Tarea1_Perez_Macaya

April 30, 2025

[1]:
```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
import numpy as np
```

[2]:
```python
#Leer CSV
df = pd.read_csv("C:/Users/edins/OneDrive/Documentos/LAB-MAA/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]
```

Estadística descriptiva:

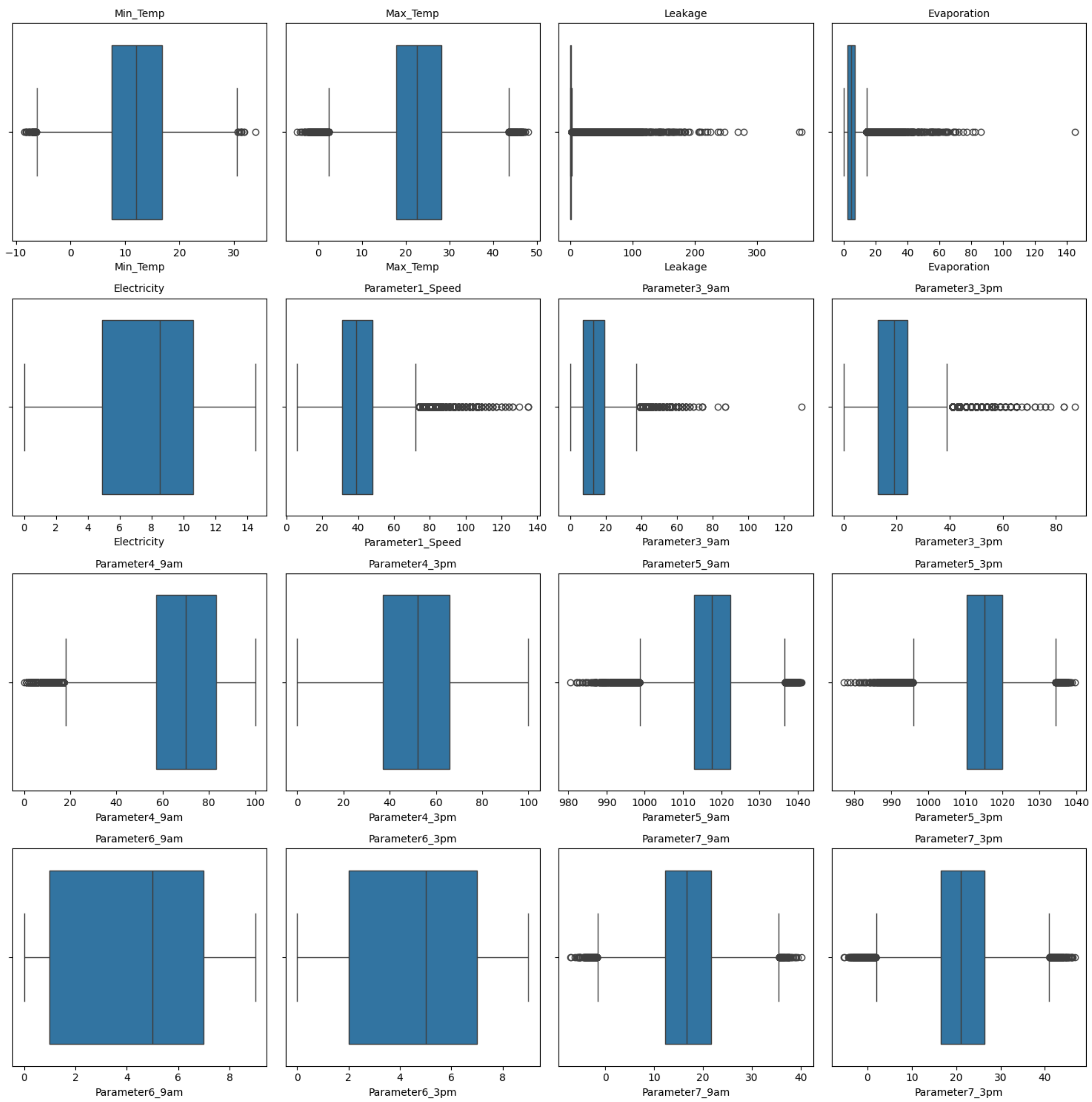
[3]:
```
df.describe()
```

[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
```

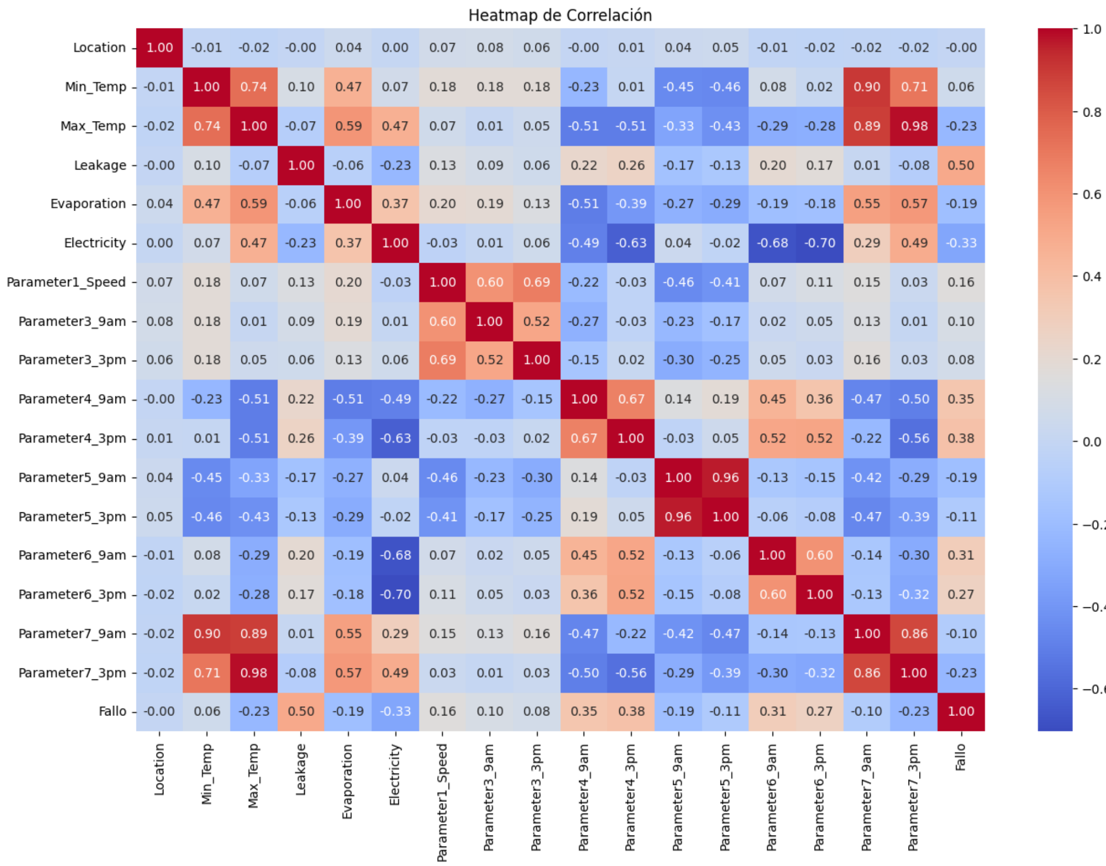
[4]:
```python
fig, axes = plt.subplots(4, 4,figsize=(15, 15))

sns.boxplot(x=df['Min_Temp'],data=df,ax=axes[0, 0])
axes[0, 0].set_title('Min_Temp',fontsize=10)

sns.boxplot(x=df['Max_Temp'],data=df,ax=axes[0, 1])
axes[0, 1].set_title('Max_Temp',fontsize=10)

sns.boxplot(x=df['Leakage'],data=df,ax=axes[0, 2])
axes[0, 2].set_title('Leakage',fontsize=10)

sns.boxplot(x=df['Evaporation'],data=df,ax=axes[0, 3])
axes[0, 3].set_title('Evaporation',fontsize=10)

sns.boxplot(x=df['Electricity'],data=df,ax=axes[1, 0])
axes[1, 0].set_title('Electricity',fontsize=10)

sns.boxplot(x=df['Parameter1_Speed'],data=df,ax=axes[1, 1])
axes[1, 1].set_title('Parameter1_Speed',fontsize=10)

sns.boxplot(x=df['Parameter3_9am'],data=df,ax=axes[1, 2])
axes[1, 2].set_title('Parameter3_9am',fontsize=10)

sns.boxplot(x=df['Parameter3_3pm'],data=df,ax=axes[1, 3])
axes[1, 3].set_title('Parameter3_3pm',fontsize=10)

sns.boxplot(x=df['Parameter4_9am'],data=df,ax=axes[2, 0])
axes[2, 0].set_title('Parameter4_9am',fontsize=10)

sns.boxplot(x=df['Parameter4_3pm'],data=df,ax=axes[2, 1])
axes[2, 1].set_title('Parameter4_3pm',fontsize=10)

sns.boxplot(x=df['Parameter5_9am'],data=df,ax=axes[2, 2])
axes[2, 2].set_title('Parameter5_9am',fontsize=10)

sns.boxplot(x=df['Parameter5_3pm'],data=df,ax=axes[2, 3])
axes[2, 3].set_title('Parameter5_3pm',fontsize=10)

sns.boxplot(x=df['Parameter6_9am'],data=df,ax=axes[3, 0])
axes[3, 0].set_title('Parameter6_9am',fontsize=10)

sns.boxplot(x=df['Parameter6_3pm'],data=df,ax=axes[3, 1])
axes[3, 1].set_title('Parameter6_3pm',fontsize=10)

sns.boxplot(x=df['Parameter7_9am'],data=df,ax=axes[3, 2])
axes[3, 2].set_title('Parameter7_9am',fontsize=10)
```

```
sns.boxplot(x=df['Parameter7_3pm'],data=df,ax=axes[3, 3])
axes[3, 3].set_title('Parameter7_3pm',fontsize=10)

plt.tight_layout()
plt.show()
```



[5]:
```
df_numerico = df.select_dtypes(include='number')
df_numerico['Fallo'] = df['Failure_today'].map({'Yes': 1, 'No': 0})
corr = df_numerico.corr()
plt.figure(figsize=(15, 10))
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Heatmap de Correlación')
```

```python
plt.show()
```



Podemos observar que Leakage esta altamente relacionado a la variable Fallo, por lo que se concluirá que es sobre explicativa

Limpieza de datos:

[6]:
```python
#vemos cuantos NaN hay
df.isnull().sum().sum()
```

[6]: np.int64(316559)

Vemos que existen muchos valores NaN (316 559), pero al aplicar .dropna() se nos elimina más de la mitad del DataFrame, por lo que la presencia de NaN podría estar relacionada con la variable de fallo. Se proseguira de la siguiente forma:

[7]:
```python
#Calculamos el porcentaje de NaN por columna

porcentaje_nan = (df.isnull().sum() / len(df)) * 100
porcentaje_nan = porcentaje_nan[porcentaje_nan > 0].sort_values(ascending=False)
print(porcentaje_nan)
```

```
Electricity         47.692924
Evaporation         42.789026
Parameter6_3pm      40.152469
Parameter6_9am      37.735332
Parameter5_9am       9.855619
Parameter5_3pm       9.832411
Parameter2_9am       7.041838
Parameter1_Dir       6.561504
Parameter1_Speed     6.519308
Parameter2_3pm       2.656952
Parameter4_3pm       2.538803
Parameter7_3pm       1.917113
Parameter3_3pm       1.849599
Parameter4_9am       1.247600
Leakage              0.988797
Failure_today        0.988797
Parameter3_9am       0.948007
Parameter7_9am       0.635756
Min_Temp             0.447983
Max_Temp             0.226453
dtype: float64
```

Las columnas con mayor porcentaje de NaN son Electricity, Evaporation, Parameter6_3pm y Parameter6_9am. Transformamos estas variables a binarias (0 = variable no se midió (NaN) , 1 = variable se midió) para luego aplicar .dropna()

[8]:
```python
data = df
data['Electricity'] = data['Electricity'].notnull().astype(int)
data['Evaporation'] = data['Evaporation'].notnull().astype(int)
data['Parameter6_3pm'] = data['Parameter6_3pm'].notnull().astype(int)
data['Parameter6_9am'] = data['Parameter6_9am'].notnull().astype(int)
data =data.dropna()
data
```

[8]:
```
                 Date  Location  Min_Temp  Max_Temp  Leakage  Evaporation  \
0       12/1/2008         3      13.4      22.9      0.6            0
1       12/2/2008         3       7.4      25.1      0.0            0
2       12/3/2008         3      12.9      25.7      0.0            0
3       12/4/2008         3       9.2      28.0      0.0            0
4       12/5/2008         3      17.5      32.3      1.0            0
…             …       …        …        …        …          …
142188  6/20/2017        42       3.5      21.8      0.0            0
142189  6/21/2017        42       2.8      23.4      0.0            0
142190  6/22/2017        42       3.6      25.3      0.0            0
142191  6/23/2017        42       5.4      26.9      0.0            0
142192  6/24/2017        42       7.8      27.0      0.0            0

        Electricity Parameter1_Dir  Parameter1_Speed Parameter2_9am  …  \
```

```
0                  0             W             44.0           W  …
1                  0           WNW             44.0         NNW  …
2                  0           WSW             46.0           W  …
3                  0            NE             24.0          SE  …
4                  0             W             41.0         ENE  …
…             …           …             …            … …
142188             0             E             31.0         ESE  …
142189             0             E             31.0          SE  …
142190             0           NNW             22.0          SE  …
142191             0             N             37.0          SE  …
142192             0            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               1               0            16.9
1               1007.8               0               0            17.2
2               1008.7               0               1            21.0
3               1012.8               0               0            18.1
4               1006.0               1               1            17.8
…                …             …             …             …
142188          1021.2               0               0             9.4
142189          1020.3               0               0            10.1
142190          1019.1               0               0            10.9
142191          1016.8               0               0            12.5
142192          1016.5               1               1            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

[112925 rows x 22 columns]
```

[9]:
```
data['Fallo'] = data['Failure_today'].map({'Yes': 1, 'No': 0})
data['mes'] = pd.to_datetime(data['Date'], format='%m/%d/%Y').dt.month␣
 ↪#extraemos el mes
data['Estacion'] = data['mes'].map({1:'I',2:'I',3:'I',4:'P',5:'P',6:'P',7:'V',8:
 ↪'V',9:'V',10:'O',11:'O',12:'O'}) #mapeamos numero de mes a estacion

data = pd.get_dummies(data,columns=['Estacion'],drop_first=True,dtype=int)
data['Parametro1_Dir']=data['Parameter1_Dir'].map({'NW':'N','NNW':'N','N':
 ↪'N','NNE':'N','NE':'N','ENE':'E','E':'E','ESE':'E','SE':'S',
                                                  'SSE':'S','S':'S','SSW':
 ↪'S','SW':'S','WSW':'W','W':'W','WNW':'W'})
data = pd.get_dummies(data,columns=['Parametro1_Dir'],drop_first=True,dtype=int)
data['Parametro2_9am']=data['Parameter2_9am'].map({'NW':'N','NNW':'N','N':
 ↪'N','NNE':'N','NE':'N','ENE':'E','E':'E','ESE':'E','SE':'S',
                                                  'SSE':'S','S':'S','SSW':
 ↪'S','SW':'S','WSW':'W','W':'W','WNW':'W'})
data = pd.get_dummies(data,columns=['Parametro2_9am'],drop_first=True,dtype=int)
data['Parametro2_3pm']=data['Parameter2_3pm'].map({'NW':'N','NNW':'N','N':
 ↪'N','NNE':'N','NE':'N','ENE':'E','E':'E','ESE':'E','SE':'S',
                                                  'SSE':'S','S':'S','SSW':
 ↪'S','SW':'S','WSW':'W','W':'W','WNW':'W'})
data = pd.get_dummies(data,columns=['Parametro2_3pm'],drop_first=True,dtype=int)

data = pd.get_dummies(data, columns=['Location'], prefix='loc',␣
 ↪drop_first=True, dtype=int)

data
```

```
C:\Users\edins\AppData\Local\Temp\ipykernel_12548\1657808883.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data['Fallo'] = data['Failure_today'].map({'Yes': 1, 'No': 0})
C:\Users\edins\AppData\Local\Temp\ipykernel_12548\1657808883.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data['mes'] = pd.to_datetime(data['Date'], format='%m/%d/%Y').dt.month
#extraemos el mes
C:\Users\edins\AppData\Local\Temp\ipykernel_12548\1657808883.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data['Estacion'] = data['mes'].map({1:'I',2:'I',3:'I',4:'P',5:'P',6:'P',7:'V',
8:'V',9:'V',10:'O',11:'O',12:'O'}) #mapeamos numero de mes a estacion
```

[9]:

```
              Date  Min_Temp  Max_Temp  Leakage  Evaporation  Electricity  \
0        12/1/2008      13.4      22.9      0.6            0            0
1        12/2/2008       7.4      25.1      0.0            0            0
2        12/3/2008      12.9      25.7      0.0            0            0
3        12/4/2008       9.2      28.0      0.0            0            0
4        12/5/2008      17.5      32.3      1.0            0            0
…              …       …        …         …          …            …
142188   6/20/2017       3.5      21.8      0.0            0            0
142189   6/21/2017       2.8      23.4      0.0            0            0
142190   6/22/2017       3.6      25.3      0.0            0            0
142191   6/23/2017       5.4      26.9      0.0            0            0
142192   6/24/2017       7.8      27.0      0.0            0            0

       Parameter1_Dir  Parameter1_Speed Parameter2_9am Parameter2_3pm  …  \
0                   W              44.0              W            WNW  …
1                 WNW              44.0            NNW            WSW  …
2                 WSW              46.0              W            WSW  …
3                  NE              24.0             SE              E  …
4                   W              41.0            ENE             NW  …
…               …                …               …              …   …
142188              E              31.0            ESE              E  …
142189              E              31.0             SE            ENE  …
142190            NNW              22.0             SE              N  …
142191              N              37.0             SE            WNW  …
142192             SE              28.0            SSE              N  …

        loc_40  loc_41  loc_42  loc_43  loc_44  loc_45  loc_46  loc_47  \
0            0       0       0       0       0       0       0       0
1            0       0       0       0       0       0       0       0
2            0       0       0       0       0       0       0       0
3            0       0       0       0       0       0       0       0
4            0       0       0       0       0       0       0       0
…          …     …     …     …     …     …     …     …     …
```

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

        loc_48  loc_49
0            0       0
1            0       0
2            0       0
3            0       0
4            0       0
…        …     …
142188       0       0
142189       0       0
142190       0       0
142191       0       0
142192       0       0

[112925 rows x 78 columns]
```

2. Regresion OLS

[11]:
```
y = data['Fallo']
X = data.drop(['Fallo','Failure_today','Leakage', 'Date',␣
  ↪'mes','Parameter1_Dir','Parameter2_9am','Parameter2_3pm'], axis=1)
X=sm.add_constant(X)
model = sm.OLS(y, X)
results = model.fit(cov_type='HC0')
print(results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Fallo   R-squared:                       0.302
Model:                            OLS   Adj. R-squared:                  0.301
Method:                 Least Squares   F-statistic:                     718.4
Date:                Thu, 24 Apr 2025   Prob (F-statistic):               0.00
Time:                        22:53:56   Log-Likelihood:                -41295.
No. Observations:              112925   AIC:                         8.273e+04
Df Residuals:                  112854   BIC:                         8.342e+04
Df Model:                          70
Covariance Type:                  HC0
================================================================================
====
                       coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----
```

const	7.8297	0.226	34.582	0.000	7.386	8.274
Min_Temp	0.0097	0.001	19.197	0.000	0.009	0.011
Max_Temp	-0.0329	0.001	-32.326	0.000	-0.035	-0.031
Evaporation	-0.0159	0.005	-3.144	0.002	-0.026	-0.006
Electricity	-0.0053	0.005	-1.085	0.278	-0.015	0.004
Parameter1_Speed	0.0054	0.000	38.559	0.000	0.005	0.006
Parameter3_9am	0.0027	0.000	15.174	0.000	0.002	0.003
Parameter3_3pm	-0.0040	0.000	-21.351	0.000	-0.004	-0.004
Parameter4_9am	0.0074	0.000	62.607	0.000	0.007	0.008
Parameter4_3pm	0.0023	0.000	16.707	0.000	0.002	0.003
Parameter5_9am	-0.0389	0.001	-50.299	0.000	-0.040	-0.037
Parameter5_3pm	0.0308	0.001	39.574	0.000	0.029	0.032
Parameter6_9am	0.0277	0.005	6.027	0.000	0.019	0.037
Parameter6_3pm	0.0215	0.004	5.158	0.000	0.013	0.030
Parameter7_9am	-0.0011	0.001	-1.397	0.162	-0.003	0.000
Parameter7_3pm	0.0272	0.001	24.135	0.000	0.025	0.029
Estacion_O	0.0604	0.003	18.871	0.000	0.054	0.067
Estacion_P	0.0277	0.004	7.101	0.000	0.020	0.035
Estacion_V	0.0674	0.004	15.300	0.000	0.059	0.076
Parametro1_Dir_N	-0.0097	0.004	-2.676	0.007	-0.017	-0.003
Parametro1_Dir_S	-0.0005	0.004	-0.140	0.889	-0.007	0.006
Parametro1_Dir_W	0.0018	0.004	0.436	0.663	-0.006	0.010
Parametro2_9am_N	-0.0016	0.003	-0.509	0.610	-0.008	0.005
Parametro2_9am_S	0.0214	0.003	6.722	0.000	0.015	0.028

Parametro2_9am_W	0.0324	0.004	7.691	0.000	0.024	0.041
Parametro2_3pm_N	0.0026	0.004	0.731	0.465	-0.004	0.010
Parametro2_3pm_S	0.0171	0.004	4.874	0.000	0.010	0.024
Parametro2_3pm_W	0.0241	0.004	5.739	0.000	0.016	0.032
loc_3	-0.0922	0.010	-9.141	0.000	-0.112	-0.072
loc_4	0.0704	0.009	7.831	0.000	0.053	0.088
loc_5	-0.1012	0.010	-9.838	0.000	-0.121	-0.081
loc_6	-0.2546	0.011	-23.126	0.000	-0.276	-0.233
loc_7	-0.1411	0.010	-14.119	0.000	-0.161	-0.121
loc_8	-0.0324	0.010	-3.159	0.002	-0.052	-0.012
loc_9	-0.0961	0.011	-8.594	0.000	-0.118	-0.074
loc_10	-0.1106	0.010	-11.202	0.000	-0.130	-0.091
loc_11	-0.0615	0.010	-6.302	0.000	-0.081	-0.042
loc_12	-0.0643	0.011	-5.915	0.000	-0.086	-0.043
loc_13	-0.1383	0.011	-12.952	0.000	-0.159	-0.117
loc_14	-0.1320	0.011	-12.340	0.000	-0.153	-0.111
loc_15	-0.0739	0.010	-7.229	0.000	-0.094	-0.054
loc_16	-0.1537	0.010	-15.417	0.000	-0.173	-0.134
loc_17	-0.1278	0.016	-8.158	0.000	-0.159	-0.097
loc_18	-0.1417	0.012	-12.051	0.000	-0.165	-0.119
loc_19	-0.1334	0.011	-12.093	0.000	-0.155	-0.112
loc_20	-0.1881	0.010	-18.186	0.000	-0.208	-0.168
loc_21	-0.1278	0.009	-13.729	0.000	-0.146	-0.110
loc_22	-0.0776	0.009	-8.248	0.000	-0.096	-0.059

```
loc_23               -0.1265      0.011    -12.029      0.000      -0.147
-0.106
loc_26               -0.1731      0.011    -15.784      0.000      -0.195
-0.152
loc_27               -0.1502      0.010    -14.448      0.000      -0.171
-0.130
loc_28               -0.1850      0.011    -17.049      0.000      -0.206
-0.164
loc_29               -0.1076      0.010    -11.129      0.000      -0.127
-0.089
loc_30               -0.0669      0.010     -6.383      0.000      -0.087
-0.046
loc_32               -0.0681      0.010     -7.168      0.000      -0.087
-0.050
loc_33               -0.0699      0.010     -7.242      0.000      -0.089
-0.051
loc_34               -0.1474      0.011    -13.827      0.000      -0.168
-0.127
loc_35               -0.0940      0.011     -8.824      0.000      -0.115
-0.073
loc_36               -0.2286      0.011    -21.525      0.000      -0.249
-0.208
loc_38               -0.1126      0.011    -10.205      0.000      -0.134
-0.091
loc_39               -0.1065      0.010    -10.170      0.000      -0.127
-0.086
loc_40               -0.1453      0.010    -14.170      0.000      -0.165
-0.125
loc_41               -0.0646      0.010     -6.407      0.000      -0.084
-0.045
loc_42                0.0468      0.010      4.762      0.000       0.028
0.066
loc_43               -0.0910      0.010     -9.317      0.000      -0.110
-0.072
loc_44               -0.1077      0.011    -10.087      0.000      -0.129
-0.087
loc_45               -0.1769      0.010    -17.044      0.000      -0.197
-0.157
loc_46               -0.0945      0.011     -8.391      0.000      -0.117
-0.072
loc_47               -0.0600      0.011     -5.641      0.000      -0.081
-0.039
loc_48               -0.1940      0.010    -18.721      0.000      -0.214
-0.174
loc_49               -0.1159      0.009    -12.918      0.000      -0.133
-0.098
==============================================================================
Omnibus:                     8422.921   Durbin-Watson:                   1.799
```

```
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            10418.269
Skew:                           0.741   Prob(JB):                         0.00
Kurtosis:                       2.864   Cond. No.                     2.95e+05
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 2.95e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

3. Probit

[12]:
```python
model = sm.Probit(y, X)
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: 0.352421
         Iterations 7
                          Probit Regression Results
==============================================================================
Dep. Variable:                  Fallo   No. Observations:               112925
Model:                         Probit   Df Residuals:                   112854
Method:                           MLE   Df Model:                           70
Date:                Thu, 24 Apr 2025   Pseudo R-squ.:                  0.3385
Time:                        22:53:57   Log-Likelihood:                -39797.
converged:                       True   LL-Null:                       -60159.
Covariance Type:                  HC0   LLR p-value:                     0.000
================================================================================
====
                       coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----
const               28.1011      1.004     27.987      0.000      26.133
30.069
Min_Temp             0.0752      0.003     25.360      0.000       0.069
0.081
Max_Temp            -0.1438      0.005    -26.787      0.000      -0.154
-0.133
Evaporation         -0.0181      0.025     -0.718      0.473      -0.067
0.031
Electricity         -0.0540      0.025     -2.150      0.032      -0.103
-0.005
Parameter1_Speed     0.0200      0.001     31.699      0.000       0.019
0.021
```

Parameter3_9am	0.0092	0.001	10.367	0.000	0.007	0.011
Parameter3_3pm	-0.0136	0.001	-15.137	0.000	-0.015	-0.012
Parameter4_9am	0.0407	0.001	63.869	0.000	0.039	0.042
Parameter4_3pm	0.0020	0.001	3.223	0.001	0.001	0.003
Parameter5_9am	-0.1369	0.004	-37.841	0.000	-0.144	-0.130
Parameter5_3pm	0.1062	0.004	29.550	0.000	0.099	0.113
Parameter6_9am	0.0717	0.025	2.833	0.005	0.022	0.121
Parameter6_3pm	0.1363	0.024	5.726	0.000	0.090	0.183
Parameter7_9am	-0.0050	0.005	-1.096	0.273	-0.014	0.004
Parameter7_3pm	0.0630	0.006	10.664	0.000	0.051	0.075
Estacion_O	0.2241	0.016	13.657	0.000	0.192	0.256
Estacion_P	-0.0126	0.019	-0.669	0.504	-0.049	0.024
Estacion_V	0.1072	0.022	4.952	0.000	0.065	0.150
Parametro1_Dir_N	-0.0914	0.020	-4.466	0.000	-0.132	-0.051
Parametro1_Dir_S	-0.0191	0.019	-1.026	0.305	-0.056	0.017
Parametro1_Dir_W	0.0035	0.022	0.160	0.873	-0.039	0.046
Parametro2_9am_N	0.0169	0.019	0.889	0.374	-0.020	0.054
Parametro2_9am_S	0.1647	0.018	9.218	0.000	0.130	0.200
Parametro2_9am_W	0.1770	0.021	8.616	0.000	0.137	0.217
Parametro2_3pm_N	-0.0064	0.020	-0.322	0.748	-0.045	0.033
Parametro2_3pm_S	0.0282	0.018	1.537	0.124	-0.008	0.064
Parametro2_3pm_W	0.0544	0.022	2.496	0.013	0.012	0.097
loc_3	-0.3516	0.056	-6.305	0.000	-0.461	-0.242
loc_4	0.1703	0.066	2.577	0.010	0.041	0.300

loc_5	-0.2281	0.050	-4.587	0.000	-0.325	-0.131
loc_6	-1.1824	0.057	-20.820	0.000	-1.294	-1.071
loc_7	-0.6240	0.056	-11.215	0.000	-0.733	-0.515
loc_8	0.2017	0.052	3.909	0.000	0.101	0.303
loc_9	-0.0846	0.052	-1.621	0.105	-0.187	0.018
loc_10	-0.3337	0.053	-6.323	0.000	-0.437	-0.230
loc_11	-0.3257	0.061	-5.320	0.000	-0.446	-0.206
loc_12	-0.0337	0.051	-0.659	0.510	-0.134	0.067
loc_13	-0.6092	0.048	-12.659	0.000	-0.704	-0.515
loc_14	-0.1870	0.054	-3.456	0.001	-0.293	-0.081
loc_15	-0.0159	0.049	-0.327	0.744	-0.111	0.079
loc_16	-0.4961	0.050	-9.915	0.000	-0.594	-0.398
loc_17	-0.1745	0.085	-2.044	0.041	-0.342	-0.007
loc_18	-0.4679	0.054	-8.611	0.000	-0.574	-0.361
loc_19	-0.4000	0.053	-7.506	0.000	-0.504	-0.296
loc_20	-0.7009	0.053	-13.248	0.000	-0.805	-0.597
loc_21	-0.7528	0.058	-13.018	0.000	-0.866	-0.639
loc_22	-0.1304	0.057	-2.287	0.022	-0.242	-0.019
loc_23	-0.5246	0.051	-10.213	0.000	-0.625	-0.424
loc_26	-0.8828	0.062	-14.351	0.000	-1.003	-0.762
loc_27	-0.4745	0.048	-9.980	0.000	-0.568	-0.381
loc_28	-0.6004	0.050	-11.909	0.000	-0.699	-0.502
loc_29	-0.6098	0.055	-11.017	0.000	-0.718	-0.501
loc_30	-0.1004	0.059	-1.696	0.090	-0.216	0.016

```
loc_32              -0.1536      0.052     -2.962      0.003      -0.255
-0.052
loc_33              -0.1242      0.054     -2.318      0.020      -0.229
-0.019
loc_34              -0.6285      0.050    -12.641      0.000      -0.726
-0.531
loc_35              -0.2420      0.053     -4.589      0.000      -0.345
-0.139
loc_36              -0.8198      0.053    -15.348      0.000      -0.924
-0.715
loc_38              -0.2714      0.051     -5.320      0.000      -0.371
-0.171
loc_39              -0.2795      0.053     -5.283      0.000      -0.383
-0.176
loc_40              -0.2793      0.055     -5.117      0.000      -0.386
-0.172
loc_41              -0.1088      0.050     -2.175      0.030      -0.207
-0.011
loc_42               0.0790      0.081      0.971      0.332      -0.081
0.239
loc_43              -0.3431      0.056     -6.093      0.000      -0.453
-0.233
loc_44              -0.3534      0.048     -7.429      0.000      -0.447
-0.260
loc_45              -0.7068      0.053    -13.450      0.000      -0.810
-0.604
loc_46              -0.1734      0.054     -3.219      0.001      -0.279
-0.068
loc_47              -0.1085      0.049     -2.205      0.027      -0.205
-0.012
loc_48              -0.6694      0.051    -13.001      0.000      -0.770
-0.568
loc_49              -0.8366      0.065    -12.809      0.000      -0.965
-0.709
================================================================================
====
       Probit Marginal Effects
=====================================
Dep. Variable:                  Fallo
Method:                          dydx
At:                           overall
================================================================================
====
                      dy/dx    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----
Min_Temp             0.0149      0.001     25.612      0.000       0.014
```

```
0.016
Max_Temp           -0.0284      0.001    -27.096      0.000      -0.030
-0.026
Evaporation        -0.0036      0.005     -0.718      0.473      -0.013
0.006
Electricity        -0.0107      0.005     -2.150      0.032      -0.020
-0.001
Parameter1_Speed    0.0040      0.000     32.257      0.000       0.004
0.004
Parameter3_9am      0.0018      0.000     10.384      0.000       0.001
0.002
Parameter3_3pm     -0.0027      0.000    -15.193      0.000      -0.003
-0.002
Parameter4_9am      0.0081      0.000     69.026      0.000       0.008
0.008
Parameter4_3pm      0.0004      0.000      3.223      0.001       0.000
0.001
Parameter5_9am     -0.0271      0.001    -38.661      0.000      -0.028
-0.026
Parameter5_3pm      0.0210      0.001     29.938      0.000       0.020
0.022
Parameter6_9am      0.0142      0.005      2.834      0.005       0.004
0.024
Parameter6_3pm      0.0270      0.005      5.729      0.000       0.018
0.036
Parameter7_9am     -0.0010      0.001     -1.096      0.273      -0.003
0.001
Parameter7_3pm      0.0125      0.001     10.685      0.000       0.010
0.015
Estacion_O          0.0443      0.003     13.671      0.000       0.038
0.051
Estacion_P         -0.0025      0.004     -0.669      0.504      -0.010
0.005
Estacion_V          0.0212      0.004      4.953      0.000       0.013
0.030
Parametro1_Dir_N   -0.0181      0.004     -4.470      0.000      -0.026
-0.010
Parametro1_Dir_S   -0.0038      0.004     -1.026      0.305      -0.011
0.003
Parametro1_Dir_W    0.0007      0.004      0.160      0.873      -0.008
0.009
Parametro2_9am_N    0.0033      0.004      0.889      0.374      -0.004
0.011
Parametro2_9am_S    0.0326      0.004      9.224      0.000       0.026
0.039
Parametro2_9am_W    0.0350      0.004      8.622      0.000       0.027
0.043
Parametro2_3pm_N   -0.0013      0.004     -0.322      0.748      -0.009
```

						0.006
Parametro2_3pm_S	0.0056	0.004	1.537	0.124	-0.002	0.013
Parametro2_3pm_W	0.0108	0.004	2.496	0.013	0.002	0.019
loc_3	-0.0695	0.011	-6.316	0.000	-0.091	-0.048
loc_4	0.0337	0.013	2.577	0.010	0.008	0.059
loc_5	-0.0451	0.010	-4.589	0.000	-0.064	-0.026
loc_6	-0.2337	0.011	-21.063	0.000	-0.255	-0.212
loc_7	-0.1234	0.011	-11.253	0.000	-0.145	-0.102
loc_8	0.0399	0.010	3.910	0.000	0.020	0.060
loc_9	-0.0167	0.010	-1.621	0.105	-0.037	0.003
loc_10	-0.0660	0.010	-6.331	0.000	-0.086	-0.046
loc_11	-0.0644	0.012	-5.326	0.000	-0.088	-0.041
loc_12	-0.0067	0.010	-0.659	0.510	-0.026	0.013
loc_13	-0.1204	0.009	-12.704	0.000	-0.139	-0.102
loc_14	-0.0370	0.011	-3.455	0.001	-0.058	-0.016
loc_15	-0.0031	0.010	-0.327	0.744	-0.022	0.016
loc_16	-0.0981	0.010	-9.948	0.000	-0.117	-0.079
loc_17	-0.0345	0.017	-2.044	0.041	-0.068	-0.001
loc_18	-0.0925	0.011	-8.627	0.000	-0.114	-0.071
loc_19	-0.0791	0.011	-7.519	0.000	-0.100	-0.058
loc_20	-0.1386	0.010	-13.306	0.000	-0.159	-0.118
loc_21	-0.1488	0.011	-13.070	0.000	-0.171	-0.127
loc_22	-0.0258	0.011	-2.288	0.022	-0.048	-0.004
loc_23	-0.1037	0.010	-10.239	0.000	-0.124	-0.084
loc_26	-0.1745	0.012	-14.410	0.000	-0.198	

```
-0.151
loc_27                -0.0938      0.009    -10.000      0.000       -0.112
-0.075
loc_28                -0.1187      0.010    -11.942      0.000       -0.138
-0.099
loc_29                -0.1206      0.011    -11.057      0.000       -0.142
-0.099
loc_30                -0.0199      0.012     -1.697      0.090       -0.043
0.003
loc_32                -0.0304      0.010     -2.963      0.003       -0.050
-0.010
loc_33                -0.0246      0.011     -2.318      0.020       -0.045
-0.004
loc_34                -0.1242      0.010    -12.686      0.000       -0.143
-0.105
loc_35                -0.0478      0.010     -4.590      0.000       -0.068
-0.027
loc_36                -0.1621      0.010    -15.445      0.000       -0.183
-0.141
loc_38                -0.0537      0.010     -5.324      0.000       -0.073
-0.034
loc_39                -0.0553      0.010     -5.287      0.000       -0.076
-0.035
loc_40                -0.0552      0.011     -5.116      0.000       -0.076
-0.034
loc_41                -0.0215      0.010     -2.176      0.030       -0.041
-0.002
loc_42                 0.0156      0.016      0.971      0.332       -0.016
0.047
loc_43                -0.0678      0.011     -6.103      0.000       -0.090
-0.046
loc_44                -0.0699      0.009     -7.438      0.000       -0.088
-0.051
loc_45                -0.1397      0.010    -13.513      0.000       -0.160
-0.119
loc_46                -0.0343      0.011     -3.220      0.001       -0.055
-0.013
loc_47                -0.0214      0.010     -2.205      0.027       -0.041
-0.002
loc_48                -0.1323      0.010    -13.044      0.000       -0.152
-0.112
loc_49                -0.1654      0.013    -12.876      0.000       -0.191
-0.140
================================================================================
====
```

4. Logit

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

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

```
Optimization terminated successfully.
         Current function value: 0.351250
         Iterations 8
                           Logit Regression Results
==============================================================================
Dep. Variable:                  Fallo   No. Observations:               112925
Model:                          Logit   Df Residuals:                   112854
Method:                           MLE   Df Model:                           70
Date:                Thu, 24 Apr 2025   Pseudo R-squ.:                  0.3407
Time:                        22:54:05   Log-Likelihood:                -39665.
converged:                       True   LL-Null:                       -60159.
Covariance Type:                  HC0   LLR p-value:                     0.000
================================================================================
====
                       coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----
const               48.6035      1.772     27.426      0.000      45.130
52.077
Min_Temp             0.1371      0.005     26.080      0.000       0.127
0.147
Max_Temp            -0.2601      0.010    -27.218      0.000      -0.279
-0.241
Evaporation         -0.0310      0.044     -0.698      0.485      -0.118
0.056
Electricity         -0.0986      0.044     -2.219      0.027      -0.186
-0.011
Parameter1_Speed     0.0352      0.001     31.429      0.000       0.033
0.037
Parameter3_9am       0.0155      0.002      9.767      0.000       0.012
0.019
Parameter3_3pm      -0.0231      0.002    -14.397      0.000      -0.026
-0.020
Parameter4_9am       0.0741      0.001     65.391      0.000       0.072
0.076
Parameter4_3pm       0.0025      0.001      2.307      0.021       0.000
0.005
Parameter5_9am      -0.2417      0.006    -37.504      0.000      -0.254
-0.229
Parameter5_3pm       0.1885      0.006     29.508      0.000       0.176
```

```
0.201
Parameter6_9am       0.1424      0.046      3.091      0.002       0.052
0.233
Parameter6_3pm       0.2531      0.043      5.851      0.000       0.168
0.338
Parameter7_9am      -0.0099      0.008     -1.221      0.222      -0.026
0.006
Parameter7_3pm       0.1086      0.010     10.386      0.000       0.088
0.129
Estacion_O           0.3771      0.030     12.763      0.000       0.319
0.435
Estacion_P          -0.0207      0.033     -0.622      0.534      -0.086
0.045
Estacion_V           0.1751      0.038      4.552      0.000       0.100
0.251
Parametro1_Dir_N    -0.1773      0.036     -4.885      0.000      -0.249
-0.106
Parametro1_Dir_S    -0.0517      0.033     -1.565      0.118      -0.116
0.013
Parametro1_Dir_W    -0.0121      0.039     -0.313      0.754      -0.088
0.064
Parametro2_9am_N     0.0270      0.034      0.796      0.426      -0.039
0.093
Parametro2_9am_S     0.2940      0.032      9.253      0.000       0.232
0.356
Parametro2_9am_W     0.3180      0.036      8.729      0.000       0.247
0.389
Parametro2_3pm_N    -0.0104      0.035     -0.294      0.769      -0.080
0.059
Parametro2_3pm_S     0.0355      0.033      1.091      0.275      -0.028
0.099
Parametro2_3pm_W     0.0848      0.039      2.198      0.028       0.009
0.160
loc_3               -0.7087      0.100     -7.120      0.000      -0.904
-0.514
loc_4                0.2260      0.119      1.900      0.057      -0.007
0.459
loc_5               -0.3752      0.089     -4.213      0.000      -0.550
-0.201
loc_6               -2.2007      0.100    -21.947      0.000      -2.397
-2.004
loc_7               -1.1739      0.099    -11.801      0.000      -1.369
-0.979
loc_8                0.4044      0.092      4.377      0.000       0.223
0.585
loc_9               -0.0715      0.093     -0.769      0.442      -0.254
0.111
loc_10              -0.6328      0.095     -6.634      0.000      -0.820
```

						-0.446
loc_11	-0.6832	0.110	-6.210	0.000	-0.899	-0.468
loc_12	-0.0351	0.092	-0.383	0.702	-0.215	0.145
loc_13	-1.0976	0.085	-12.911	0.000	-1.264	-0.931
loc_14	-0.2466	0.097	-2.545	0.011	-0.437	-0.057
loc_15	0.0403	0.087	0.465	0.642	-0.130	0.210
loc_16	-0.9468	0.091	-10.434	0.000	-1.125	-0.769
loc_17	-0.2018	0.152	-1.328	0.184	-0.500	0.096
loc_18	-0.8390	0.097	-8.676	0.000	-1.029	-0.649
loc_19	-0.7447	0.096	-7.778	0.000	-0.932	-0.557
loc_20	-1.2845	0.095	-13.518	0.000	-1.471	-1.098
loc_21	-1.4018	0.104	-13.508	0.000	-1.605	-1.198
loc_22	-0.2645	0.105	-2.514	0.012	-0.471	-0.058
loc_23	-0.9686	0.092	-10.555	0.000	-1.148	-0.789
loc_26	-1.5795	0.110	-14.412	0.000	-1.794	-1.365
loc_27	-0.8312	0.085	-9.776	0.000	-0.998	-0.665
loc_28	-1.0656	0.090	-11.782	0.000	-1.243	-0.888
loc_29	-1.1637	0.099	-11.700	0.000	-1.359	-0.969
loc_30	-0.2052	0.105	-1.948	0.051	-0.412	0.001
loc_32	-0.2677	0.092	-2.894	0.004	-0.449	-0.086
loc_33	-0.2215	0.096	-2.317	0.020	-0.409	-0.034
loc_34	-1.1395	0.089	-12.795	0.000	-1.314	-0.965
loc_35	-0.4109	0.094	-4.350	0.000	-0.596	-0.226
loc_36	-1.5101	0.096	-15.722	0.000	-1.698	-1.322
loc_38	-0.4746	0.091	-5.198	0.000	-0.654	

```
-0.296
loc_39              -0.5194      0.096     -5.393      0.000      -0.708
-0.331
loc_40              -0.4027      0.098     -4.117      0.000      -0.594
-0.211
loc_41              -0.1965      0.090     -2.195      0.028      -0.372
-0.021
loc_42               0.0721      0.147      0.489      0.625      -0.217
0.361
loc_43              -0.7069      0.101     -6.966      0.000      -0.906
-0.508
loc_44              -0.6303      0.085     -7.418      0.000      -0.797
-0.464
loc_45              -1.3060      0.094    -13.894      0.000      -1.490
-1.122
loc_46              -0.3202      0.097     -3.309      0.001      -0.510
-0.131
loc_47              -0.1858      0.088     -2.122      0.034      -0.357
-0.014
loc_48              -1.1975      0.093    -12.853      0.000      -1.380
-1.015
loc_49              -1.5994      0.116    -13.834      0.000      -1.826
-1.373
================================================================================
====
        Logit Marginal Effects
=====================================
Dep. Variable:                  Fallo
Method:                          dydx
At:                           overall
================================================================================
====
                      dy/dx    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----
Min_Temp             0.0152      0.001     26.320      0.000       0.014
0.016
Max_Temp            -0.0288      0.001    -27.575      0.000      -0.031
-0.027
Evaporation         -0.0034      0.005     -0.698      0.485      -0.013
0.006
Electricity         -0.0109      0.005     -2.219      0.027      -0.021
-0.001
Parameter1_Speed     0.0039      0.000     32.141      0.000       0.004
0.004
Parameter3_9am       0.0017      0.000      9.783      0.000       0.001
0.002
```

Parameter3_3pm	-0.0026	0.000	-14.454	0.000	-0.003	-0.002
Parameter4_9am	0.0082	0.000	71.073	0.000	0.008	0.008
Parameter4_3pm	0.0003	0.000	2.308	0.021	4.25e-05	0.001
Parameter5_9am	-0.0268	0.001	-38.526	0.000	-0.028	-0.025
Parameter5_3pm	0.0209	0.001	30.022	0.000	0.020	0.022
Parameter6_9am	0.0158	0.005	3.092	0.002	0.006	0.026
Parameter6_3pm	0.0281	0.005	5.855	0.000	0.019	0.037
Parameter7_9am	-0.0011	0.001	-1.222	0.222	-0.003	0.001
Parameter7_3pm	0.0120	0.001	10.419	0.000	0.010	0.014
Estacion_O	0.0418	0.003	12.784	0.000	0.035	0.048
Estacion_P	-0.0023	0.004	-0.622	0.534	-0.010	0.005
Estacion_V	0.0194	0.004	4.556	0.000	0.011	0.028
Parametro1_Dir_N	-0.0197	0.004	-4.888	0.000	-0.028	-0.012
Parametro1_Dir_S	-0.0057	0.004	-1.565	0.118	-0.013	0.001
Parametro1_Dir_W	-0.0013	0.004	-0.313	0.754	-0.010	0.007
Parametro2_9am_N	0.0030	0.004	0.796	0.426	-0.004	0.010
Parametro2_9am_S	0.0326	0.004	9.256	0.000	0.026	0.040
Parametro2_9am_W	0.0353	0.004	8.734	0.000	0.027	0.043
Parametro2_3pm_N	-0.0012	0.004	-0.294	0.769	-0.009	0.007
Parametro2_3pm_S	0.0039	0.004	1.091	0.275	-0.003	0.011
Parametro2_3pm_W	0.0094	0.004	2.198	0.028	0.001	0.018
loc_3	-0.0786	0.011	-7.134	0.000	-0.100	-0.057
loc_4	0.0251	0.013	1.900	0.057	-0.001	0.051
loc_5	-0.0416	0.010	-4.216	0.000	-0.061	-0.022

loc_6	-0.2440	0.011	-22.218	0.000	-0.266	-0.222
loc_7	-0.1302	0.011	-11.845	0.000	-0.152	-0.109
loc_8	0.0448	0.010	4.378	0.000	0.025	0.065
loc_9	-0.0079	0.010	-0.769	0.442	-0.028	0.012
loc_10	-0.0702	0.011	-6.644	0.000	-0.091	-0.049
loc_11	-0.0758	0.012	-6.219	0.000	-0.100	-0.052
loc_12	-0.0039	0.010	-0.383	0.702	-0.024	0.016
loc_13	-0.1217	0.009	-12.958	0.000	-0.140	-0.103
loc_14	-0.0273	0.011	-2.546	0.011	-0.048	-0.006
loc_15	0.0045	0.010	0.465	0.642	-0.014	0.023
loc_16	-0.1050	0.010	-10.478	0.000	-0.125	-0.085
loc_17	-0.0224	0.017	-1.328	0.184	-0.055	0.011
loc_18	-0.0930	0.011	-8.694	0.000	-0.114	-0.072
loc_19	-0.0826	0.011	-7.795	0.000	-0.103	-0.062
loc_20	-0.1424	0.010	-13.586	0.000	-0.163	-0.122
loc_21	-0.1554	0.011	-13.568	0.000	-0.178	-0.133
loc_22	-0.0293	0.012	-2.515	0.012	-0.052	-0.006
loc_23	-0.1074	0.010	-10.585	0.000	-0.127	-0.088
loc_26	-0.1751	0.012	-14.473	0.000	-0.199	-0.151
loc_27	-0.0922	0.009	-9.800	0.000	-0.111	-0.074
loc_28	-0.1182	0.010	-11.825	0.000	-0.138	-0.099
loc_29	-0.1290	0.011	-11.740	0.000	-0.151	-0.107
loc_30	-0.0228	0.012	-1.948	0.051	-0.046	0.000
loc_32	-0.0297	0.010	-2.895	0.004	-0.050	-0.010

```
loc_33              -0.0246      0.011     -2.318      0.020      -0.045
-0.004
loc_34              -0.1263      0.010    -12.842      0.000      -0.146
-0.107
loc_35              -0.0456      0.010     -4.353      0.000      -0.066
-0.025
loc_36              -0.1674      0.011    -15.835      0.000      -0.188
-0.147
loc_38              -0.0526      0.010     -5.203      0.000      -0.072
-0.033
loc_39              -0.0576      0.011     -5.399      0.000      -0.078
-0.037
loc_40              -0.0446      0.011     -4.117      0.000      -0.066
-0.023
loc_41              -0.0218      0.010     -2.195      0.028      -0.041
-0.002
loc_42               0.0080      0.016      0.489      0.625      -0.024
0.040
loc_43              -0.0784      0.011     -6.979      0.000      -0.100
-0.056
loc_44              -0.0699      0.009     -7.429      0.000      -0.088
-0.051
loc_45              -0.1448      0.010    -13.968      0.000      -0.165
-0.124
loc_46              -0.0355      0.011     -3.311      0.001      -0.057
-0.014
loc_47              -0.0206      0.010     -2.123      0.034      -0.040
-0.002
loc_48              -0.1328      0.010    -12.905      0.000      -0.153
-0.113
loc_49              -0.1773      0.013    -13.897      0.000      -0.202
-0.152
================================================================================
====
Odds Ratios
                  Odds Ratio        5%       95%
Min_Temp            1.135137  1.158763  1.146889
Max_Temp            0.756629  0.785515  0.770937
Evaporation         0.888487  1.057742  0.969428
Electricity         0.830560  0.988578  0.906131
Parameter1_Speed    1.033514  1.038055  1.035782
Parameter3_9am      1.012473  1.018794  1.015629
Parameter3_3pm      0.974068  0.980221  0.977140
Parameter4_9am      1.074554  1.079339  1.076944
Parameter4_3pm      1.000383  1.004713  1.002546
Parameter5_9am      0.775399  0.795241  0.785258
Parameter5_3pm      1.192437  1.222677  1.207462
Parameter6_9am      1.053471  1.261970  1.153017
```

```
Parameter6_3pm      1.183299  1.401928  1.287982
Parameter7_9am      0.974548  1.006004  0.990151
Parameter7_3pm      1.092101  1.137792  1.114712
Estacion_0          1.376000  1.544964  1.458037
```

5. Al ser la variable dependiente una variable dicotómica, los modelos de Probit y Logit entregan resultados más adecuados. Además, los valores de log-likelihood y de R2 son más relevantes en estos modelos por lo que se ajustan mejor al problema. Max Temp y Electricity destacan como variables significativas.

6. Agrupacion por mes-año:

[22]:
```python
df1 = df

#volvemos a limpiar los NaN como se hizo antes:
df1['Electricity'] = df1['Electricity'].notnull().astype(int)
df1['Evaporation'] = df1['Evaporation'].notnull().astype(int)
df1['Parameter6_3pm'] = df1['Parameter6_3pm'].notnull().astype(int)
df1['Parameter6_9am'] = df1['Parameter6_9am'].notnull().astype(int)
df1 =df1.dropna()
df1
#volvemos a mapear fallo y transformamos fecha a datetime
df1['Fallo'] = df['Failure_today'].map({'Yes': 1, 'No': 0})
df1['Date'] = pd.to_datetime(df1['Date'])

df1['año_mes'] = df1['Date'].dt.to_period('M')

df2 = df1.groupby(['Location', 'año_mes']).agg({
    'Min_Temp' : 'mean',
    'Max_Temp':'mean',
    'Leakage': 'mean',
    'Evaporation': lambda x: (x == 1).sum() ,
    'Electricity':lambda x: (x == 1).sum() ,
    'Parameter1_Speed': 'mean',
    'Parameter3_9am' :'mean',
    'Parameter3_3pm' :'mean',
    'Parameter4_9am' :'mean',
    'Parameter4_3pm' :'mean',
    'Parameter5_9am' :'mean',
    'Parameter5_3pm' :'mean',
    'Parameter6_9am' :'mean',
    'Parameter6_3pm' :'mean',
    'Parameter7_9am' :'mean',
    'Parameter7_3pm' :'mean',
    'Fallo': lambda x: (x == 1).sum()
}).reset_index()
```

```
C:\Users\edins\AppData\Local\Temp\ipykernel_12548\3691603292.py:11:
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df1['Fallo'] = df['Failure_today'].map({'Yes': 1, 'No': 0})
C:\Users\edins\AppData\Local\Temp\ipykernel_12548\3691603292.py:12:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df1['Date'] = pd.to_datetime(df1['Date'])
C:\Users\edins\AppData\Local\Temp\ipykernel_12548\3691603292.py:14:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df1['año_mes'] = df1['Date'].dt.to_period('M')
```

[23]:

```
df2
```

[23]:

```
      Location  año_mes   Min_Temp   Max_Temp   Leakage  Evaporation  \
0            1  2008-07   7.000000  14.550000  3.530000           20
1            1  2008-08   5.936842  14.600000  4.242105           19
2            1  2008-09   9.461538  20.234615  0.615385           26
3            1  2008-10  12.383333  25.045833  0.200000           24
4            1  2008-11  14.210714  24.642857  0.492857           28
…          …       …        …          …         …          …
4132        49  2017-02  19.546429  34.232143  0.000000           28
4133        49  2017-03  18.745161  33.732258  0.000000           31
4134        49  2017-04  13.572414  24.796552  1.403448           29
4135        49  2017-05   9.277419  20.938710  0.341935           31
4136        49  2017-06   5.952174  18.747826  0.008696           23

      Electricity  Parameter1_Speed  Parameter3_9am  Parameter3_3pm  \
0              20         39.450000       11.950000       16.250000
1              19         36.105263        9.315789       15.631579
2              26         39.846154       14.730769       17.807692
3              24         37.291667       11.875000       17.458333
4              28         42.142857       12.607143       18.678571
…             …              …              …              …
4132           28         46.464286       23.178571       20.928571
4133           31         43.612903       20.387097       18.419355
```

```
4134            29          35.758621         18.586207         17.172414
4135            31          33.580645         14.741935         17.290323
4136            23          28.000000         11.391304         13.391304

      Parameter4_9am  Parameter4_3pm  Parameter5_9am  Parameter5_3pm  \
0          73.300000       58.800000     1020.545000     1019.020000
1          74.947368       58.000000     1026.763158     1025.205263
2          51.423077       37.115385     1019.834615     1017.573077
3          45.125000       29.958333     1020.979167     1018.900000
4          52.678571       35.428571     1012.867857     1011.203571
…                …               …               …               …
4132       49.964286       24.285714     1013.971429     1011.989286
4133       49.387097       21.806452     1014.780645     1012.367742
4134       56.034483       38.379310     1022.668966     1019.606897
4135       65.258065       37.677419     1022.958065     1020.187097
4136       66.565217       36.608696     1029.586957     1026.939130

      Parameter6_9am  Parameter6_3pm  Parameter7_9am  Parameter7_3pm  Fallo
0                1.0             1.0       10.795000       13.615000     10
1                1.0             1.0        9.973684       13.484211     10
2                1.0             1.0       15.188462       19.211538      4
3                1.0             1.0       17.933333       23.941667      2
4                1.0             1.0       18.492857       23.110714      5
…                …               …               …               …    …
4132             1.0             1.0       23.560714       32.203571      0
4133             1.0             1.0       22.170968       32.074194      0
4134             1.0             1.0       18.596552       23.644828      4
4135             1.0             1.0       13.806452       20.267742      1
4136             1.0             1.0       10.556522       18.052174      0

[4137 rows x 19 columns]
```

Poisson:

[24]:
```
y = df2['Fallo']
X1 = df2.drop(['Fallo','año_mes','Leakage','Parameter6_9am','Parameter6_3pm'],␣
 ↪axis=1)
poisson=sm.GLM(y,X1,family=sm.families.Poisson()).fit()
print(poisson.summary())
```

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                  Fallo   No. Observations:                 4137
Model:                            GLM   Df Residuals:                     4124
Model Family:                 Poisson   Df Model:                           12
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -9318.7
Date:                jue, 24 abr. 2025   Deviance:                       4898.1
```

```
Time:                        23:36:44   Pearson chi2:                 4.37e+03
No. Iterations:                     5   Pseudo R-squ. (CS):             0.8490
Covariance Type:            nonrobust
================================================================================
====
                     coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----
Location          -0.0029      0.000     -6.299      0.000      -0.004
-0.002
Min_Temp          -0.0077      0.007     -1.135      0.256      -0.021
0.006
Max_Temp          -0.0918      0.021     -4.463      0.000      -0.132
-0.051
Evaporation        0.0199      0.001     21.497      0.000       0.018
0.022
Electricity        0.0199      0.001     21.497      0.000       0.018
0.022
Parameter1_Speed   0.0560      0.002     29.262      0.000       0.052
0.060
Parameter3_9am    -0.0054      0.003     -1.950      0.051      -0.011
2.87e-05
Parameter3_3pm    -0.0650      0.003    -22.735      0.000      -0.071
-0.059
Parameter4_9am     0.0383      0.002     19.632      0.000       0.035
0.042
Parameter4_3pm    -0.0022      0.002     -0.940      0.347      -0.007
0.002
Parameter5_9am    -0.0307      0.012     -2.531      0.011      -0.055
-0.007
Parameter5_3pm     0.0286      0.012      2.349      0.019       0.005
0.052
Parameter7_9am     0.1787      0.012     15.476      0.000       0.156
0.201
Parameter7_3pm    -0.0634      0.023     -2.704      0.007      -0.109
-0.017
================================================================================
====
```

7. Sobredispersión y valor alpha:

[25]:
```python
aux=((y-poisson.mu)**2-poisson.mu)/poisson.mu
auxr=sm.OLS(aux,poisson.mu).fit()
print(auxr.summary())
```

```
                            OLS Regression Results
================================================================================
=======
```

```
Dep. Variable:                  Fallo   R-squared (uncentered):
0.002
Model:                            OLS   Adj. R-squared (uncentered):
0.001
Method:                 Least Squares   F-statistic:
6.489
Date:               jue, 24 abr. 2025   Prob (F-statistic):
0.0109
Time:                        23:36:49   Log-Likelihood:
-7419.9
No. Observations:                4137   AIC:
1.484e+04
Df Residuals:                    4136   BIC:
1.485e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0081      0.003      2.547      0.011       0.002       0.014
==============================================================================
Omnibus:                     3000.887   Durbin-Watson:                   1.782
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            70722.467
Skew:                           3.207   Prob(JB):                         0.00
Kurtosis:                      22.213   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.
```

[26]:
```python
alfa = np.exp(auxr.params[0])
print(alfa)
```

```
1.0081757831083749
```

```
C:\Users\edins\AppData\Local\Temp\ipykernel_12548\2180082010.py:1:
FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a
future version, integer keys will always be treated as labels (consistent with
DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
  alfa = np.exp(auxr.params[0])
```

Dado que Alfa $= 1.008$, existe una sobredispersión moderada

8. Binomial Negativa

[27]:
```
negbin=sm.GLM(y,X1,family=sm.families.NegativeBinomial(alpha=alfa)).fit()
print(negbin.summary())
```

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                  Fallo   No. Observations:                 4137
Model:                            GLM   Df Residuals:                     4124
Model Family:        NegativeBinomial   Df Model:                           12
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -11350.
Date:               jue, 24 abr. 2025   Deviance:                       1154.4
Time:                        23:37:31   Pearson chi2:                     804.
No. Iterations:                     8   Pseudo R-squ. (CS):             0.2584
Covariance Type:            nonrobust
================================================================================
====
                      coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----
Location           -0.0030      0.001     -2.482      0.013      -0.005
-0.001
Min_Temp            0.0029      0.017      0.171      0.864      -0.030
0.036
Max_Temp           -0.0442      0.055     -0.801      0.423      -0.152
0.064
Evaporation         0.0195      0.002      8.603      0.000       0.015
0.024
Electricity         0.0195      0.002      8.603      0.000       0.015
0.024
Parameter1_Speed    0.0639      0.006     11.605      0.000       0.053
0.075
Parameter3_9am      0.0006      0.007      0.088      0.930      -0.013
0.015
Parameter3_3pm     -0.0798      0.008    -10.242      0.000      -0.095
-0.065
Parameter4_9am      0.0460      0.005      9.064      0.000       0.036
0.056
Parameter4_3pm     -0.0093      0.006     -1.484      0.138      -0.022
0.003
Parameter5_9am     -0.0530      0.032     -1.635      0.102      -0.117
0.011
Parameter5_3pm      0.0508      0.033      1.562      0.118      -0.013
0.115
Parameter7_9am      0.2028      0.030      6.775      0.000       0.144
0.262
Parameter7_3pm     -0.1461      0.062     -2.338      0.019      -0.268
-0.024
```

==
====

9. Ambos modelos entregan coeficientes similares, sin embargo, se podría interpretar que el modelo de binomial negativa entrega un mejor resultado al tener en cuenta el alpha y la sobre dispersion. Muchas variables resultaron ser robustas para el modelo, tales como Evaporation, Electricity o Parameter 1 pues muestran una alta significancia sobre la variable de fallo.