8.500000      -4.800000       0.000000
25%        12.000000       7.600000      17.900000       0.000000
50%        25.000000      12.000000      22.600000       0.000000
75%        37.000000      16.800000      28.200000       0.800000
max        49.000000      33.900000      48.100000     371.000000

        Evaporation   Electricity  Parameter1_Speed  Parameter3_9am  \
count  81350.000000  74377.000000     132923.000000   140845.000000
mean       5.469824      7.624853         39.984292       14.001988
std        4.188537      3.781525         13.588801        8.893337
min        0.000000      0.000000          6.000000        0.000000
25%        2.600000      4.900000         31.000000        7.000000
50%        4.800000      8.500000         39.000000       13.000000
75%        7.400000     10.600000         48.000000       19.000000
max      145.000000     14.500000        135.000000      130.000000

       Parameter3_3pm  Parameter4_9am  Parameter4_3pm  Parameter5_9am  \
count   139563.000000   140419.000000   138583.000000   128179.000000
mean        18.637576       68.843810       51.482606     1017.653758
std          8.803345       19.051293       20.797772        7.105476
min          0.000000        0.000000        0.000000      980.500000
25%         13.000000       57.000000       37.000000     1012.900000
50%         19.000000       70.000000       52.000000     1017.600000
75%         24.000000       83.000000       66.000000     1022.400000
max         87.000000      100.000000      100.000000     1041.000000

       Parameter5_3pm  Parameter6_9am  Parameter6_3pm  Parameter7_9am  \
count   128212.000000    88536.000000    85099.000000   141289.000000
mean      1015.258204        4.437189        4.503167       16.987509
std          7.036677        2.887016        2.720633        6.492838
min        977.100000        0.000000        0.000000       -7.200000
```

```
25%        1010.400000         1.000000         2.000000        12.300000
50%        1015.200000         5.000000         5.000000        16.700000
75%        1020.000000         7.000000         7.000000        21.600000
max        1039.600000         9.000000         9.000000        40.200000

       Parameter7_3pm
count   139467.000000
mean        21.687235
std          6.937594
min         -5.400000
25%         16.600000
50%         21.100000
75%         26.400000
max         46.700000
```

1.2.1 Convirtiendo las fechas de str a formato datetime

[4]:
```
df['Date'] = pd.to_datetime(df['Date'])
df['Date'].head()
```

[4]:
```
0   2008-12-01
1   2008-12-02
2   2008-12-03
3   2008-12-04
4   2008-12-05
Name: Date, dtype: datetime64[ns]
```

1.2.2 Convirtiendo en binario el Failure Today

[5]:
```
df['Failure_today'] = df['Failure_today'].map({'No':0, 'Yes':1})
df['Failure_today'].unique()
```

[5]:
```
array([ 0.,  1., nan])
```

1.2.3 Reduciendo la cantidad de direcciones para construir dummies

[6]:
```
direction_map = {
    'N': 'Norte', 'NNE': 'Norte', 'NE': 'Norte', 'ENE': 'Norte',
    'E': 'Este', 'ESE': 'Este', 'SE': 'Este', 'SSE': 'Este',
    'S': 'Sur', 'SSW': 'Sur', 'SW': 'Sur', 'WSW': 'Sur',
    'W': 'Oeste', 'WNW': 'Oeste', 'NW': 'Oeste', 'NNW': 'Oeste'
}
df['Parameter1_Dir'] = df['Parameter1_Dir'].map(direction_map)
df['Parameter2_9am'] = df['Parameter2_9am'].map(direction_map)
df['Parameter2_3pm'] = df['Parameter2_3pm'].map(direction_map)
df
```

[6]:

```
             Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
0      2008-12-01         3      13.4      22.9      0.6          NaN
1      2008-12-02         3       7.4      25.1      0.0          NaN
2      2008-12-03         3      12.9      25.7      0.0          NaN
3      2008-12-04         3       9.2      28.0      0.0          NaN
4      2008-12-05         3      17.5      32.3      1.0          NaN
…             …       …       …       …       …        …
142188 2017-06-20        42       3.5      21.8      0.0          NaN
142189 2017-06-21        42       2.8      23.4      0.0          NaN
142190 2017-06-22        42       3.6      25.3      0.0          NaN
142191 2017-06-23        42       5.4      26.9      0.0          NaN
142192 2017-06-24        42       7.8      27.0      0.0          NaN

        Electricity Parameter1_Dir  Parameter1_Speed Parameter2_9am  …  \
0               NaN          Oeste              44.0          Oeste  …
1               NaN          Oeste              44.0          Oeste  …
2               NaN            Sur              46.0          Oeste  …
3               NaN          Norte              24.0           Este  …
4               NaN          Oeste              41.0          Norte  …
…             …            …                 …             …  …
142188          NaN           Este              31.0           Este  …
142189          NaN           Este              31.0           Este  …
142190          NaN          Oeste              22.0           Este  …
142191          NaN          Norte              37.0           Este  …
142192          NaN           Este              28.0           Este  …

       Parameter3_3pm  Parameter4_9am  Parameter4_3pm  Parameter5_9am  \
0                24.0            71.0            22.0          1007.7
1                22.0            44.0            25.0          1010.6
2                26.0            38.0            30.0          1007.6
3                 9.0            45.0            16.0          1017.6
4                20.0            82.0            33.0          1010.8
…                …             …             …             …
142188           13.0            59.0            27.0          1024.7
142189           11.0            51.0            24.0          1024.6
142190            9.0            56.0            21.0          1023.5
142191            9.0            53.0            24.0          1021.0
142192            7.0            51.0            24.0          1019.4

        Parameter5_3pm  Parameter6_9am  Parameter6_3pm  Parameter7_9am  \
0               1007.1             8.0             NaN            16.9
1               1007.8             NaN             NaN            17.2
2               1008.7             NaN             2.0            21.0
3               1012.8             NaN             NaN            18.1
4               1006.0             7.0             8.0            17.8
…                 …             …             …             …
142188          1021.2             NaN             NaN             9.4
```

```
142189          1020.3             NaN            NaN             10.1
142190          1019.1             NaN            NaN             10.9
142191          1016.8             NaN            NaN             12.5
142192          1016.5             3.0            2.0             15.1

        Parameter7_3pm  Failure_today
0                 21.8            0.0
1                 24.3            0.0
2                 23.2            0.0
3                 26.5            0.0
4                 29.7            0.0
…                  …             …
142188            20.9            0.0
142189            22.4            0.0
142190            24.5            0.0
142191            26.1            0.0
142192            26.0            0.0

[142193 rows x 22 columns]
```

1.2.4 Ahora haré la columna para mes y año

[7]:
```python
df['Month'] = df['Date'].dt.month
df['Year'] = df['Date'].dt.year
df
```

[7]:
```
             Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
0      2008-12-01         3      13.4      22.9      0.6          NaN
1      2008-12-02         3       7.4      25.1      0.0          NaN
2      2008-12-03         3      12.9      25.7      0.0          NaN
3      2008-12-04         3       9.2      28.0      0.0          NaN
4      2008-12-05         3      17.5      32.3      1.0          NaN
…             …       …       …       …       …       …
142188 2017-06-20        42       3.5      21.8      0.0          NaN
142189 2017-06-21        42       2.8      23.4      0.0          NaN
142190 2017-06-22        42       3.6      25.3      0.0          NaN
142191 2017-06-23        42       5.4      26.9      0.0          NaN
142192 2017-06-24        42       7.8      27.0      0.0          NaN

        Electricity Parameter1_Dir  Parameter1_Speed Parameter2_9am  …  \
0               NaN          Oeste              44.0          Oeste  …
1               NaN          Oeste              44.0          Oeste  …
2               NaN            Sur              46.0          Oeste  …
3               NaN          Norte              24.0           Este  …
4               NaN          Oeste              41.0          Norte  …
…              …            …                 …             …   …
142188          NaN           Este              31.0           Este  …
```

```
142189          NaN          Este            31.0          Este  …
142190          NaN         Oeste            22.0          Este  …
142191          NaN         Norte            37.0          Este  …
142192          NaN          Este            28.0          Este  …

       Parameter4_3pm  Parameter5_9am  Parameter5_3pm  Parameter6_9am  \
0                22.0          1007.7          1007.1             8.0
1                25.0          1010.6          1007.8             NaN
2                30.0          1007.6          1008.7             NaN
3                16.0          1017.6          1012.8             NaN
4                33.0          1010.8          1006.0             7.0
…                 …             …             …             …
142188           27.0          1024.7          1021.2             NaN
142189           24.0          1024.6          1020.3             NaN
142190           21.0          1023.5          1019.1             NaN
142191           24.0          1021.0          1016.8             NaN
142192           24.0          1019.4          1016.5             3.0

        Parameter6_3pm  Parameter7_9am  Parameter7_3pm  Failure_today  Month  \
0                  NaN            16.9            21.8            0.0     12
1                  NaN            17.2            24.3            0.0     12
2                  2.0            21.0            23.2            0.0     12
3                  NaN            18.1            26.5            0.0     12
4                  8.0            17.8            29.7            0.0     12
…                 …             …             …             …    …
142188             NaN             9.4            20.9            0.0      6
142189             NaN            10.1            22.4            0.0      6
142190             NaN            10.9            24.5            0.0      6
142191             NaN            12.5            26.1            0.0      6
142192             2.0            15.1            26.0            0.0      6

        Year
0       2008
1       2008
2       2008
3       2008
4       2008
…      …
142188  2017
142189  2017
142190  2017
142191  2017
142192  2017

[142193 rows x 24 columns]
```

[8]:
```
df = df[~df['Year'].isin([2007, 2008])]
df
```

[8]:
```
                 Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
30      2009-01-01         3      11.3      26.5      0.0          NaN
31      2009-01-02         3       9.6      23.9      0.0          NaN
32      2009-01-03         3      10.5      28.8      0.0          NaN
33      2009-01-04         3      12.3      34.6      0.0          NaN
34      2009-01-05         3      12.9      35.8      0.0          NaN
…              …         …       …         …         …          …
142188  2017-06-20        42       3.5      21.8      0.0          NaN
142189  2017-06-21        42       2.8      23.4      0.0          NaN
142190  2017-06-22        42       3.6      25.3      0.0          NaN
142191  2017-06-23        42       5.4      26.9      0.0          NaN
142192  2017-06-24        42       7.8      27.0      0.0          NaN

        Electricity Parameter1_Dir  Parameter1_Speed Parameter2_9am  …  \
30              NaN          Oeste              56.0          Oeste  …
31              NaN          Oeste              41.0            Sur  …
32              NaN           Este              26.0           Este  …
33              NaN          Oeste              37.0           Este  …
34              NaN          Oeste              41.0          Norte  …
…             …            …                …             …   …
142188          NaN           Este              31.0           Este  …
142189          NaN           Este              31.0           Este  …
142190          NaN          Oeste              22.0           Este  …
142191          NaN          Norte              37.0           Este  …
142192          NaN           Este              28.0           Este  …

       Parameter4_3pm  Parameter5_9am  Parameter5_3pm  Parameter6_9am  \
30               26.0          1004.5          1003.2             NaN
31               22.0          1014.4          1013.1             NaN
32               22.0          1018.7          1014.8             NaN
33               12.0          1015.1          1010.3             NaN
34                9.0          1012.6          1009.2             NaN
…                 …             …             …             …
142188           27.0          1024.7          1021.2             NaN
142189           24.0          1024.6          1020.3             NaN
142190           21.0          1023.5          1019.1             NaN
142191           24.0          1021.0          1016.8             NaN
142192           24.0          1019.4          1016.5             3.0

        Parameter6_3pm  Parameter7_9am  Parameter7_3pm  Failure_today  Month  \
30                 NaN            19.7            25.7            0.0      1
31                 NaN            14.9            22.1            0.0      1
32                 NaN            17.1            26.5            0.0      1
33                 NaN            20.7            33.9            0.0      1
```

```
34              NaN        22.4        34.4        0.0     1
…             …          …          …          …    …
142188          NaN         9.4        20.9        0.0     6
142189          NaN        10.1        22.4        0.0     6
142190          NaN        10.9        24.5        0.0     6
142191          NaN        12.5        26.1        0.0     6
142192          2.0        15.1        26.0        0.0     6

        Year
30      2009
31      2009
32      2009
33      2009
34      2009
…      …
142188  2017
142189  2017
142190  2017
142191  2017
142192  2017

[139886 rows x 24 columns]
```

1.2.5 Ahora para las variables que tienen significativamente muchos valores nan, haré variables para estas para evidenciar el efecto de que estas no estén siendo medidas

[9]:
```python
columnas = list(df.columns)
for i in columnas:
    print(f'La variable {i} tiene un {(df[i].isna().sum()/len(df))*100}%')
```

```
La variable Date tiene un 0.0%
La variable Location tiene un 0.0%
La variable Min_Temp tiene un 0.44893699155026234%
La variable Max_Temp tiene un 0.22732796705889083%
La variable Leakage tiene un 0.9929514032855324%
La variable Evaporation tiene un 43.262370787641366%
La variable Electricity tiene un 48.25214817780192%
La variable Parameter1_Dir tiene un 6.350170853409205%
La variable Parameter1_Speed tiene un 6.30799365197375%
La variable Parameter2_9am tiene un 7.09363338718671%
La variable Parameter2_3pm tiene un 2.687188138913115%
La variable Parameter3_9am tiene un 0.9436255236406788%
La variable Parameter3_3pm tiene un 1.8686644839369202%
La variable Parameter4_9am tiene un 1.2517335544657793%
La variable Parameter4_3pm tiene un 2.5663754771742706%
La variable Parameter5_9am tiene un 9.950960067483521%
La variable Parameter5_3pm tiene un 9.928084297213445%
```

```
La variable Parameter6_9am tiene un 38.10316972391804%
La variable Parameter6_3pm tiene un 40.553021746279114%
La variable Parameter7_9am tiene un 0.6340877571736986%
La variable Parameter7_3pm tiene un 1.939436398209971%
La variable Failure_today tiene un 0.9929514032855324%
La variable Month tiene un 0.0%
La variable Year tiene un 0.0%
```

Destacan Parameter6_9am, Parameter6_3pm, por las descartaré a tener muy pocos datos, por otro lado, para Electricity y Evaporation haré un indicador para medir el efecto que sea medida o no

[10]:
```
df = df.drop(['Parameter6_9am', 'Parameter6_3pm'], axis=1)
```

Ahora generaré estaciones

[11]:
```
mapa_estaciones = {
    1: 1, 2: 1, 12: 1,
    3: 2, 4: 2, 5: 2,
    6: 3, 7: 3, 8: 3,
    9: 4, 10: 4, 11: 4
}
df['Estacion'] = df['Month'].map(mapa_estaciones)
df
```

[11]:
```
              Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
30      2009-01-01         3      11.3      26.5      0.0          NaN
31      2009-01-02         3       9.6      23.9      0.0          NaN
32      2009-01-03         3      10.5      28.8      0.0          NaN
33      2009-01-04         3      12.3      34.6      0.0          NaN
34      2009-01-05         3      12.9      35.8      0.0          NaN
…              …       …       …         …         …          …
142188  2017-06-20        42       3.5      21.8      0.0          NaN
142189  2017-06-21        42       2.8      23.4      0.0          NaN
142190  2017-06-22        42       3.6      25.3      0.0          NaN
142191  2017-06-23        42       5.4      26.9      0.0          NaN
142192  2017-06-24        42       7.8      27.0      0.0          NaN

        Electricity Parameter1_Dir  Parameter1_Speed Parameter2_9am  …  \
30              NaN          Oeste              56.0          Oeste  …
31              NaN          Oeste              41.0            Sur  …
32              NaN           Este              26.0           Este  …
33              NaN          Oeste              37.0           Este  …
34              NaN          Oeste              41.0          Norte  …
…             …            …                …            …   …
142188          NaN           Este              31.0           Este  …
142189          NaN           Este              31.0           Este  …
142190          NaN          Oeste              22.0           Este  …
142191          NaN          Norte              37.0           Este  …
```

```
142192          NaN          Este             28.0          Este  …

       Parameter4_9am  Parameter4_3pm  Parameter5_9am  Parameter5_3pm  \
30               46.0            26.0          1004.5          1003.2
31               44.0            22.0          1014.4          1013.1
32               43.0            22.0          1018.7          1014.8
33               41.0            12.0          1015.1          1010.3
34               41.0             9.0          1012.6          1009.2
…                 …              …              …              …
142188           59.0            27.0          1024.7          1021.2
142189           51.0            24.0          1024.6          1020.3
142190           56.0            21.0          1023.5          1019.1
142191           53.0            24.0          1021.0          1016.8
142192           51.0            24.0          1019.4          1016.5

        Parameter7_9am  Parameter7_3pm  Failure_today  Month  Year  Estacion
30                19.7            25.7            0.0      1  2009         1
31                14.9            22.1            0.0      1  2009         1
32                17.1            26.5            0.0      1  2009         1
33                20.7            33.9            0.0      1  2009         1
34                22.4            34.4            0.0      1  2009         1
…                  …              …              …     …   …       …
142188             9.4            20.9            0.0      6  2017         3
142189            10.1            22.4            0.0      6  2017         3
142190            10.9            24.5            0.0      6  2017         3
142191            12.5            26.1            0.0      6  2017         3
142192            15.1            26.0            0.0      6  2017         3

[139886 rows x 23 columns]
```

[12]:
```
df_modelo = df
df_modelo
```

[12]:
```
              Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
30      2009-01-01         3      11.3      26.5      0.0          NaN
31      2009-01-02         3       9.6      23.9      0.0          NaN
32      2009-01-03         3      10.5      28.8      0.0          NaN
33      2009-01-04         3      12.3      34.6      0.0          NaN
34      2009-01-05         3      12.9      35.8      0.0          NaN
…              …        …       …        …        …          …
142188  2017-06-20        42       3.5      21.8      0.0          NaN
142189  2017-06-21        42       2.8      23.4      0.0          NaN
142190  2017-06-22        42       3.6      25.3      0.0          NaN
142191  2017-06-23        42       5.4      26.9      0.0          NaN
142192  2017-06-24        42       7.8      27.0      0.0          NaN

        Electricity Parameter1_Dir  Parameter1_Speed Parameter2_9am  …  \
```

```
30              NaN         Oeste             56.0          Oeste  …
31              NaN         Oeste             41.0            Sur  …
32              NaN          Este             26.0           Este  …
33              NaN         Oeste             37.0           Este  …
34              NaN         Oeste             41.0          Norte  …
…                 …             …                …              …  …
142188          NaN          Este             31.0           Este  …
142189          NaN          Este             31.0           Este  …
142190          NaN         Oeste             22.0           Este  …
142191          NaN         Norte             37.0           Este  …
142192          NaN          Este             28.0           Este  …

       Parameter4_9am  Parameter4_3pm  Parameter5_9am  Parameter5_3pm  \
30               46.0            26.0          1004.5          1003.2
31               44.0            22.0          1014.4          1013.1
32               43.0            22.0          1018.7          1014.8
33               41.0            12.0          1015.1          1010.3
34               41.0             9.0          1012.6          1009.2
…                   …               …               …               …
142188           59.0            27.0          1024.7          1021.2
142189           51.0            24.0          1024.6          1020.3
142190           56.0            21.0          1023.5          1019.1
142191           53.0            24.0          1021.0          1016.8
142192           51.0            24.0          1019.4          1016.5

        Parameter7_9am  Parameter7_3pm  Failure_today  Month  Year  Estacion
30                19.7            25.7            0.0      1  2009         1
31                14.9            22.1            0.0      1  2009         1
32                17.1            26.5            0.0      1  2009         1
33                20.7            33.9            0.0      1  2009         1
34                22.4            34.4            0.0      1  2009         1
…                    …               …              …      …     …         …
142188             9.4            20.9            0.0      6  2017         3
142189            10.1            22.4            0.0      6  2017         3
142190            10.9            24.5            0.0      6  2017         3
142191            12.5            26.1            0.0      6  2017         3
142192            15.1            26.0            0.0      6  2017         3

[139886 rows x 23 columns]
```

[13]:

```python
df_modelo['I_Electricity'] = df_modelo['Electricity'].notna().astype(int)
df_modelo['I_Evaporation'] = df_modelo['Evaporation'].notna().astype(int)
df_modelo['Electricity'] = df_modelo['Electricity'].fillna(value=0)
df_modelo['Evaporation'] = df_modelo['Evaporation'].fillna(value=0)
df_modelo
```

[13]:

```
              Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
30      2009-01-01         3      11.3      26.5      0.0          0.0
31      2009-01-02         3       9.6      23.9      0.0          0.0
32      2009-01-03         3      10.5      28.8      0.0          0.0
33      2009-01-04         3      12.3      34.6      0.0          0.0
34      2009-01-05         3      12.9      35.8      0.0          0.0
…              …       …       …         …       …         …
142188  2017-06-20        42       3.5      21.8      0.0          0.0
142189  2017-06-21        42       2.8      23.4      0.0          0.0
142190  2017-06-22        42       3.6      25.3      0.0          0.0
142191  2017-06-23        42       5.4      26.9      0.0          0.0
142192  2017-06-24        42       7.8      27.0      0.0          0.0

        Electricity Parameter1_Dir  Parameter1_Speed Parameter2_9am  …  \
30              0.0          Oeste              56.0          Oeste  …
31              0.0          Oeste              41.0            Sur  …
32              0.0           Este              26.0           Este  …
33              0.0          Oeste              37.0           Este  …
34              0.0          Oeste              41.0          Norte  …
…                 …              …                 …            …  …
142188          0.0           Este              31.0           Este  …
142189          0.0           Este              31.0           Este  …
142190          0.0          Oeste              22.0           Este  …
142191          0.0          Norte              37.0           Este  …
142192          0.0           Este              28.0           Este  …

       Parameter5_9am  Parameter5_3pm  Parameter7_9am  Parameter7_3pm  \
30             1004.5          1003.2            19.7            25.7
31             1014.4          1013.1            14.9            22.1
32             1018.7          1014.8            17.1            26.5
33             1015.1          1010.3            20.7            33.9
34             1012.6          1009.2            22.4            34.4
…                   …               …               …               …
142188         1024.7          1021.2             9.4            20.9
142189         1024.6          1020.3            10.1            22.4
142190         1023.5          1019.1            10.9            24.5
142191         1021.0          1016.8            12.5            26.1
142192         1019.4          1016.5            15.1            26.0

        Failure_today  Month  Year  Estacion  I_Electricity  I_Evaporation
30                0.0      1  2009         1              0              0
31                0.0      1  2009         1              0              0
32                0.0      1  2009         1              0              0
33                0.0      1  2009         1              0              0
34                0.0      1  2009         1              0              0
…                   …    …     …         …              …              …
142188            0.0      6  2017         3              0              0
```

```
142189           0.0     6  2017        3           0           0
142190           0.0     6  2017        3           0           0
142191           0.0     6  2017        3           0           0
142192           0.0     6  2017        3           0           0

[139886 rows x 25 columns]
```

[14]:
```python
df_modelo = pd.get_dummies(df_modelo, columns=['Parameter1_Dir'],␣
 ↪prefix='Parameter1_Dir', drop_first=True, dtype=int)
df_modelo = pd.get_dummies(df_modelo, columns=['Parameter2_9am'],␣
 ↪prefix='Parameter2_9am', drop_first=True, dtype=int)
df_modelo = pd.get_dummies(df_modelo, columns=['Parameter2_3pm'],␣
 ↪prefix='Parameter2_3pm', drop_first=True, dtype=int)
df_modelo
```

[14]:
```
             Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
30     2009-01-01         3      11.3      26.5      0.0          0.0
31     2009-01-02         3       9.6      23.9      0.0          0.0
32     2009-01-03         3      10.5      28.8      0.0          0.0
33     2009-01-04         3      12.3      34.6      0.0          0.0
34     2009-01-05         3      12.9      35.8      0.0          0.0
…             …         …         …         …        …          …
142188 2017-06-20        42       3.5      21.8      0.0          0.0
142189 2017-06-21        42       2.8      23.4      0.0          0.0
142190 2017-06-22        42       3.6      25.3      0.0          0.0
142191 2017-06-23        42       5.4      26.9      0.0          0.0
142192 2017-06-24        42       7.8      27.0      0.0          0.0

        Electricity  Parameter1_Speed  Parameter3_9am  Parameter3_3pm  …  \
30              0.0              56.0            19.0            31.0  …
31              0.0              41.0            19.0            11.0  …
32              0.0              26.0            11.0             7.0  …
33              0.0              37.0             6.0            17.0  …
34              0.0              41.0             6.0            26.0  …
…                 …                 …               …               …  …
142188          0.0              31.0            15.0            13.0  …
142189          0.0              31.0            13.0            11.0  …
142190          0.0              22.0            13.0             9.0  …
142191          0.0              37.0             9.0             9.0  …
142192          0.0              28.0            13.0             7.0  …

        I_Evaporation  Parameter1_Dir_Norte  Parameter1_Dir_Oeste  \
30                  0                     0                     1
31                  0                     0                     1
32                  0                     0                     0
33                  0                     0                     1
34                  0                     0                     1
```

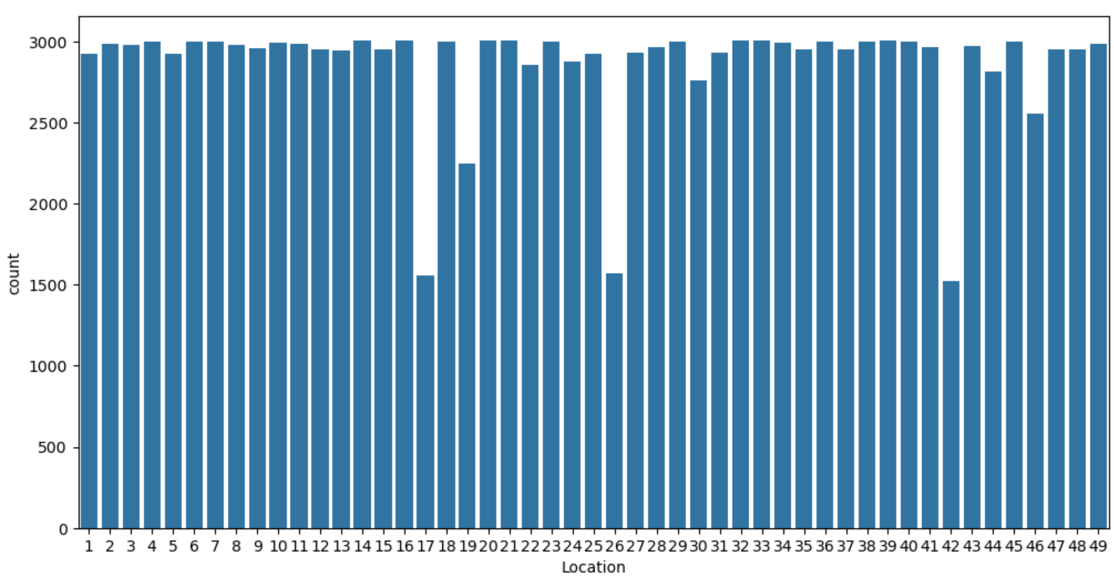
```
…                   …                   …                   …
142188              0                   0                   0
142189              0                   0                   0
142190              0                   0                   1
142191              0                   1                   0
142192              0                   0                   0

        Parameter1_Dir_Sur  Parameter2_9am_Norte  Parameter2_9am_Oeste  \
30                       0                     0                     1
31                       0                     0                     0
32                       0                     0                     0
33                       0                     0                     0
34                       0                     1                     0
…                     …                   …                   …
142188                   0                     0                     0
142189                   0                     0                     0
142190                   0                     0                     0
142191                   0                     0                     0
142192                   0                     0                     0

        Parameter2_9am_Sur  Parameter2_3pm_Norte  Parameter2_3pm_Oeste  \
30                       0                     0                     1
31                       1                     0                     0
32                       0                     0                     0
33                       0                     0                     1
34                       0                     0                     1
…                     …                   …                   …
142188                   0                     0                     0
142189                   0                     1                     0
142190                   0                     1                     0
142191                   0                     0                     1
142192                   0                     1                     0

        Parameter2_3pm_Sur
30                       0
31                       1
32                       0
33                       0
34                       0
…                     …
142188                   0
142189                   0
142190                   0
142191                   0
142192                   0

[139886 rows x 31 columns]
```

[15]:
```python
df_modelo = pd.get_dummies(df_modelo, columns=['Estacion'], prefix='Estacion',
    drop_first=True, dtype=int)
df_modelo
```

[15]:
```
              Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
30      2009-01-01         3      11.3      26.5      0.0          0.0
31      2009-01-02         3       9.6      23.9      0.0          0.0
32      2009-01-03         3      10.5      28.8      0.0          0.0
33      2009-01-04         3      12.3      34.6      0.0          0.0
34      2009-01-05         3      12.9      35.8      0.0          0.0
…              …         …         …         …        …            …
142188  2017-06-20        42       3.5      21.8      0.0          0.0
142189  2017-06-21        42       2.8      23.4      0.0          0.0
142190  2017-06-22        42       3.6      25.3      0.0          0.0
142191  2017-06-23        42       5.4      26.9      0.0          0.0
142192  2017-06-24        42       7.8      27.0      0.0          0.0

        Electricity  Parameter1_Speed  Parameter3_9am  Parameter3_3pm  …  \
30              0.0              56.0            19.0            31.0  …
31              0.0              41.0            19.0            11.0  …
32              0.0              26.0            11.0             7.0  …
33              0.0              37.0             6.0            17.0  …
34              0.0              41.0             6.0            26.0  …
…                 …                 …               …               …  …
142188          0.0              31.0            15.0            13.0  …
142189          0.0              31.0            13.0            11.0  …
142190          0.0              22.0            13.0             9.0  …
142191          0.0              37.0             9.0             9.0  …
142192          0.0              28.0            13.0             7.0  …

        Parameter1_Dir_Sur  Parameter2_9am_Norte  Parameter2_9am_Oeste  \
30                       0                     0                     1
31                       0                     0                     0
32                       0                     0                     0
33                       0                     0                     0
34                       0                     1                     0
…                        …                     …                     …
142188                   0                     0                     0
142189                   0                     0                     0
142190                   0                     0                     0
142191                   0                     0                     0
142192                   0                     0                     0

        Parameter2_9am_Sur  Parameter2_3pm_Norte  Parameter2_3pm_Oeste  \
30                       0                     0                     1
31                       1                     0                     0
32                       0                     0                     0
```

```
33                        0                      0                      1
34                        0                      0                      1
…                     …                    …                    …
142188                    0                      0                      0
142189                    0                      1                      0
142190                    0                      1                      0
142191                    0                      0                      1
142192                    0                      1                      0

        Parameter2_3pm_Sur  Estacion_2  Estacion_3  Estacion_4
30                       0           0           0           0
31                       1           0           0           0
32                       0           0           0           0
33                       0           0           0           0
34                       0           0           0           0
…                     …         …         …         …
142188                   0           0           1           0
142189                   0           0           1           0
142190                   0           0           1           0
142191                   0           0           1           0
142192                   0           0           1           0

[139886 rows x 33 columns]
```

[16]:
```python
plt.figure(figsize=(12, 6))
sns.countplot(data=df_modelo, x='Location')
plt.show()
```



1.2.6 No contaré con la Location 17, 26 y 42

[17]:
```
df_modelo = pd.get_dummies(df_modelo, columns=['Location'], prefix='Location',
 ↪drop_first=True, dtype=int)
df_modelo
```

[17]:
```
              Date  Min_Temp  Max_Temp  Leakage  Evaporation  Electricity  \
30      2009-01-01      11.3      26.5      0.0          0.0          0.0
31      2009-01-02       9.6      23.9      0.0          0.0          0.0
32      2009-01-03      10.5      28.8      0.0          0.0          0.0
33      2009-01-04      12.3      34.6      0.0          0.0          0.0
34      2009-01-05      12.9      35.8      0.0          0.0          0.0
…              …         …         …        …            …            …
142188  2017-06-20       3.5      21.8      0.0          0.0          0.0
142189  2017-06-21       2.8      23.4      0.0          0.0          0.0
142190  2017-06-22       3.6      25.3      0.0          0.0          0.0
142191  2017-06-23       5.4      26.9      0.0          0.0          0.0
142192  2017-06-24       7.8      27.0      0.0          0.0          0.0

        Parameter1_Speed  Parameter3_9am  Parameter3_3pm  Parameter4_9am  …  \
30                  56.0            19.0            31.0            46.0  …
31                  41.0            19.0            11.0            44.0  …
32                  26.0            11.0             7.0            43.0  …
33                  37.0             6.0            17.0            41.0  …
34                  41.0             6.0            26.0            41.0  …
…                      …               …               …               …  …
142188              31.0            15.0            13.0            59.0  …
142189              31.0            13.0            11.0            51.0  …
142190              22.0            13.0             9.0            56.0  …
142191              37.0             9.0             9.0            53.0  …
142192              28.0            13.0             7.0            51.0  …

        Location_40  Location_41  Location_42  Location_43  Location_44  \
30                0            0            0            0            0
31                0            0            0            0            0
32                0            0            0            0            0
33                0            0            0            0            0
34                0            0            0            0            0
…                 …            …            …            …            …
142188            0            0            1            0            0
142189            0            0            1            0            0
142190            0            0            1            0            0
142191            0            0            1            0            0
142192            0            0            1            0            0

        Location_45  Location_46  Location_47  Location_48  Location_49
30                0            0            0            0            0
```

```
31                 0          0          0          0          0
32                 0          0          0          0          0
33                 0          0          0          0          0
34                 0          0          0          0          0
…             …          …          …          …          …
142188             0          0          0          0          0
142189             0          0          0          0          0
142190             0          0          0          0          0
142191             0          0          0          0          0
142192             0          0          0          0          0

[139886 rows x 80 columns]
```

[18]:
```
proporcion = (df_modelo.isna().sum().sum() / df_modelo.size)*100
c = len(df_modelo)
print(f'El df_modelo tiene un {proporcion}% de nulos con {c} columnas.')
```

```
El df_modelo tiene un 0.47566411220565313% de nulos con 139886 columnas.
```

[19]:
```
df_modelo = df_modelo.dropna()
proporcion = (df_modelo.isna().sum().sum() / df_modelo.size)*100
c = len(df_modelo)
print(f'El df_modelo tiene un {proporcion}% de nulos con {c} columnas.')
```

```
El df_modelo tiene un 0.0% de nulos con 117793 columnas.
```

1.3 2. Ejecute un modelo de probabilidad lineal (MCO) que permita explicar la probabilidad de que un dia se reporte fallo medido por sensor, a partir de las informacion disponible. Seleccione las variables dependientes a incluir en el modelo final e interprete su significado.

[20]:
```
y = df_modelo['Failure_today']
X = df_modelo.drop(['Failure_today','Date', 'Leakage', 'Month', 'Year',
 ↪'Location_17', 'Location_26', 'Location_42'], axis=1)
X=sm.add_constant(X)
model = sm.OLS(y,X)
results_ols = model.fit(cov_type='HC0')
print(results_ols.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Failure_today   R-squared:                       0.300
Model:                            OLS   Adj. R-squared:                  0.299
Method:                 Least Squares   F-statistic:                     761.0
Date:                Thu, 24 Apr 2025   Prob (F-statistic):               0.00
Time:                        23:37:07   Log-Likelihood:                -42553.
No. Observations:              117793   AIC:                         8.524e+04
Df Residuals:                  117725   BIC:                         8.590e+04
Df Model:                          67
```

```
Covariance Type:                  HC0
================================================================================
========
                         coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
--------
const                  7.7128      0.223     34.518      0.000       7.275
8.151
Min_Temp               0.0114      0.000     22.918      0.000       0.010
0.012
Max_Temp              -0.0339      0.001    -33.934      0.000      -0.036
-0.032
Evaporation           -0.0062      0.000    -14.002      0.000      -0.007
-0.005
Electricity           -0.0037      0.000     -8.120      0.000      -0.005
-0.003
Parameter1_Speed       0.0053      0.000     38.683      0.000       0.005
0.006
Parameter3_9am         0.0028      0.000     16.124      0.000       0.002
0.003
Parameter3_3pm        -0.0040      0.000    -22.130      0.000      -0.004
-0.004
Parameter4_9am         0.0068      0.000     58.807      0.000       0.007
0.007
Parameter4_3pm         0.0019      0.000     13.605      0.000       0.002
0.002
Parameter5_9am        -0.0372      0.001    -49.550      0.000      -0.039
-0.036
Parameter5_3pm         0.0293      0.001     38.928      0.000       0.028
0.031
Parameter7_9am         0.0002      0.001      0.301      0.763      -0.001
0.002
Parameter7_3pm         0.0254      0.001     22.893      0.000       0.023
0.028
I_Electricity          0.0300      0.006      4.874      0.000       0.018
0.042
I_Evaporation          0.0270      0.005      5.175      0.000       0.017
0.037
Parameter1_Dir_Norte  -0.0090      0.003     -2.694      0.007      -0.016
-0.002
Parameter1_Dir_Oeste  -0.0015      0.004     -0.385      0.700      -0.009
0.006
Parameter1_Dir_Sur     0.0032      0.004      0.898      0.369      -0.004
0.010
Parameter2_9am_Norte   0.0025      0.003      0.881      0.378      -0.003
0.008
Parameter2_9am_Oeste   0.0168      0.004      4.748      0.000       0.010
```

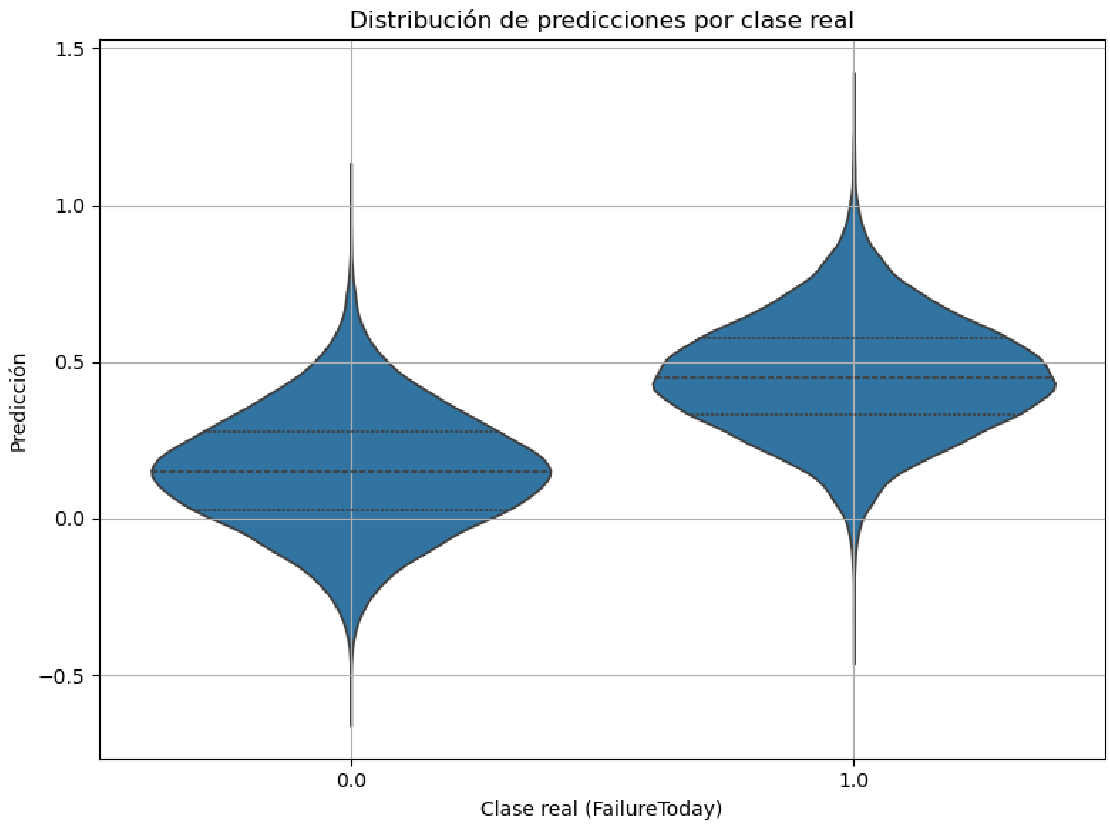
						0.024
Parameter2_9am_Sur	0.0428	0.003	12.702	0.000	0.036	0.049
Parameter2_3pm_Norte	-0.0170	0.003	-4.996	0.000	-0.024	-0.010
Parameter2_3pm_Oeste	0.0172	0.004	4.551	0.000	0.010	0.025
Parameter2_3pm_Sur	0.0126	0.003	3.599	0.000	0.006	0.019
Estacion_2	-0.0043	0.003	-1.295	0.195	-0.011	0.002
Estacion_3	0.0052	0.004	1.187	0.235	-0.003	0.014
Estacion_4	0.0520	0.003	15.432	0.000	0.045	0.059
Location_2	1.305e-14	1.16e-15	11.211	0.000	1.08e-14	1.53e-14
Location_3	-0.0244	0.007	-3.290	0.001	-0.039	-0.010
Location_4	0.1578	0.007	23.509	0.000	0.145	0.171
Location_5	-0.0618	0.008	-7.738	0.000	-0.077	-0.046
Location_6	-0.1553	0.008	-18.533	0.000	-0.172	-0.139
Location_7	-0.0665	0.007	-9.163	0.000	-0.081	-0.052
Location_8	0.0442	0.009	5.042	0.000	0.027	0.061
Location_9	0.0011	0.009	0.121	0.904	-0.017	0.020
Location_10	-0.0491	0.008	-6.200	0.000	-0.065	-0.034
Location_11	0.0147	0.007	2.048	0.041	0.001	0.029
Location_12	-0.0009	0.009	-0.099	0.921	-0.019	0.017
Location_13	-0.1102	0.009	-12.317	0.000	-0.128	-0.093
Location_14	-0.0450	0.009	-5.006	0.000	-0.063	-0.027
Location_15	-0.0266	0.009	-3.100	0.002	-0.043	-0.010
Location_16	-0.1252	0.009	-13.959	0.000	-0.143	-0.108
Location_18	-0.1131	0.010	-11.685	0.000	-0.132	-0.094
Location_19	-0.0672	0.010	-6.696	0.000	-0.087	

						-0.048
Location_20	-0.1173	0.009	-13.659	0.000	-0.134	-0.100
Location_21	-0.0483	0.007	-6.675	0.000	-0.062	-0.034
Location_22	0.0074	0.007	1.020	0.308	-0.007	0.022
Location_23	-0.0558	0.009	-6.436	0.000	-0.073	-0.039
Location_24	-1.743e-17	4.15e-18	-4.196	0.000	-2.56e-17	-9.29e-18
Location_25	9.727e-18	4.24e-18	2.292	0.022	1.41e-18	1.8e-17
Location_27	-0.1023	0.009	-11.598	0.000	-0.120	-0.085
Location_28	-0.1118	0.009	-11.986	0.000	-0.130	-0.094
Location_29	-0.0347	0.008	-4.414	0.000	-0.050	-0.019
Location_30	0.0057	0.009	0.645	0.519	-0.012	0.023
Location_31	-1.259e-19	4.54e-18	-0.028	0.978	-9.02e-18	8.77e-18
Location_32	0.0171	0.008	2.218	0.027	0.002	0.032
Location_33	0.0103	0.008	1.351	0.177	-0.005	0.025
Location_34	-0.0800	0.009	-8.684	0.000	-0.098	-0.062
Location_35	-0.0496	0.008	-6.402	0.000	-0.065	-0.034
Location_36	-0.1434	0.009	-16.825	0.000	-0.160	-0.127
Location_37	4.584e-18	3.34e-18	1.372	0.170	-1.96e-18	1.11e-17
Location_38	-0.0419	0.010	-4.327	0.000	-0.061	-0.023
Location_39	-0.0369	0.009	-4.125	0.000	-0.054	-0.019
Location_40	-0.0473	0.008	-5.683	0.000	-0.064	-0.031
Location_41	-0.0231	0.008	-3.016	0.003	-0.038	-0.008
Location_43	-0.0115	0.008	-1.494	0.135	-0.027	0.004
Location_44	-0.0534	0.009	-5.998	0.000	-0.071	-0.036
Location_45	-0.1045	0.009	-12.167	0.000	-0.121	

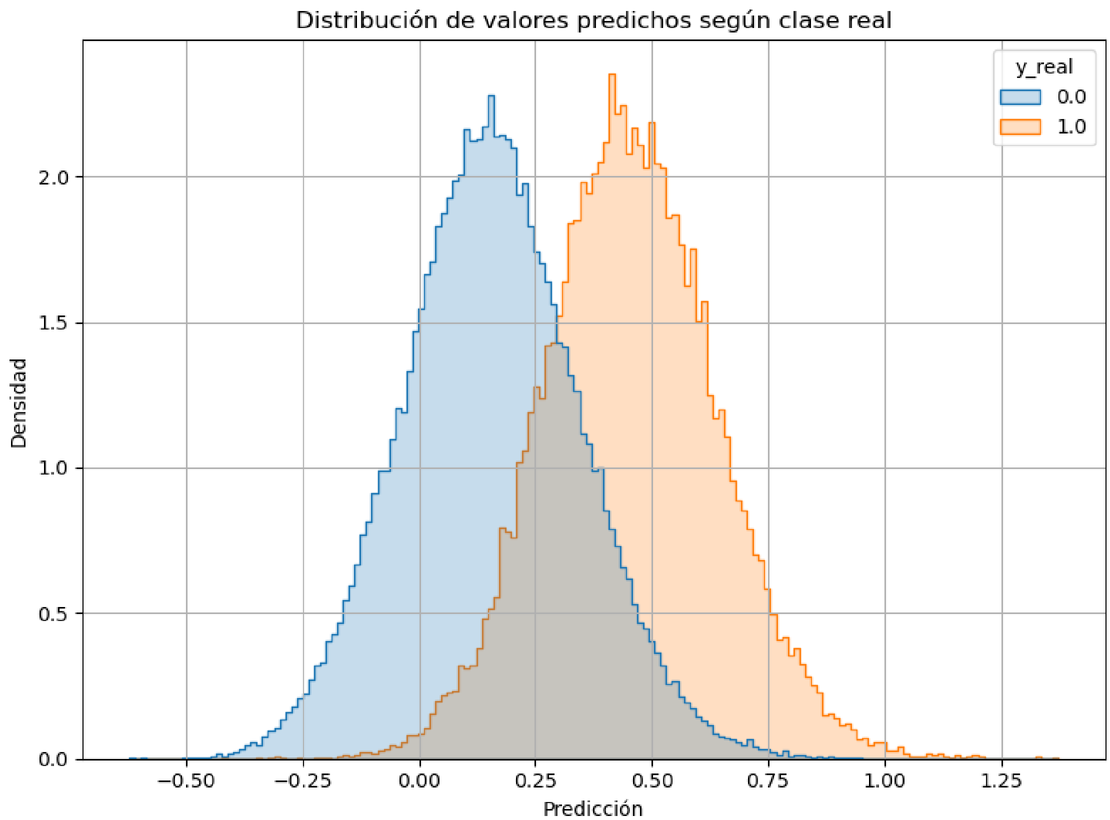
```
-0.088
Location_46              -0.0067      0.010     -0.694      0.488      -0.026
0.012
Location_47              -0.0098      0.009     -1.135      0.256      -0.027
0.007
Location_48              -0.1292      0.009    -14.812      0.000      -0.146
-0.112
Location_49              -0.0227      0.007     -3.316      0.001      -0.036
-0.009
==============================================================================
Omnibus:                     8967.498   Durbin-Watson:                   1.786
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            11165.658
Skew:                           0.752   Prob(JB):                         0.00
Kurtosis:                       2.887   Cond. No.                     7.75e+19
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The smallest eigenvalue is 4.08e-29. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

[21]:

```
df_pred_ols = pd.DataFrame({'y_real': y,
                            'y_pred': results_ols.fittedvalues})
plt.figure(figsize=(8,6))
sns.violinplot(x='y_real', y='y_pred', data=df_pred_ols, inner='quartile')
plt.title('Distribución de predicciones por clase real')
plt.xlabel('Clase real (FailureToday)')
plt.ylabel('Predicción')
plt.grid(True)
plt.tight_layout()
plt.show()
```



[22]:
```
plt.figure(figsize=(8,6))
sns.histplot(data=df_pred_ols, x='y_pred', hue='y_real', element='step',␣
↪stat='density', common_norm=False)
plt.title('Distribución de valores predichos según clase real')
plt.xlabel('Predicción')
plt.ylabel('Densidad')
plt.grid(True)
plt.tight_layout()
plt.show()
```



R: Vemos un ajuste limitado dado el R cuadrado, pero las variables incluidas son todas significativas, por lo que esas serán las variables que incluiré.

1.4 3. Ejecute un modelo *probit* para responder a la pregunta 2. Seleccione las variables dependientes a incluir en el modelo final e interprete su significado.

[23]:
```
X.corr()
```

[23]:
```
              const  Min_Temp  Max_Temp  Evaporation  Electricity  \
const           NaN       NaN       NaN          NaN          NaN
Min_Temp        NaN  1.000000  0.729620     0.371431     0.128364
Max_Temp        NaN  0.729620  1.000000     0.430861     0.276225
Evaporation     NaN  0.371431  0.430861     1.000000     0.492649
Electricity     NaN  0.128364  0.276225     0.492649     1.000000
…             …       …       …          …          …
Location_45     NaN -0.057469 -0.060288     0.045298     0.066551
Location_46     NaN  0.010224  0.015523     0.057595    -0.009404
Location_47     NaN -0.036112 -0.036627    -0.119094    -0.137098
Location_48     NaN  0.062690 -0.045696    -0.126126    -0.145193
Location_49     NaN  0.024489  0.070315     0.198933     0.055342
```

	Parameter1_Speed	Parameter3_9am	Parameter3_3pm	Parameter4_9am	\
const	NaN	NaN	NaN	NaN	
Min_Temp	0.198963	0.203872	0.175876	-0.248006	
Max_Temp	0.091001	0.033257	0.044063	-0.526967	
Evaporation	0.137571	0.162519	0.091016	-0.388722	
Electricity	0.027632	0.081771	0.069427	-0.307824	
…	…	…	…	…	
Location_45	-0.023408	-0.080644	-0.070903	0.083828	
Location_46	0.017938	0.033137	0.055322	0.019851	
Location_47	0.000195	-0.005202	0.003837	0.022429	
Location_48	0.064948	0.039718	0.051529	-0.006798	
Location_49	0.046735	0.103253	0.029875	-0.126754	

	Parameter4_3pm	…	Location_39	Location_40	Location_41	\
const	NaN	…	NaN	NaN	NaN	
Min_Temp	0.020072	…	0.062633	0.201145	-0.128682	
Max_Temp	-0.502128	…	-0.002336	0.134392	-0.061372	
Evaporation	-0.306175	…	0.066250	0.134354	-0.127395	
Electricity	-0.336867	…	0.093368	0.098900	-0.146654	
…	…	…	…	…	…	
Location_45	0.029010	…	-0.025607	-0.025763	-0.025414	
Location_46	0.015249	…	-0.022340	-0.022476	-0.022172	
Location_47	0.049577	…	-0.023725	-0.023869	-0.023546	
Location_48	0.106951	…	-0.025126	-0.025279	-0.024936	
Location_49	-0.173698	…	-0.025450	-0.025605	-0.025259	

	Location_43	Location_44	Location_45	Location_46	Location_47	\
const	NaN	NaN	NaN	NaN	NaN	
Min_Temp	-0.070878	-0.010710	-0.057469	0.010224	-0.036112	
Max_Temp	-0.012150	-0.063434	-0.060288	0.015523	-0.036627	
Evaporation	0.058642	-0.123343	0.045298	0.057595	-0.119094	
Electricity	0.092242	-0.141989	0.066551	-0.009404	-0.137098	
…	…	…	…	…	…	
Location_45	-0.025562	-0.024606	1.000000	-0.022371	-0.023758	
Location_46	-0.022301	-0.021467	-0.022371	1.000000	-0.020727	
Location_47	-0.023683	-0.022797	-0.023758	-0.020727	1.000000	
Location_48	-0.025082	-0.024143	-0.025161	-0.021951	-0.023312	
Location_49	-0.025406	-0.024455	-0.025486	-0.022235	-0.023613	

	Location_48	Location_49
const	NaN	NaN
Min_Temp	0.062690	0.024489
Max_Temp	-0.045696	0.070315
Evaporation	-0.126126	0.198933
Electricity	-0.145193	0.055342
…	…	…

```
Location_45    -0.025161    -0.025486
Location_46    -0.021951    -0.022235
Location_47    -0.023312    -0.023613
Location_48     1.000000    -0.025007
Location_49    -0.025007     1.000000

[73 rows x 73 columns]
```

[24]:
```python
X[['Location_2', 'Location_24', 'Location_25', 'Location_31', 'Location_37']]
```

[24]:
```
        Location_2  Location_24  Location_25  Location_31  Location_37
30               0            0            0            0            0
31               0            0            0            0            0
32               0            0            0            0            0
33               0            0            0            0            0
34               0            0            0            0            0
…              …           …           …           …           …
142188           0            0            0            0            0
142189           0            0            0            0            0
142190           0            0            0            0            0
142191           0            0            0            0            0
142192           0            0            0            0            0

[117793 rows x 5 columns]
```

1.4.1 Nos damos cuenta que estas locaciones generan error al estar llenas de "0", por lo que las descarto.

[25]:
```python
X1 = X.drop(['Location_2', 'Location_24', 'Location_25', 'Location_31',␣
 ↪'Location_37'], axis=1)
model = sm.Probit(y, X1)
probit_model = model.fit(cov_type='HC0')
print(probit_model.summary())

mfxp = probit_model.get_margeff()
print(mfxp.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.348503
         Iterations 7
                          Probit Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:               117793
Model:                         Probit   Df Residuals:                   117725
Method:                           MLE   Df Model:                           67
Date:                Thu, 24 Apr 2025   Pseudo R-squ.:                  0.3402
Time:                        23:37:20   Log-Likelihood:                -41051.
converged:                       True   LL-Null:                       -62216.
```

```
Covariance Type:                  HC0   LLR p-value:                     0.000
================================================================================
========
                          coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
--------
const                  26.8802      0.987     27.228      0.000      24.945
28.815
Min_Temp                0.0934      0.003     30.953      0.000       0.087
0.099
Max_Temp               -0.1482      0.005    -27.802      0.000      -0.159
-0.138
Evaporation            -0.0464      0.004    -11.128      0.000      -0.055
-0.038
Electricity             0.0082      0.002      3.654      0.000       0.004
0.013
Parameter1_Speed        0.0210      0.001     33.682      0.000       0.020
0.022
Parameter3_9am          0.0088      0.001     10.276      0.000       0.007
0.010
Parameter3_3pm         -0.0148      0.001    -16.738      0.000      -0.017
-0.013
Parameter4_9am          0.0388      0.001     60.780      0.000       0.038
0.040
Parameter4_3pm          0.0013      0.001      2.121      0.034    9.93e-05
0.003
Parameter5_9am         -0.1340      0.004    -38.147      0.000      -0.141
-0.127
Parameter5_3pm          0.1045      0.003     30.056      0.000       0.098
0.111
Parameter7_9am         -0.0048      0.005     -1.061      0.289      -0.014
0.004
Parameter7_3pm          0.0555      0.006      9.491      0.000       0.044
0.067
I_Electricity          -0.1158      0.028     -4.081      0.000      -0.171
-0.060
I_Evaporation           0.2681      0.029      9.102      0.000       0.210
0.326
Parameter1_Dir_Norte   -0.0689      0.019     -3.603      0.000      -0.106
-0.031
Parameter1_Dir_Oeste   -0.0160      0.020     -0.808      0.419      -0.055
0.023
Parameter1_Dir_Sur      0.0144      0.017      0.841      0.401      -0.019
0.048
Parameter2_9am_Norte   -0.0275      0.017     -1.633      0.103      -0.061
0.006
Parameter2_9am_Oeste    0.0711      0.017      4.228      0.000       0.038
```

						0.104
Parameter2_9am_Sur	0.1747	0.015	11.420	0.000	0.145	0.205
Parameter2_3pm_Norte	-0.0636	0.019	-3.344	0.001	-0.101	-0.026
Parameter2_3pm_Oeste	0.0585	0.020	2.965	0.003	0.020	0.097
Parameter2_3pm_Sur	0.0190	0.017	1.108	0.268	-0.015	0.053
Estacion_2	-0.0963	0.017	-5.772	0.000	-0.129	-0.064
Estacion_3	-0.1665	0.022	-7.457	0.000	-0.210	-0.123
Estacion_4	0.1492	0.018	8.208	0.000	0.114	0.185
Location_3	0.0432	0.042	1.030	0.303	-0.039	0.125
Location_4	0.5972	0.060	9.949	0.000	0.480	0.715
Location_5	-0.0115	0.041	-0.277	0.781	-0.093	0.070
Location_6	-0.6714	0.043	-15.750	0.000	-0.755	-0.588
Location_7	-0.2300	0.042	-5.436	0.000	-0.313	-0.147
Location_8	0.5841	0.043	13.592	0.000	0.500	0.668
Location_9	0.3694	0.043	8.674	0.000	0.286	0.453
Location_10	0.0105	0.043	0.244	0.807	-0.074	0.095
Location_11	0.0684	0.051	1.336	0.181	-0.032	0.169
Location_12	0.2647	0.041	6.400	0.000	0.184	0.346
Location_13	-0.4354	0.041	-10.638	0.000	-0.516	-0.355
Location_14	0.1690	0.045	3.752	0.000	0.081	0.257
Location_15	0.1937	0.042	4.609	0.000	0.111	0.276
Location_16	-0.3255	0.044	-7.437	0.000	-0.411	-0.240
Location_18	-0.3088	0.046	-6.760	0.000	-0.398	-0.219
Location_19	-0.0493	0.046	-1.072	0.284	-0.139	0.041
Location_20	-0.3176	0.043	-7.372	0.000	-0.402	

```
-0.233
Location_21          -0.3389      0.049     -6.923      0.000      -0.435
-0.243
Location_22           0.3399      0.048      7.022      0.000       0.245
0.435
Location_23          -0.1623      0.041     -4.005      0.000      -0.242
-0.083
Location_27          -0.2731      0.041     -6.710      0.000      -0.353
-0.193
Location_28          -0.2730      0.041     -6.693      0.000      -0.353
-0.193
Location_29          -0.2207      0.046     -4.756      0.000      -0.312
-0.130
Location_30           0.3412      0.051      6.672      0.000       0.241
0.441
Location_32           0.2362      0.043      5.541      0.000       0.153
0.320
Location_33           0.2691      0.044      6.094      0.000       0.183
0.356
Location_34          -0.2947      0.040     -7.373      0.000      -0.373
-0.216
Location_35          -0.0293      0.042     -0.699      0.484      -0.111
0.053
Location_36          -0.3894      0.042     -9.326      0.000      -0.471
-0.308
Location_38           0.0832      0.044      1.895      0.058      -0.003
0.169
Location_39           0.0770      0.044      1.761      0.078      -0.009
0.163
Location_40           0.1855      0.046      4.073      0.000       0.096
0.275
Location_41           0.1294      0.041      3.144      0.002       0.049
0.210
Location_43           0.0573      0.047      1.226      0.220      -0.034
0.149
Location_44          -0.1134      0.039     -2.881      0.004      -0.191
-0.036
Location_45          -0.3383      0.042     -7.992      0.000      -0.421
-0.255
Location_46           0.2841      0.045      6.365      0.000       0.197
0.372
Location_47           0.1390      0.041      3.406      0.001       0.059
0.219
Location_48          -0.3704      0.042     -8.778      0.000      -0.453
-0.288
Location_49          -0.3410      0.058     -5.857      0.000      -0.455
-0.227
==============================================================================
```

```
========
       Probit Marginal Effects
=====================================
Dep. Variable:          Failure_today
Method:                          dydx
At:                           overall
================================================================================
========
                          dy/dx    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
--------
Min_Temp                 0.0183      0.001     31.509      0.000       0.017
0.019
Max_Temp                -0.0290      0.001    -28.127      0.000      -0.031
-0.027
Evaporation             -0.0091      0.001    -11.233      0.000      -0.011
-0.007
Electricity              0.0016      0.000      3.658      0.000       0.001
0.002
Parameter1_Speed         0.0041      0.000     34.395      0.000       0.004
0.004
Parameter3_9am           0.0017      0.000     10.293      0.000       0.001
0.002
Parameter3_3pm          -0.0029      0.000    -16.812      0.000      -0.003
-0.003
Parameter4_9am           0.0076      0.000     64.797      0.000       0.007
0.008
Parameter4_3pm           0.0003      0.000      2.121      0.034    1.94e-05
0.000
Parameter5_9am          -0.0262      0.001    -39.037      0.000      -0.028
-0.025
Parameter5_3pm           0.0205      0.001     30.497      0.000       0.019
0.022
Parameter7_9am          -0.0009      0.001     -1.061      0.289      -0.003
0.001
Parameter7_3pm           0.0109      0.001      9.503      0.000       0.009
0.013
I_Electricity           -0.0227      0.006     -4.085      0.000      -0.034
-0.012
I_Evaporation            0.0525      0.006      9.144      0.000       0.041
0.064
Parameter1_Dir_Norte    -0.0135      0.004     -3.604      0.000      -0.021
-0.006
Parameter1_Dir_Oeste    -0.0031      0.004     -0.808      0.419      -0.011
0.004
Parameter1_Dir_Sur       0.0028      0.003      0.841      0.401      -0.004
0.009
```

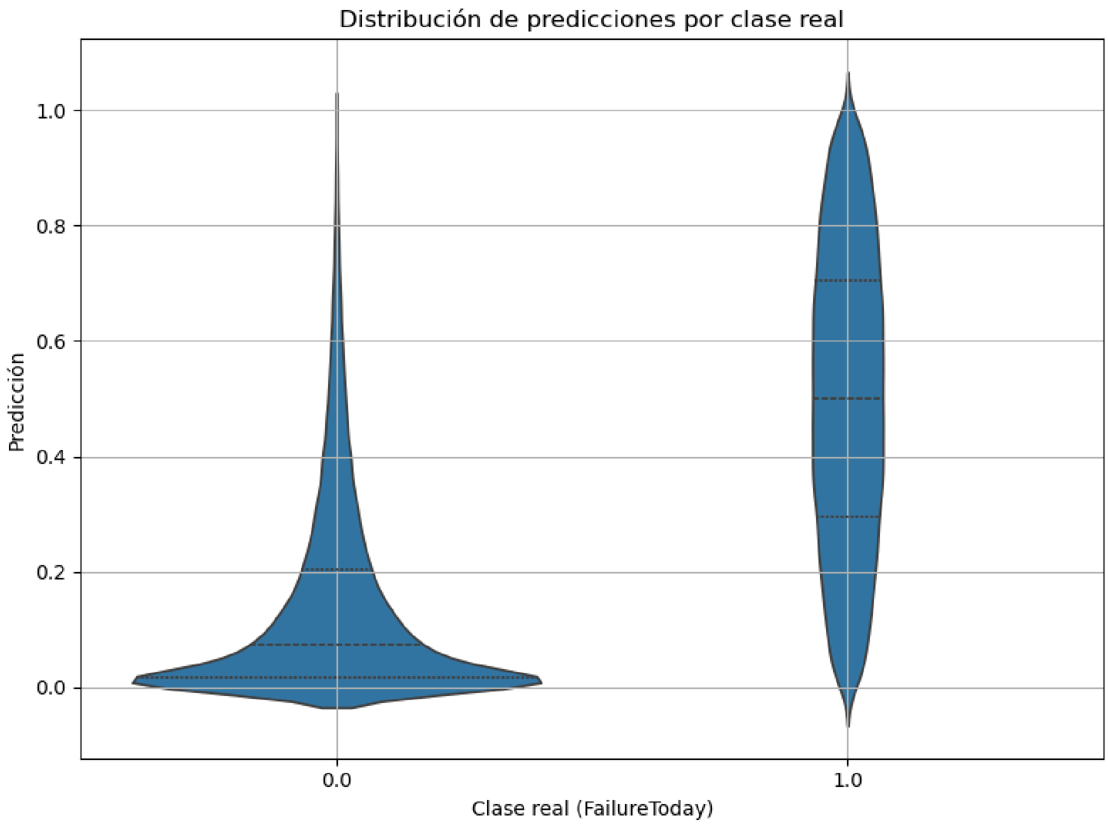
Parameter2_9am_Norte	-0.0054	0.003	-1.633	0.103	-0.012	0.001
Parameter2_9am_Oeste	0.0139	0.003	4.228	0.000	0.007	0.020
Parameter2_9am_Sur	0.0342	0.003	11.434	0.000	0.028	0.040
Parameter2_3pm_Norte	-0.0125	0.004	-3.344	0.001	-0.020	-0.005
Parameter2_3pm_Oeste	0.0114	0.004	2.966	0.003	0.004	0.019
Parameter2_3pm_Sur	0.0037	0.003	1.108	0.268	-0.003	0.010
Estacion_2	-0.0189	0.003	-5.776	0.000	-0.025	-0.012
Estacion_3	-0.0326	0.004	-7.472	0.000	-0.041	-0.024
Estacion_4	0.0292	0.004	8.210	0.000	0.022	0.036
Location_3	0.0085	0.008	1.030	0.303	-0.008	0.025
Location_4	0.1169	0.012	9.971	0.000	0.094	0.140
Location_5	-0.0023	0.008	-0.277	0.781	-0.018	0.014
Location_6	-0.1314	0.008	-15.858	0.000	-0.148	-0.115
Location_7	-0.0450	0.008	-5.441	0.000	-0.061	-0.029
Location_8	0.1143	0.008	13.655	0.000	0.098	0.131
Location_9	0.0723	0.008	8.700	0.000	0.056	0.089
Location_10	0.0021	0.008	0.244	0.807	-0.014	0.019
Location_11	0.0134	0.010	1.336	0.181	-0.006	0.033
Location_12	0.0518	0.008	6.406	0.000	0.036	0.068
Location_13	-0.0852	0.008	-10.667	0.000	-0.101	-0.070
Location_14	0.0331	0.009	3.756	0.000	0.016	0.050
Location_15	0.0379	0.008	4.612	0.000	0.022	0.054
Location_16	-0.0637	0.009	-7.453	0.000	-0.080	-0.047
Location_18	-0.0604	0.009	-6.766	0.000	-0.078	-0.043

Location_19	-0.0097	0.009	-1.072	0.284	-0.027	0.008
Location_20	-0.0622	0.008	-7.382	0.000	-0.079	-0.046
Location_21	-0.0663	0.010	-6.930	0.000	-0.085	-0.048
Location_22	0.0665	0.009	7.022	0.000	0.048	0.085
Location_23	-0.0318	0.008	-4.007	0.000	-0.047	-0.016
Location_27	-0.0535	0.008	-6.717	0.000	-0.069	-0.038
Location_28	-0.0534	0.008	-6.697	0.000	-0.069	-0.038
Location_29	-0.0432	0.009	-4.760	0.000	-0.061	-0.025
Location_30	0.0668	0.010	6.676	0.000	0.047	0.086
Location_32	0.0462	0.008	5.547	0.000	0.030	0.063
Location_33	0.0527	0.009	6.100	0.000	0.036	0.070
Location_34	-0.0577	0.008	-7.384	0.000	-0.073	-0.042
Location_35	-0.0057	0.008	-0.699	0.484	-0.022	0.010
Location_36	-0.0762	0.008	-9.347	0.000	-0.092	-0.060
Location_38	0.0163	0.009	1.896	0.058	-0.001	0.033
Location_39	0.0151	0.009	1.761	0.078	-0.002	0.032
Location_40	0.0363	0.009	4.080	0.000	0.019	0.054
Location_41	0.0253	0.008	3.144	0.002	0.010	0.041
Location_43	0.0112	0.009	1.226	0.220	-0.007	0.029
Location_44	-0.0222	0.008	-2.882	0.004	-0.037	-0.007
Location_45	-0.0662	0.008	-8.005	0.000	-0.082	-0.050
Location_46	0.0556	0.009	6.371	0.000	0.039	0.073
Location_47	0.0272	0.008	3.406	0.001	0.012	0.043
Location_48	-0.0725	0.008	-8.792	0.000	-0.089	-0.056

```
Location_49               -0.0668      0.011     -5.863      0.000      -0.089
-0.044
================================================================================
========
```

[26]:

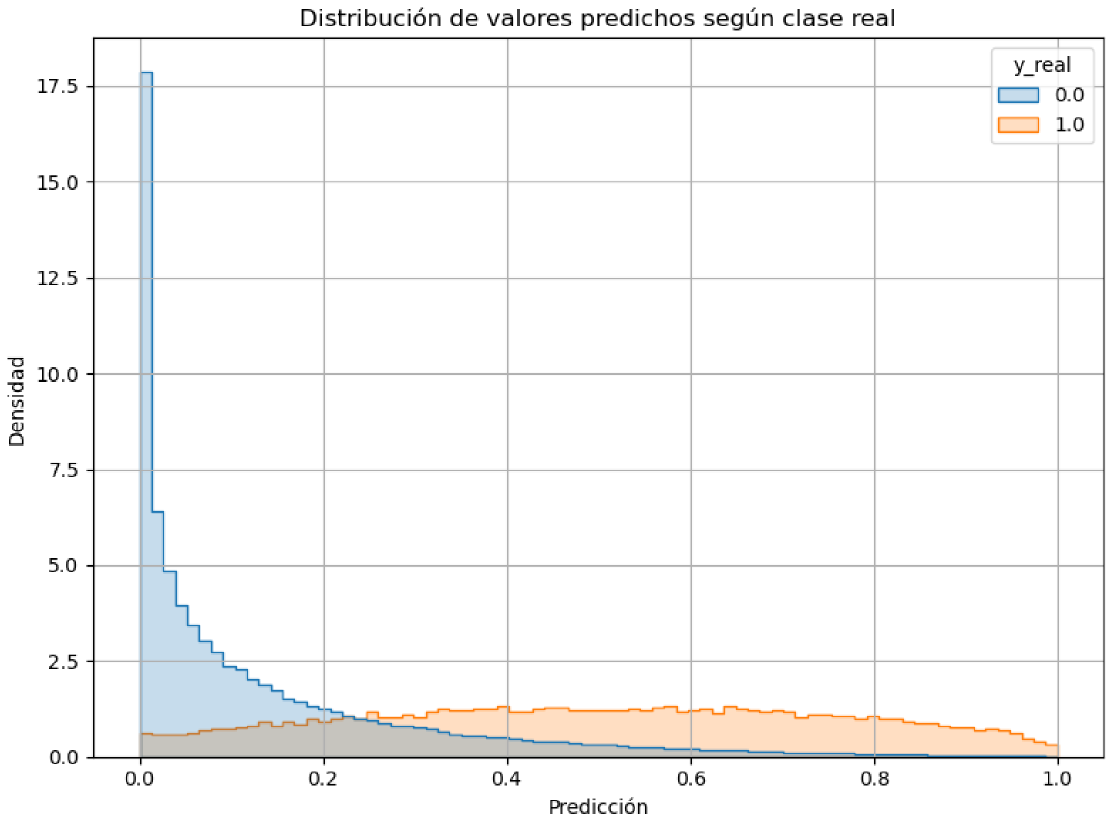
```
df_pred_probit = pd.DataFrame({'y_real': y,
                            'y_pred': probit_model.predict(X1)})
plt.figure(figsize=(8,6))
sns.violinplot(x='y_real', y='y_pred', data=df_pred_probit, inner='quartile')
plt.title('Distribución de predicciones por clase real')
plt.xlabel('Clase real (FailureToday)')
plt.ylabel('Predicción')
plt.grid(True)
plt.tight_layout()
plt.show()
```



[27]:

```
plt.figure(figsize=(8,6))
sns.histplot(data=df_pred_probit, x='y_pred', hue='y_real', element='step',␣
 ↪stat='density', common_norm=False)
plt.title('Distribución de valores predichos según clase real')
plt.xlabel('Predicción')
```

```
plt.ylabel('Densidad')
plt.grid(True)
plt.tight_layout()
plt.show()
```



R: Vemos que Probit se ajusta mejor a los datos en comparación a OLS, por otro lado, vemos que las variables explicativas siguen siendo robustas, a excepción de 'Parameter4_3pm' que pierde significancia.

1.5 4. Ejecute un modelo *logit* para responder a la pregunta 2. Seleccione las variables dependientes a incluir en el modelo final e interprete su significado.

[28]:
```
model = sm.Logit(y, X1)
logit_model = model.fit(cov_type='HC0')
print(logit_model.summary())

mfxl = logit_model.get_margeff()
print(mfxl.summary())

params = logit_model.params
```

```
conf = logit_model.conf_int()
conf['Odds Ratio'] = params
conf.columns = ['Odds Ratio', '5%', '95%']
print("Odds Ratios")
print(np.exp(conf).iloc[1:17 , ])
```

```
Optimization terminated successfully.
         Current function value: 0.346766
         Iterations 8
                           Logit Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:               117793
Model:                          Logit   Df Residuals:                   117725
Method:                           MLE   Df Model:                           67
Date:                Thu, 24 Apr 2025   Pseudo R-squ.:                  0.3435
Time:                        23:38:20   Log-Likelihood:                -40847.
converged:                       True   LL-Null:                       -62216.
Covariance Type:                  HC0   LLR p-value:                     0.000
================================================================================
========
                           coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
--------
const                   46.6643      1.742     26.788      0.000      43.250
50.078
Min_Temp                 0.1736      0.005     32.593      0.000       0.163
0.184
Max_Temp                -0.2680      0.010    -28.175      0.000      -0.287
-0.249
Evaporation             -0.1080      0.007    -15.178      0.000      -0.122
-0.094
Electricity              0.0225      0.004      5.696      0.000       0.015
0.030
Parameter1_Speed         0.0375      0.001     33.757      0.000       0.035
0.040
Parameter3_9am           0.0147      0.002      9.691      0.000       0.012
0.018
Parameter3_3pm          -0.0255      0.002    -16.154      0.000      -0.029
-0.022
Parameter4_9am           0.0704      0.001     62.147      0.000       0.068
0.073
Parameter4_3pm           0.0016      0.001      1.474      0.141      -0.001
0.004
Parameter5_9am          -0.2378      0.006    -37.970      0.000      -0.250
-0.226
Parameter5_3pm           0.1867      0.006     30.095      0.000       0.175
0.199
```

Parameter7_9am	-0.0114	0.008	-1.420	0.156	-0.027	0.004
Parameter7_3pm	0.0938	0.010	9.053	0.000	0.074	0.114
I_Electricity	-0.2544	0.050	-5.111	0.000	-0.352	-0.157
I_Evaporation	0.5689	0.050	11.406	0.000	0.471	0.667
Parameter1_Dir_Norte	-0.1274	0.034	-3.765	0.000	-0.194	-0.061
Parameter1_Dir_Oeste	-0.0461	0.035	-1.314	0.189	-0.115	0.023
Parameter1_Dir_Sur	0.0112	0.031	0.366	0.714	-0.049	0.071
Parameter2_9am_Norte	-0.0552	0.030	-1.841	0.066	-0.114	0.004
Parameter2_9am_Oeste	0.1114	0.030	3.739	0.000	0.053	0.170
Parameter2_9am_Sur	0.3021	0.027	11.170	0.000	0.249	0.355
Parameter2_3pm_Norte	-0.1017	0.034	-3.021	0.003	-0.168	-0.036
Parameter2_3pm_Oeste	0.1118	0.035	3.197	0.001	0.043	0.180
Parameter2_3pm_Sur	0.0324	0.031	1.057	0.290	-0.028	0.093
Estacion_2	-0.1690	0.030	-5.719	0.000	-0.227	-0.111
Estacion_3	-0.3243	0.039	-8.263	0.000	-0.401	-0.247
Estacion_4	0.2510	0.032	7.724	0.000	0.187	0.315
Location_3	0.0549	0.074	0.740	0.459	-0.090	0.200
Location_4	1.0558	0.108	9.766	0.000	0.844	1.268
Location_5	0.0225	0.075	0.302	0.762	-0.124	0.169
Location_6	-1.2512	0.075	-16.774	0.000	-1.397	-1.105
Location_7	-0.4346	0.075	-5.783	0.000	-0.582	-0.287
Location_8	1.1373	0.077	14.836	0.000	0.987	1.288
Location_9	0.7848	0.075	10.401	0.000	0.637	0.933
Location_10	0.0203	0.078	0.261	0.794	-0.132	0.173

Location_11	0.0577	0.093	0.623	0.533	-0.124	0.239
Location_12	0.5240	0.073	7.130	0.000	0.380	0.668
Location_13	-0.7949	0.072	-10.990	0.000	-0.937	-0.653
Location_14	0.4361	0.081	5.417	0.000	0.278	0.594
Location_15	0.4016	0.075	5.360	0.000	0.255	0.548
Location_16	-0.6201	0.079	-7.856	0.000	-0.775	-0.465
Location_18	-0.5364	0.081	-6.642	0.000	-0.695	-0.378
Location_19	-0.0837	0.082	-1.020	0.308	-0.244	0.077
Location_20	-0.5624	0.077	-7.303	0.000	-0.713	-0.411
Location_21	-0.6294	0.087	-7.210	0.000	-0.801	-0.458
Location_22	0.5824	0.089	6.554	0.000	0.408	0.757
Location_23	-0.3030	0.072	-4.222	0.000	-0.444	-0.162
Location_27	-0.4861	0.073	-6.628	0.000	-0.630	-0.342
Location_28	-0.4593	0.073	-6.306	0.000	-0.602	-0.317
Location_29	-0.4393	0.082	-5.326	0.000	-0.601	-0.278
Location_30	0.6187	0.091	6.826	0.000	0.441	0.796
Location_32	0.4658	0.075	6.184	0.000	0.318	0.614
Location_33	0.5163	0.078	6.586	0.000	0.363	0.670
Location_34	-0.5378	0.071	-7.587	0.000	-0.677	-0.399
Location_35	-0.0123	0.076	-0.162	0.871	-0.161	0.136
Location_36	-0.6951	0.075	-9.267	0.000	-0.842	-0.548
Location_38	0.2162	0.078	2.762	0.006	0.063	0.370
Location_39	0.1785	0.080	2.244	0.025	0.023	0.334
Location_40	0.5042	0.081	6.196	0.000	0.345	0.664

```
Location_41              0.2399      0.073      3.269      0.001       0.096
0.384
Location_43              0.0247      0.083      0.298      0.766      -0.138
0.187
Location_44             -0.2015      0.070     -2.870      0.004      -0.339
-0.064
Location_45             -0.6116      0.075     -8.144      0.000      -0.759
-0.464
Location_46              0.5397      0.080      6.761      0.000       0.383
0.696
Location_47              0.2569      0.072      3.577      0.000       0.116
0.398
Location_48             -0.6597      0.077     -8.609      0.000      -0.810
-0.510
Location_49             -0.6590      0.102     -6.444      0.000      -0.859
-0.459
================================================================================
========
        Logit Marginal Effects
=====================================
Dep. Variable:          Failure_today
Method:                          dydx
At:                           overall
================================================================================
========
                          dy/dx    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
--------
Min_Temp                 0.0190      0.001     33.135      0.000       0.018
0.020
Max_Temp                -0.0293      0.001    -28.563      0.000      -0.031
-0.027
Evaporation             -0.0118      0.001    -15.363      0.000      -0.013
-0.010
Electricity              0.0025      0.000      5.703      0.000       0.002
0.003
Parameter1_Speed         0.0041      0.000     34.633      0.000       0.004
0.004
Parameter3_9am           0.0016      0.000      9.706      0.000       0.001
0.002
Parameter3_3pm          -0.0028      0.000    -16.231      0.000      -0.003
-0.002
Parameter4_9am           0.0077      0.000     66.555      0.000       0.007
0.008
Parameter4_3pm           0.0002      0.000      1.474      0.140   -5.77e-05
0.000
Parameter5_9am          -0.0260      0.001    -39.028      0.000      -0.027
```

```
-0.025
Parameter5_3pm           0.0204      0.001     30.637      0.000       0.019
0.022
Parameter7_9am          -0.0013      0.001     -1.420      0.156      -0.003
0.000
Parameter7_3pm           0.0103      0.001      9.073      0.000       0.008
0.012
I_Electricity           -0.0278      0.005     -5.113      0.000      -0.039
-0.017
I_Evaporation            0.0622      0.005     11.456      0.000       0.052
0.073
Parameter1_Dir_Norte    -0.0139      0.004     -3.765      0.000      -0.021
-0.007
Parameter1_Dir_Oeste    -0.0050      0.004     -1.314      0.189      -0.013
0.002
Parameter1_Dir_Sur       0.0012      0.003      0.366      0.714      -0.005
0.008
Parameter2_9am_Norte    -0.0060      0.003     -1.842      0.066      -0.012
0.000
Parameter2_9am_Oeste     0.0122      0.003      3.739      0.000       0.006
0.019
Parameter2_9am_Sur       0.0331      0.003     11.187      0.000       0.027
0.039
Parameter2_3pm_Norte    -0.0111      0.004     -3.021      0.003      -0.018
-0.004
Parameter2_3pm_Oeste     0.0122      0.004      3.198      0.001       0.005
0.020
Parameter2_3pm_Sur       0.0035      0.003      1.057      0.290      -0.003
0.010
Estacion_2              -0.0185      0.003     -5.718      0.000      -0.025
-0.012
Estacion_3              -0.0355      0.004     -8.271      0.000      -0.044
-0.027
Estacion_4               0.0275      0.004      7.733      0.000       0.021
0.034
Location_3               0.0060      0.008      0.740      0.459      -0.010
0.022
Location_4               0.1155      0.012      9.787      0.000       0.092
0.139
Location_5               0.0025      0.008      0.302      0.762      -0.014
0.018
Location_6              -0.1369      0.008    -16.894      0.000      -0.153
-0.121
Location_7              -0.0476      0.008     -5.787      0.000      -0.064
-0.031
Location_8               0.1244      0.008     14.905      0.000       0.108
0.141
Location_9               0.0859      0.008     10.434      0.000       0.070
```

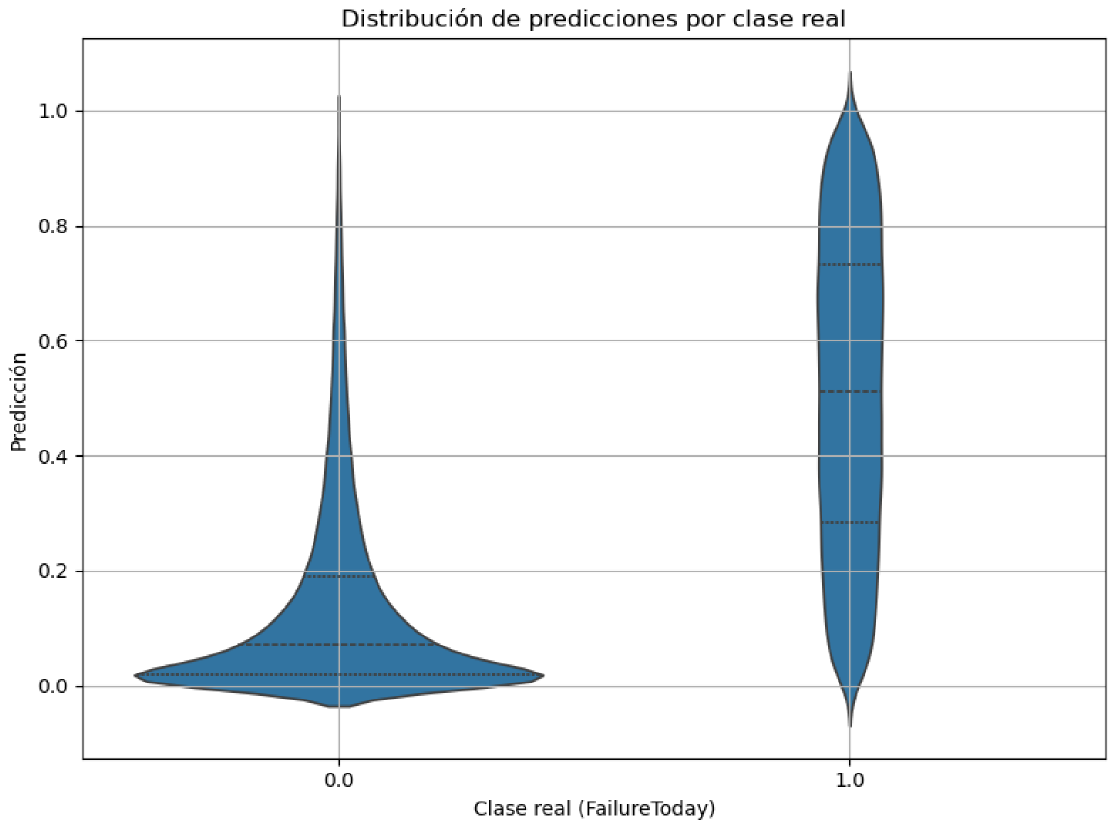
						0.102
Location_10	0.0022	0.009	0.261	0.794	-0.014	0.019
Location_11	0.0063	0.010	0.623	0.533	-0.014	0.026
Location_12	0.0573	0.008	7.138	0.000	0.042	0.073
Location_13	-0.0870	0.008	-11.020	0.000	-0.102	-0.072
Location_14	0.0477	0.009	5.424	0.000	0.030	0.065
Location_15	0.0439	0.008	5.363	0.000	0.028	0.060
Location_16	-0.0679	0.009	-7.876	0.000	-0.085	-0.051
Location_18	-0.0587	0.009	-6.649	0.000	-0.076	-0.041
Location_19	-0.0092	0.009	-1.020	0.308	-0.027	0.008
Location_20	-0.0615	0.008	-7.315	0.000	-0.078	-0.045
Location_21	-0.0689	0.010	-7.218	0.000	-0.088	-0.050
Location_22	0.0637	0.010	6.555	0.000	0.045	0.083
Location_23	-0.0332	0.008	-4.224	0.000	-0.049	-0.018
Location_27	-0.0532	0.008	-6.636	0.000	-0.069	-0.037
Location_28	-0.0503	0.008	-6.311	0.000	-0.066	-0.035
Location_29	-0.0481	0.009	-5.330	0.000	-0.066	-0.030
Location_30	0.0677	0.010	6.828	0.000	0.048	0.087
Location_32	0.0510	0.008	6.191	0.000	0.035	0.067
Location_33	0.0565	0.009	6.593	0.000	0.040	0.073
Location_34	-0.0588	0.008	-7.597	0.000	-0.074	-0.044
Location_35	-0.0013	0.008	-0.162	0.871	-0.018	0.015
Location_36	-0.0761	0.008	-9.290	0.000	-0.092	-0.060
Location_38	0.0237	0.009	2.762	0.006	0.007	0.040
Location_39	0.0195	0.009	2.244	0.025	0.002	

```
0.037
Location_40              0.0552      0.009      6.210      0.000       0.038
0.073
Location_41              0.0262      0.008      3.269      0.001       0.011
0.042
Location_43              0.0027      0.009      0.298      0.766      -0.015
0.021
Location_44             -0.0221      0.008     -2.871      0.004      -0.037
-0.007
Location_45             -0.0669      0.008     -8.158      0.000      -0.083
-0.051
Location_46              0.0591      0.009      6.764      0.000       0.042
0.076
Location_47              0.0281      0.008      3.577      0.000       0.013
0.044
Location_48             -0.0722      0.008     -8.626      0.000      -0.089
-0.056
Location_49             -0.0721      0.011     -6.449      0.000      -0.094
-0.050
================================================================================
========
Odds Ratios
                      Odds Ratio        5%       95%
Min_Temp                1.177181  1.202012  1.189531
Max_Temp                0.750745  0.779270  0.764875
Evaporation             0.885229  0.910260  0.897657
Electricity             1.014837  1.030641  1.022709
Parameter1_Speed        1.035919  1.040435  1.038174
Parameter3_9am          1.011794  1.017827  1.014806
Parameter3_3pm          0.971776  0.977815  0.974791
Parameter4_9am          1.070594  1.075361  1.072975
Parameter4_3pm          0.999473  1.003734  1.001601
Parameter5_9am          0.778703  0.798061  0.788323
Parameter5_3pm          1.190652  1.219954  1.205214
Parameter7_9am          0.973163  1.004355  0.988636
Parameter7_3pm          1.076294  1.120930  1.098385
I_Electricity           0.703305  0.854839  0.775379
I_Evaporation           1.601756  1.947609  1.766238
Parameter1_Dir_Norte    0.823907  0.940760  0.880397
```

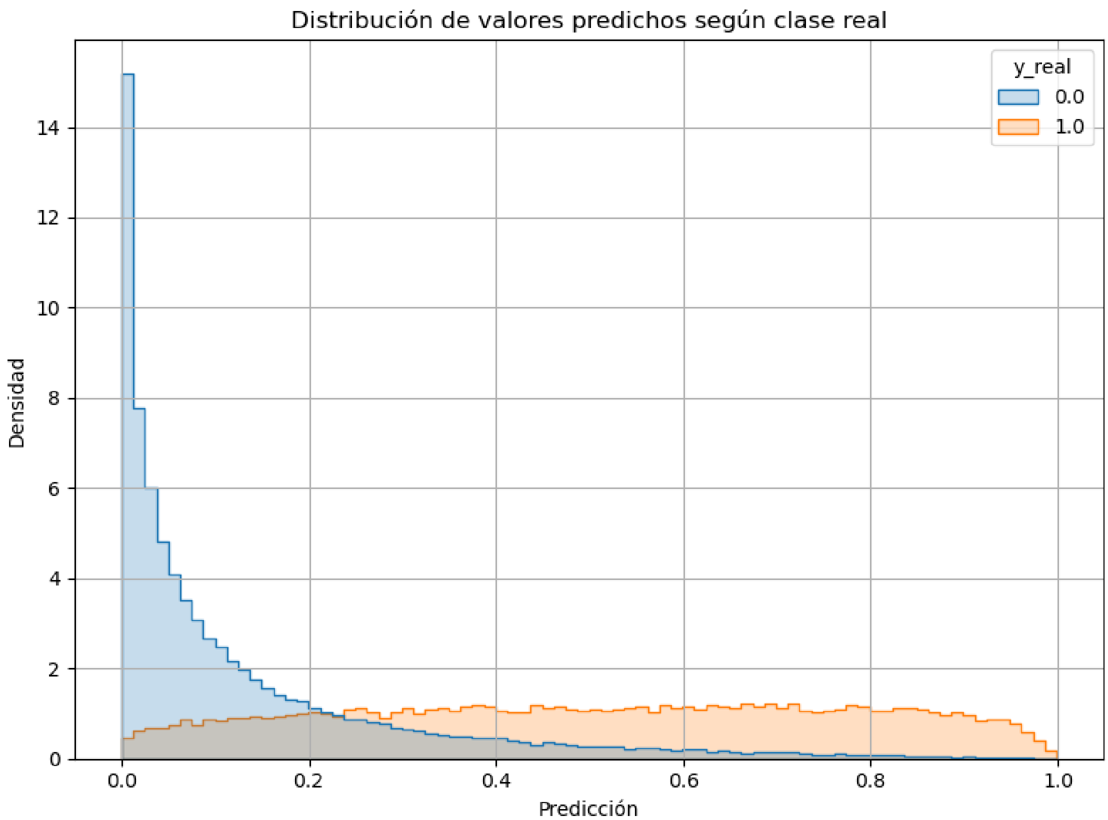
[29]:

```
df_pred_logit = pd.DataFrame({'y_real': y,
                            'y_pred': logit_model.predict(X1)})
plt.figure(figsize=(8,6))
sns.violinplot(x='y_real', y='y_pred', data=df_pred_logit, inner='quartile')
plt.title('Distribución de predicciones por clase real')
plt.xlabel('Clase real (FailureToday)')
plt.ylabel('Predicción')
```

```
plt.grid(True)
plt.tight_layout()
plt.show()
```



[30]:
```
plt.figure(figsize=(8,6))
sns.histplot(data=df_pred_logit, x='y_pred', hue='y_real', element='step',␣
↪stat='density', common_norm=False)
plt.title('Distribución de valores predichos según clase real')
plt.xlabel('Predicción')
plt.ylabel('Densidad')
plt.grid(True)
plt.tight_layout()
plt.show()
```



R: El ajuste de Logit sigue siendo muy similar a Probit, y las variables entre ellas tienen la misma robustez.

1.6 5. Comente los resultados obtenidos en 2, 3 y 4. ¿Cuáles y por qué existen las diferencias entre los resultados?. En su opinión, ¿Cuál sería el más adecuado para responder la pregunta de investgación y por qué? ¿Qué variables resultaron ser robustas a la especificación?

R: Dada las características, OLS fue la que peor ajustó, ya que, está diseñada para variables dependientes continuas, no dicotómicas como Probit o Logit. Luego dado el criterio de ajuste, puedo concluir que es ligeramente mejor Logit, por lo que lo prefiero sobre Probit. Los datos entre Probit y Logit fueron robustos entre sí, a diferencia de con OLS que uno de los parámetros no siguió siendo significativo, que debe explicarse debido a la forma en la que trabajan las funciones que operan con variables dicotómicas y la manera en la que perciben como las variables se relacionan con la dependiente.

1.7 6. Agregue la data a nivel mensual, usando la data promedio de las variables (ignorando aquellas categoricas, como la direccion del viento). En particular, genere una variable que cuente la cantidad de fallos observados en un mes, utilice un valor de 0 si en ese mes no se reporto fallos en ningun dia. Use un modelo Poisson para explicar el numero de fallas por mes. Seleccione las variables dependientes a incluir en el modelo final e interprete su significado.

[31]:
```python
df_modelo15 = df.drop(['Date','Estacion',␣
↪'Parameter1_Dir','Parameter2_9am','Parameter2_3pm'],axis=1)
df_modelo15
```

[31]:
```
        Location  Min_Temp  Max_Temp  Leakage  Evaporation  Electricity  \
30             3      11.3      26.5      0.0          0.0          0.0
31             3       9.6      23.9      0.0          0.0          0.0
32             3      10.5      28.8      0.0          0.0          0.0
33             3      12.3      34.6      0.0          0.0          0.0
34             3      12.9      35.8      0.0          0.0          0.0
…            …       …       …       …        …          …
142188        42       3.5      21.8      0.0          0.0          0.0
142189        42       2.8      23.4      0.0          0.0          0.0
142190        42       3.6      25.3      0.0          0.0          0.0
142191        42       5.4      26.9      0.0          0.0          0.0
142192        42       7.8      27.0      0.0          0.0          0.0

        Parameter1_Speed  Parameter3_9am  Parameter3_3pm  Parameter4_9am  \
30                  56.0            19.0            31.0            46.0
31                  41.0            19.0            11.0            44.0
32                  26.0            11.0             7.0            43.0
33                  37.0             6.0            17.0            41.0
34                  41.0             6.0            26.0            41.0
…                  …              …              …              …
142188              31.0            15.0            13.0            59.0
142189              31.0            13.0            11.0            51.0
142190              22.0            13.0             9.0            56.0
142191              37.0             9.0             9.0            53.0
142192              28.0            13.0             7.0            51.0

        Parameter4_3pm  Parameter5_9am  Parameter5_3pm  Parameter7_9am  \
30                26.0          1004.5          1003.2            19.7
31                22.0          1014.4          1013.1            14.9
32                22.0          1018.7          1014.8            17.1
33                12.0          1015.1          1010.3            20.7
34                 9.0          1012.6          1009.2            22.4
…                …              …              …              …
142188            27.0          1024.7          1021.2             9.4
142189            24.0          1024.6          1020.3            10.1
```

```
142190             21.0          1023.5          1019.1             10.9
142191             24.0          1021.0          1016.8             12.5
142192             24.0          1019.4          1016.5             15.1

        Parameter7_3pm  Failure_today  Month  Year  I_Electricity  \
30                25.7            0.0      1  2009              0
31                22.1            0.0      1  2009              0
32                26.5            0.0      1  2009              0
33                33.9            0.0      1  2009              0
34                34.4            0.0      1  2009              0
…                  …             …     …    …                  …
142188            20.9            0.0      6  2017              0
142189            22.4            0.0      6  2017              0
142190            24.5            0.0      6  2017              0
142191            26.1            0.0      6  2017              0
142192            26.0            0.0      6  2017              0

        I_Evaporation
30                  0
31                  0
32                  0
33                  0
34                  0
…                 …
142188              0
142189              0
142190              0
142191              0
142192              0

[139886 rows x 20 columns]
```

[32]:

```python
# Nos aseguramos de trabajar solo con las columnas relevantes
columnas_promedio = df_modelo15.drop(['Failure_today', 'Year', 'Month',␣
↪'Location'], axis=1).select_dtypes(include='number').columns

# Agrupamos: sumamos Failure_today, promediamos el resto
df_modelo2 = df_modelo15.groupby(['Year', 'Month', 'Location']).agg({
    **{col: 'mean' for col in columnas_promedio},
    'Failure_today': 'sum'
}).reset_index()

df_modelo2
```

[32]:

```
      Year  Month  Location   Min_Temp   Max_Temp   Leakage  Evaporation  \
0     2009      1         1  17.932258  32.003226  0.038710     9.090323
1     2009      1         2  16.726667  22.990323  0.496774     7.135484
```

```
2     2009      1         3  16.312903  34.658065  0.251613     0.000000
3     2009      1         4  22.422581  36.058065  0.483871    13.561290
4     2009      1         5  16.154839  32.780645  0.922581     0.000000
…     …    …        …        …          …          …          …
4687  2017      6        45   4.424000  14.744000  0.648000     1.344000
4688  2017      6        46  10.100000  18.356000  9.256000     0.000000
4689  2017      6        47   8.736000  18.616000  3.760000     0.000000
4690  2017      6        48  11.657895  17.700000  4.177778     0.000000
4691  2017      6        49   5.800000  18.754167  0.008333     2.729167

      Electricity  Parameter1_Speed  Parameter3_9am  Parameter3_3pm  \
0       11.787097         39.645161       10.161290       17.966667
1        8.958065               NaN       12.516129       24.903226
2        0.000000         42.677419       11.935484       18.548387
3       10.525806         51.258065       18.516129       25.032258
4        0.000000         41.935484        7.419355       17.466667
…               …                 …               …               …
4687     4.632000         24.040000        4.960000        9.280000
4688     0.000000         34.120000       16.440000       16.440000
4689     0.000000         34.000000        9.520000       16.320000
4690     0.000000         38.894737       15.052632       19.842105
4691     0.000000         27.666667       11.375000       12.833333

      Parameter4_9am  Parameter4_3pm  Parameter5_9am  Parameter5_3pm  \
0          37.612903       23.827586     1014.025806     1012.166667
1          72.806452       68.290323     1015.803226     1014.041935
2          41.903226       17.870968     1013.064516     1009.770968
3          37.096774       24.516129     1008.461290     1004.732258
4          65.516129       35.933333     1015.451613     1012.353333
…                  …               …               …               …
4687       97.840000       67.760000     1028.816000     1026.476000
4688       87.200000       70.880000     1025.720000     1023.492000
4689       88.520000       67.280000     1024.156000     1022.168000
4690       73.315789       69.421053     1026.163158     1024.126316
4691       66.041667       35.875000     1029.704167     1027.033333

      Parameter7_9am  Parameter7_3pm  I_Electricity  I_Evaporation  \
0          23.658065       30.750000       0.967742       0.677419
1          19.651613       21.674194       1.000000       1.000000
2          22.993548       32.964516       0.000000       0.000000
3          29.241935       34.487097       1.000000       1.000000
4          22.390323       31.156667       0.000000       0.000000
…                  …               …              …              …
4687        6.736000       13.696000       1.000000       1.000000
4688       13.168000       17.304000       0.000000       0.000000
4689       12.948000       17.360000       0.000000       0.000000
4690       14.726316       16.757895       0.000000       0.000000
```

```
4691        10.495833        18.070833        0.041667        0.916667

      Failure_today
0               0.0
1               5.0
2               1.0
3               3.0
4               3.0
…             …
4687            3.0
4688           13.0
4689            9.0
4690            4.0
4691            0.0

[4692 rows x 20 columns]
```

[33]:
```
proporcion = (df_modelo2.isna().sum().sum() / df_modelo2.size)*100
c = len(df_modelo2)
print(f'El df_modelo tiene un {proporcion}% de nulos con {c} columnas.')
```

```
El df_modelo tiene un 1.497229326513214% de nulos con 4692 columnas.
```

[34]:
```
df_modelo2 = df_modelo2.dropna()
proporcion = (df_modelo2.isna().sum().sum() / df_modelo2.size)*100
c = len(df_modelo2)
print(f'El df_modelo tiene un {proporcion}% de nulos con {c} columnas.')
```

```
El df_modelo tiene un 0.0% de nulos con 4076 columnas.
```

[37]:
```
y2 = df_modelo2['Failure_today']
X2 = df_modelo2.drop(['Failure_today','Leakage'], axis=1)
X2 = sm.add_constant(X2)
poisson=sm.GLM(y2,X2,family=sm.families.Poisson()).fit()
print(poisson.summary())
```

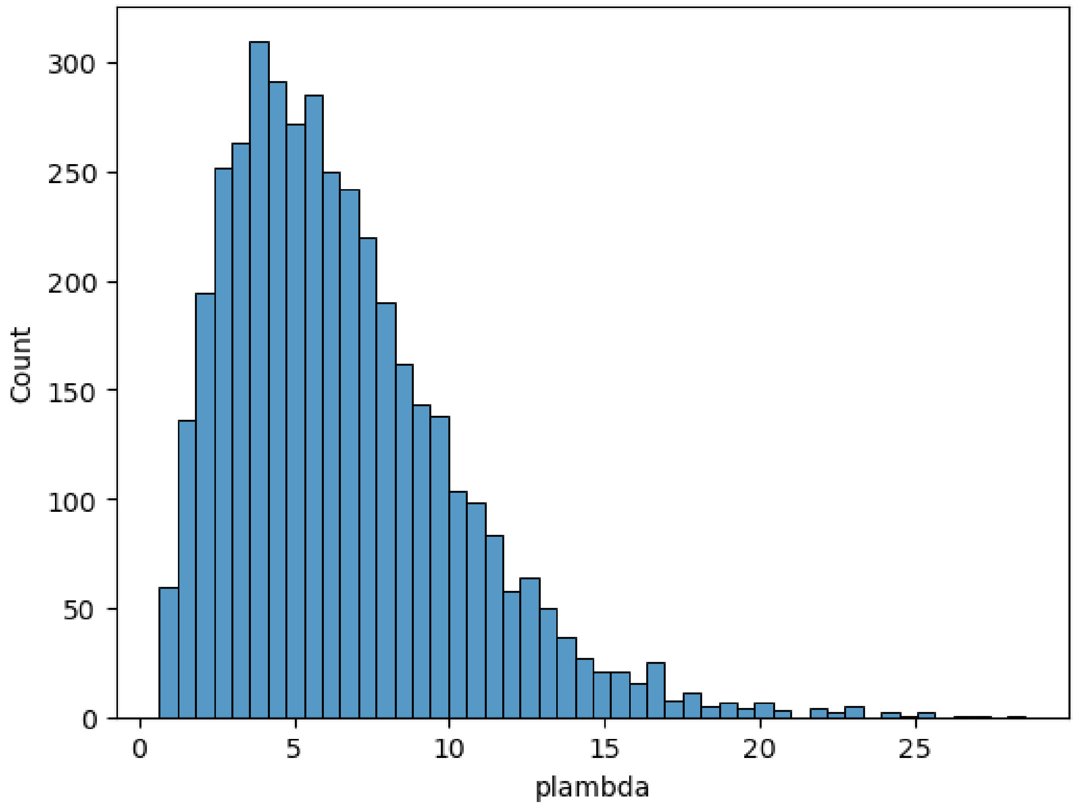
```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:                 4076
Model:                            GLM   Df Residuals:                     4057
Model Family:                 Poisson   Df Model:                           18
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -9328.7
Date:                Thu, 24 Apr 2025   Deviance:                       4847.8
Time:                        23:39:43   Pearson chi2:                 4.49e+03
No. Iterations:                     5   Pseudo R-squ. (CS):             0.8693
Covariance Type:            nonrobust
================================================================================
====
```

```
                     coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----
const             23.4878      5.507      4.265      0.000      12.695
34.281
Year              -0.0017      0.003     -0.667      0.505      -0.007
0.003
Month              0.0182      0.002      8.165      0.000       0.014
0.023
Location          -0.0021      0.000     -4.682      0.000      -0.003
-0.001
Min_Temp          -0.0025      0.007     -0.356      0.722      -0.016
0.011
Max_Temp          -0.1015      0.021     -4.883      0.000      -0.142
-0.061
Evaporation       -0.0078      0.006     -1.398      0.162      -0.019
0.003
Electricity       -0.0587      0.007     -8.335      0.000      -0.073
-0.045
Parameter1_Speed   0.0452      0.002     19.740      0.000       0.041
0.050
Parameter3_9am    -0.0045      0.003     -1.700      0.089      -0.010
0.001
Parameter3_3pm    -0.0565      0.003    -19.160      0.000      -0.062
-0.051
Parameter4_9am     0.0347      0.002     18.088      0.000       0.031
0.039
Parameter4_3pm    -0.0030      0.002     -1.279      0.201      -0.008
0.002
Parameter5_9am    -0.0666      0.012     -5.482      0.000      -0.090
-0.043
Parameter5_3pm     0.0464      0.012      3.793      0.000       0.022
0.070
Parameter7_9am     0.1654      0.012     14.313      0.000       0.143
0.188
Parameter7_3pm    -0.0451      0.024     -1.916      0.055      -0.091
0.001
I_Electricity      0.4964      0.057      8.676      0.000       0.384
0.609
I_Evaporation      0.0107      0.038      0.279      0.780      -0.064
0.086
================================================================================
====
```

[38]:

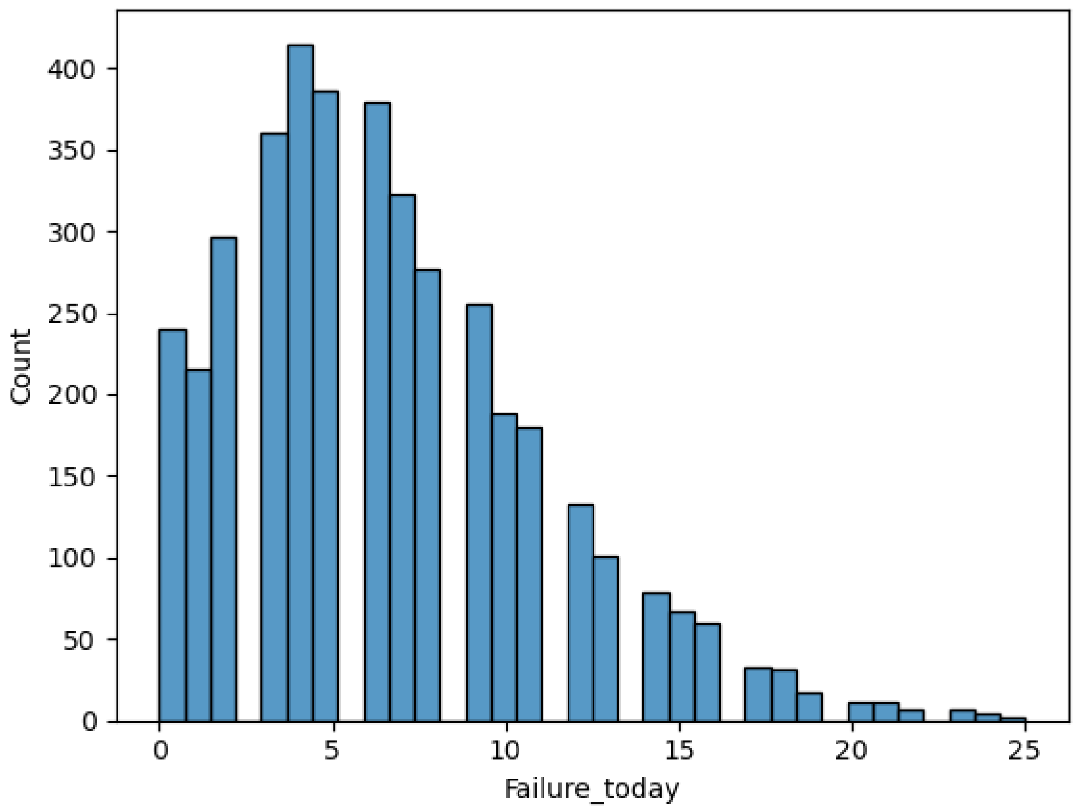
```
df_modelo2['plambda'] = poisson.mu
sns.histplot(data=df_modelo2, x="plambda")
```

[38]: <Axes: xlabel='plambda', ylabel='Count'>



[39]:
```
sns.histplot(data=df_modelo2, x="Failure_today")
```

[39]: <Axes: xlabel='Failure_today', ylabel='Count'>



Por la forma en la que distribuye, y como está el máximo, se puede apreciar un buen ajuste de parte del modelo de Poisson.

1.8 7. Determine sobre dispersion en la data y posible valor optimo de alpha para un modelo Binomial Negativa.

[40]:
```
aux=((y2-poisson.mu)**2-poisson.mu)/poisson.mu
auxr=sm.OLS(aux,poisson.mu).fit()
print(auxr.summary())
```

```
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:          Failure_today   R-squared (uncentered):
0.000
Model:                            OLS   Adj. R-squared (uncentered):
0.000
Method:                 Least Squares   F-statistic:
1.326
Date:                Thu, 24 Apr 2025   Prob (F-statistic):
0.250
```

```
Time:                        23:39:45   Log-Likelihood:
-10842.
No. Observations:                4076   AIC:
2.169e+04
Df Residuals:                    4075   BIC:
2.169e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0083      0.007      1.151      0.250      -0.006       0.022
==============================================================================
Omnibus:                    12716.173   Durbin-Watson:                   1.975
Prob(Omnibus):                  0.000   Jarque-Bera (JB):       1290050427.914
Skew:                          47.876   Prob(JB):                         0.00
Kurtosis:                    2757.417   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[41]:
```python
valalf = np.exp(auxr.params[0])
print(f"Alfa estimado manualmente: {valalf}")
```

```
Alfa estimado manualmente: 1.0082897543162326
```

Notar que el valor de alfa es mayor a 0, por lo que da indicios de que el modelo binomial negativo puede ajustarse mejor, al darse el fenómeno de la sobredispersión.

1.9 8. Usando la informacion anterior, ejecute un modelo Binomial Negativa para responder a la pregunta 6. Seleccione las variables dependientes a incluir en el modelo final e interprete su significado.

[42]:
```python
negbin=sm.GLM(y2,X2,family=sm.families.NegativeBinomial(alpha=valalf)).fit()
print(negbin.summary())
```

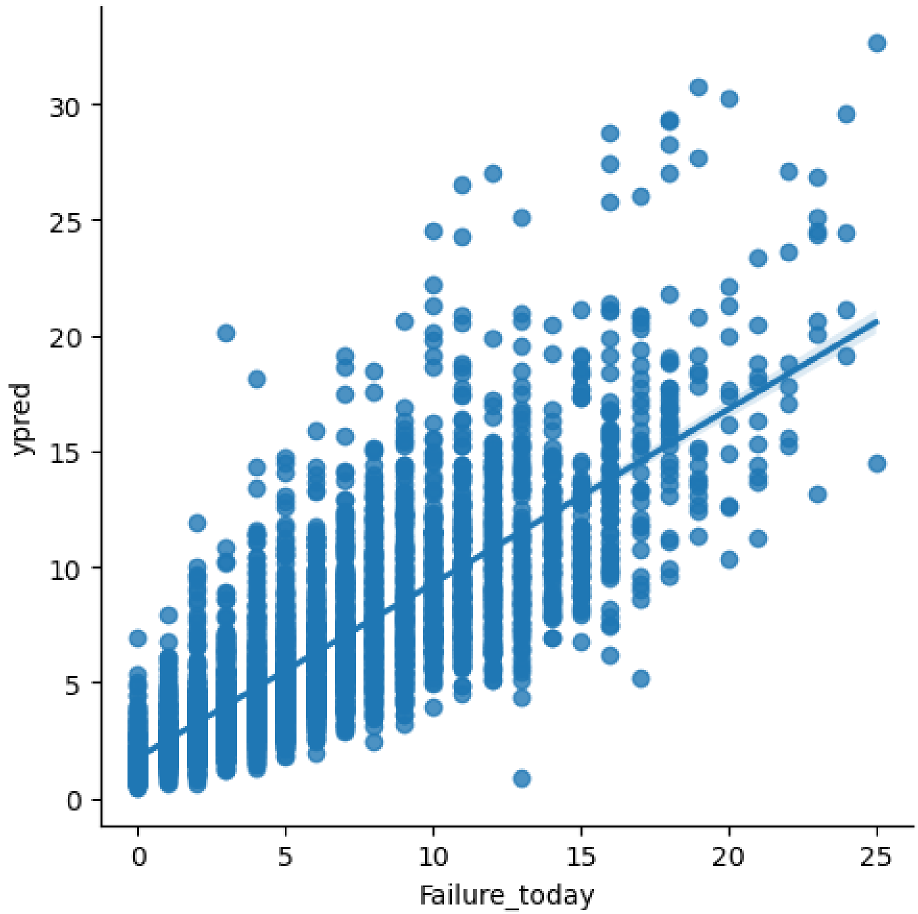
```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:                 4076
Model:                            GLM   Df Residuals:                     4057
Model Family:        NegativeBinomial   Df Model:                           18
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                 -11412.
Date:                Thu, 24 Apr 2025   Deviance:                       1094.3
```

```
Time:                        23:39:46   Pearson chi2:                     840.
No. Iterations:                     8   Pseudo R-squ. (CS):             0.2653
Covariance Type:            nonrobust
================================================================================
====
                      coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----
const              21.4178     15.748      1.360      0.174      -9.447
52.282
Year                0.0007      0.007      0.096      0.923      -0.014
0.015
Month               0.0255      0.006      4.164      0.000       0.014
0.038
Location           -0.0023      0.001     -1.823      0.068      -0.005
0.000
Min_Temp            0.0135      0.018      0.741      0.459      -0.022
0.049
Max_Temp           -0.0653      0.057     -1.147      0.251      -0.177
0.046
Evaporation        -0.0031      0.013     -0.229      0.819      -0.029
0.023
Electricity        -0.0943      0.019     -4.981      0.000      -0.131
-0.057
Parameter1_Speed    0.0502      0.007      7.611      0.000       0.037
0.063
Parameter3_9am     -0.0020      0.007     -0.283      0.777      -0.016
0.012
Parameter3_3pm     -0.0696      0.008     -8.536      0.000      -0.086
-0.054
Parameter4_9am      0.0414      0.005      8.044      0.000       0.031
0.051
Parameter4_3pm     -0.0131      0.007     -1.989      0.047      -0.026
-0.000
Parameter5_9am     -0.1140      0.034     -3.400      0.001      -0.180
-0.048
Parameter5_3pm      0.0913      0.034      2.696      0.007       0.025
0.158
Parameter7_9am      0.1933      0.031      6.169      0.000       0.132
0.255
Parameter7_3pm     -0.1241      0.064     -1.929      0.054      -0.250
0.002
I_Electricity       0.7649      0.165      4.646      0.000       0.442
1.088
I_Evaporation      -0.0307      0.102     -0.301      0.763      -0.231
0.169
================================================================================
```

====

[43]:
```
df_negbin = df_modelo2
df_negbin['ypred'] = negbin.predict(X2)
sns.lmplot(data=df_negbin, x='Failure_today', y='ypred')
```

[43]: <seaborn.axisgrid.FacetGrid at 0x131bc4fb0>



Se puede ver que el modelo sigue el patron de tendencia de la cantidad observada.

1.10 9. Comente los resultados obtenidos en 6, 7 y 8. ¿Cuáles y por qué existen las diferencias entre los resultados?. En su opinión, ¿Cuál sería el más adecuado para responder la pregunta de investgación y por qué? ¿Qué variables resultaron ser robustas a la especificación?

[44]:
```
summary_poisson = poisson.summary2().tables[1][["Coef.", "P>|z|"]]
summary_negbin = negbin.summary2().tables[1][["Coef.", "P>|z|"]]

summary_poisson.columns = ['Coef_Poisson', 'P_Poisson']
summary_negbin.columns = ['Coef_NegBin', 'P_NegBin']

comparacion = summary_poisson.join(summary_negbin, how='outer')
comparacion.round(3)
```

[44]:

	Coef_Poisson	P_Poisson	Coef_NegBin	P_NegBin
Electricity	-0.059	0.000	-0.094	0.000
Evaporation	-0.008	0.162	-0.003	0.819
I_Electricity	0.496	0.000	0.765	0.000
I_Evaporation	0.011	0.780	-0.031	0.763
Location	-0.002	0.000	-0.002	0.068
Max_Temp	-0.101	0.000	-0.065	0.251
Min_Temp	-0.003	0.722	0.013	0.459
Month	0.018	0.000	0.026	0.000
Parameter1_Speed	0.045	0.000	0.050	0.000
Parameter3_3pm	-0.057	0.000	-0.070	0.000
Parameter3_9am	-0.004	0.089	-0.002	0.777
Parameter4_3pm	-0.003	0.201	-0.013	0.047
Parameter4_9am	0.035	0.000	0.041	0.000
Parameter5_3pm	0.046	0.000	0.091	0.007
Parameter5_9am	-0.067	0.000	-0.114	0.001
Parameter7_3pm	-0.045	0.055	-0.124	0.054
Parameter7_9am	0.165	0.000	0.193	0.000
Year	-0.002	0.505	0.001	0.923
const	23.488	0.000	21.418	0.174

[45]:
```
alpha = 0.05

def evaluar_robustez(row):
    signif_poisson = row['P_Poisson'] < alpha
    signif_negbin = row['P_NegBin'] < alpha

    if signif_poisson and signif_negbin:
        return 'Robusta (significativa en ambos)'
    elif not signif_poisson and not signif_negbin:
        return 'Robusta (no significativa en ambos)'
    elif signif_poisson and not signif_negbin:
        return 'No robusta (solo significativa en Poisson)'
```

```
    elif not signif_poisson and signif_negbin:
        return 'No robusta (solo significativa en Binomial Negativa)'

def evaluar_signo(row):
    signo_poisson = row['Coef_Poisson'] >= 0
    signo_negbin = row['Coef_NegBin'] >= 0
    if signo_poisson == signo_negbin:
        return 'Mantiene el signo'
    else:
        return 'Cambia el signo'

comparacion['Robustez_P'] = comparacion.apply(evaluar_robustez, axis=1)
comparacion['Cambio_Signo'] = comparacion.apply(evaluar_signo, axis=1)
comparacion[['Robustez_P','Cambio_Signo']]
```

[45]:

```
                                                     Robustez_P  \
Electricity                      Robusta (significativa en ambos)
Evaporation                   Robusta (no significativa en ambos)
I_Electricity                    Robusta (significativa en ambos)
I_Evaporation                 Robusta (no significativa en ambos)
Location                  No robusta (solo significativa en Poisson)
Max_Temp                  No robusta (solo significativa en Poisson)
Min_Temp                      Robusta (no significativa en ambos)
Month                            Robusta (significativa en ambos)
Parameter1_Speed                 Robusta (significativa en ambos)
Parameter3_3pm                   Robusta (significativa en ambos)
Parameter3_9am                Robusta (no significativa en ambos)
Parameter4_3pm    No robusta (solo significativa en Binomial Neg…
Parameter4_9am                   Robusta (significativa en ambos)
Parameter5_3pm                   Robusta (significativa en ambos)
Parameter5_9am                   Robusta (significativa en ambos)
Parameter7_3pm                Robusta (no significativa en ambos)
Parameter7_9am                   Robusta (significativa en ambos)
Year                          Robusta (no significativa en ambos)
const                     No robusta (solo significativa en Poisson)

                       Cambio_Signo
Electricity       Mantiene el signo
Evaporation       Mantiene el signo
I_Electricity     Mantiene el signo
I_Evaporation       Cambia el signo
Location          Mantiene el signo
Max_Temp          Mantiene el signo
Min_Temp            Cambia el signo
Month             Mantiene el signo
Parameter1_Speed  Mantiene el signo
Parameter3_3pm    Mantiene el signo
```

```
Parameter3_9am    Mantiene el signo
Parameter4_3pm    Mantiene el signo
Parameter4_9am    Mantiene el signo
Parameter5_3pm    Mantiene el signo
Parameter5_9am    Mantiene el signo
Parameter7_3pm    Mantiene el signo
Parameter7_9am    Mantiene el signo
Year                Cambia el signo
const             Mantiene el signo
```

R: Las diferencias se notan al analizar como el R-cuadrado cambia entre uno y otro donde podemos ver, que dado el supuesto que relaja la sobredispersión, muestra resultados con un R-cuadrado menor en el modelo de binomial negativa en comparación al modelo de Poisson, que tiene mejor ajuste en los datos. Por lo mismo yo creo, y analizando los outputs, que gracias a percibir la sobredispersión de los datos, el modelo de binomial negativa logra ajustarse de mejor manera a los datos. Sobre la robustez de las variables, es más fácil hablar de las que no fueron, que son 'Max_Temp', 'Parameter3_9am',y la constante; y también es relevante nombrar como cambiaron de signo el efecto de ser medida la evaporación ('I_Evaporation' reduce la probabilidad de fallo en binomial negativa), y en 'Min_Temp' (aumenta la probabilidad de fallo en binomial negativa).