Tarea_1_Alan_Wilson

April 29, 2025

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.

R: Se eliminaron variables con alta cantidad de datos faltantes (P6_9am, P6_3pm, Leakage), en el caso de leakage, también se considero eliminarla dada su alta cantidad de outliers. Se reemplazaron NaN por 0 en Evaporation y Electricity, generando columnas binarias que identificar los NaN reemplazados. Se generó una variable binaria a partir de Failure_today, además, se descartaron variables altamente correlacionadas, identificadas a partir de la matriz de correlación. Posteriormente, las variables categóricas como P1_Dir, P2_9am, P2_3pm, y Location se transformaron a dummies.

[248]:
```
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 matplotlib.patches as patches
import warnings
import math
warnings.filterwarnings("ignore")
%matplotlib inline
```

[249]:
```
df = pd.read_csv('machine_failure_data.csv', delimiter=",", decimal=',')
df['Date'] = pd.to_datetime(df['Date'], errors='coerce')
df['Fecha'] = df['Date']
df['Date'] = df['Date'].dt.year.astype(int)
df = df[df['Fecha'].dt.year >= 2009]
df.columns = [col.replace('Parameter', 'P') if 'Parameter' in col else col for␣
 ↪col in df.columns]
```

[250]:
```
#Eliminamos las variables p6, que tienen muchos NaN, al igual que las␣
 ↪filtraciones que tienen muchos outliers
df = df.drop(['P6_9am', 'P6_3pm', 'Leakage'], axis=1)
```

[251]:
```
# Dejamos solo direcciones cardinal principales (N, S, E, W)
direccion = ['P1_Dir', 'P2_9am', 'P2_3pm']
for i in direccion:
    df[i] = df[i].astype(str).str[0]

df1 = df.copy()
```

[252]:
```
# Crear columnas indicadoras donde había NaN (antes de reemplazarlos)
df['Evaporation_NaN'] = df['Evaporation'].isna().astype(int)
df['Electricity_NaN'] = df['Electricity'].isna().astype(int)

# Reemplazar NaN por 0 en las columnas originales
df['Evaporation'] = df['Evaporation'].fillna(0)
df['Electricity'] = df['Electricity'].fillna(0)
```

LIMPIEZA Y CONVERSIÓN DE DATOS

[253]:
```
# ELIMINAMOS LOS NAN DE LA BASE DE DATOS
df.dropna(inplace=True)
```

[254]:
```
# CONVERTIMOS LA COLUMNA A UNA BINARIA DE 1's y 0's
df['Failure_today'] = df['Failure_today'].map({'No': 0, 'Yes': 1})

cols_a_convertir = [
    'Min_Temp', 'Max_Temp', 'Evaporation', 'Electricity',
    'P7_9am', 'P7_3pm', 'P5_9am', 'P5_3pm', 'Failure_today'
]

for col in cols_a_convertir:
    df[col] = pd.to_numeric(df[col], errors='coerce')
```

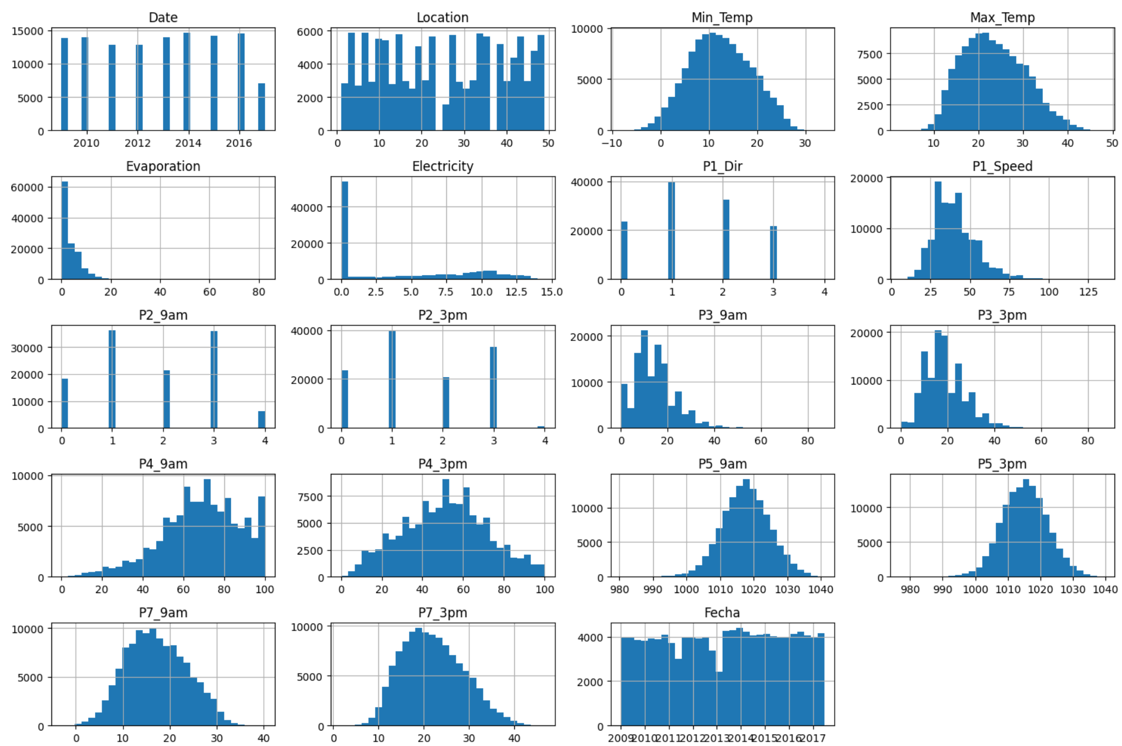
CONVERTIMOS LAS VARIABLES CATEGÓRICAS A NÚMEROS PARA GRÁFICAR (POSTERIMENTE SE PASARÁN A DUMMI)

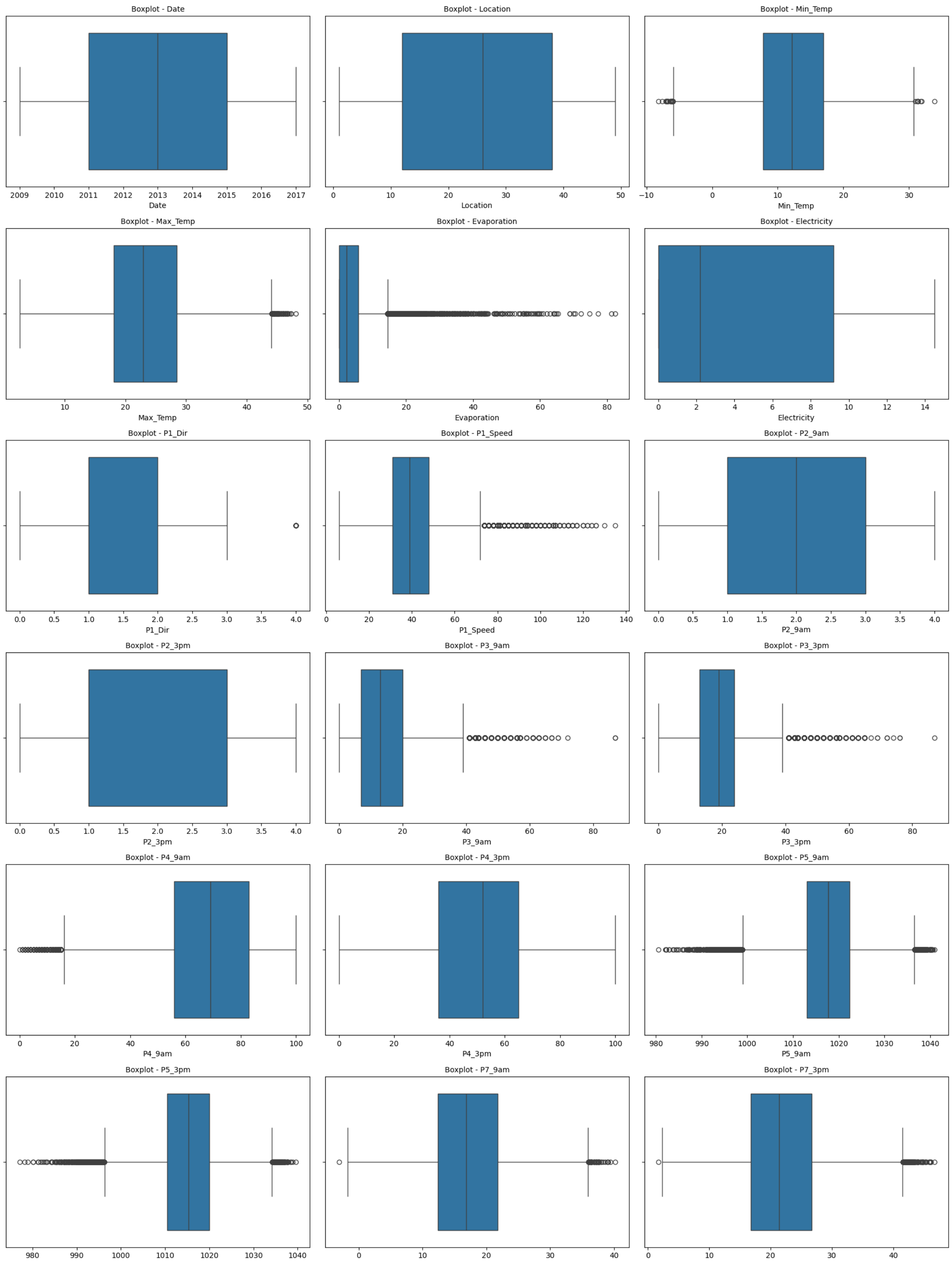
[255]:
```
for col in ['P1_Dir', 'P2_9am', 'P2_3pm']:
    df[col], _ = pd.factorize(df[col])
```

REVISIÓN DE DISTRIBUCIÓN Y OUTLIERS (HISTOGRAMAS Y BOXPLOT)

[256]:
```
# Excluir columnas binarias (con solo 2 valores únicos)
df_non_binary = df.loc[:, df.nunique() > 2]

df_non_binary.hist(bins=30, figsize=(15, 10))
plt.tight_layout()
plt.show()

numeric_cols = df_non_binary.select_dtypes(include=[np.number]).columns
```

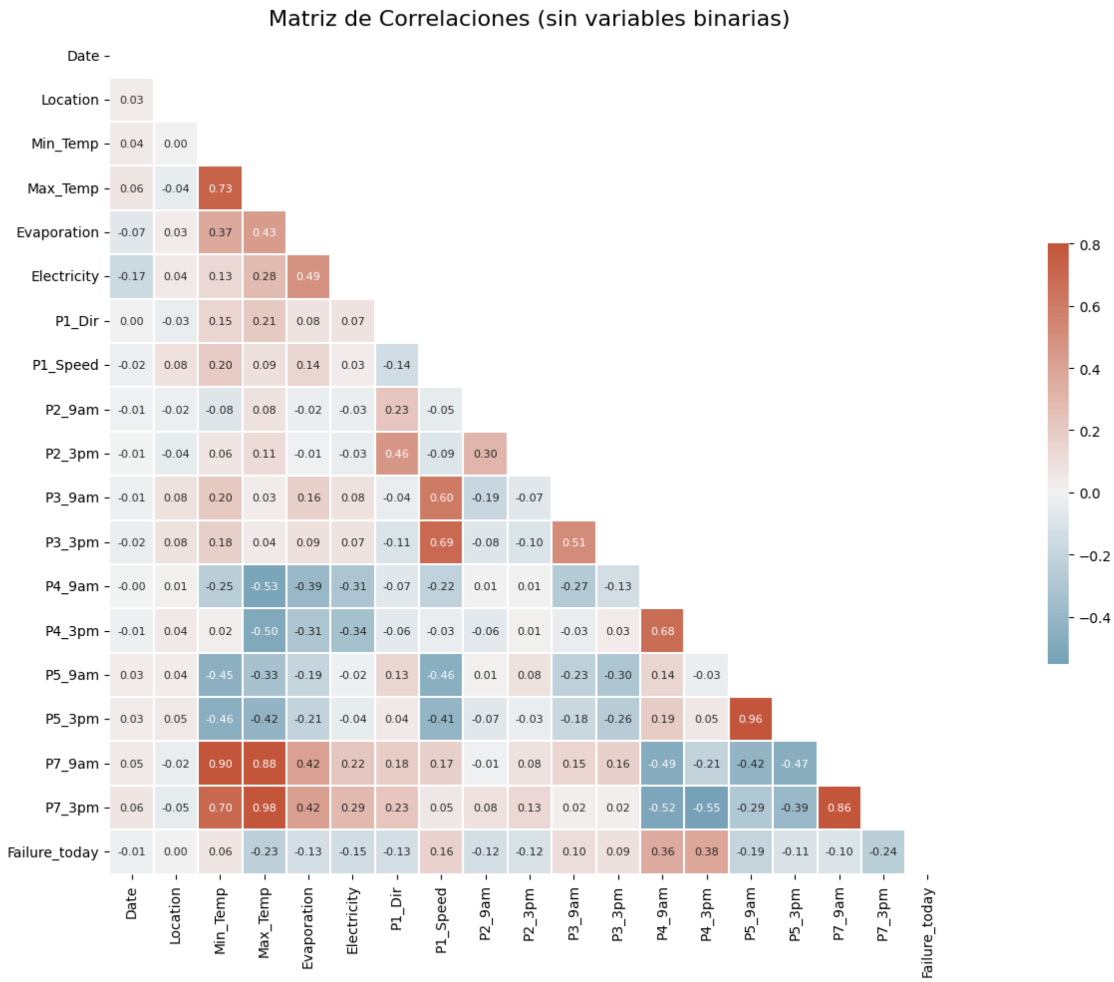
```python
cols_per_row = 3
total = len(numeric_cols)
rows = math.ceil(total / cols_per_row)

fig, axes = plt.subplots(rows, cols_per_row, figsize=(6 * cols_per_row, 4 * rows))
axes = axes.flatten()

for i, col in enumerate(numeric_cols):
    sns.boxplot(x=df_non_binary[col], ax=axes[i])
    axes[i].set_title(f'Boxplot - {col}', fontsize=10)

for j in range(i + 1, len(axes)):
    fig.delaxes(axes[j])

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





MATRIZ DE CORRELACIONES INCLUYENDO TODAS LAS VARIABBLES (SIN DUMMIES, CON CATEGÓRICAS)

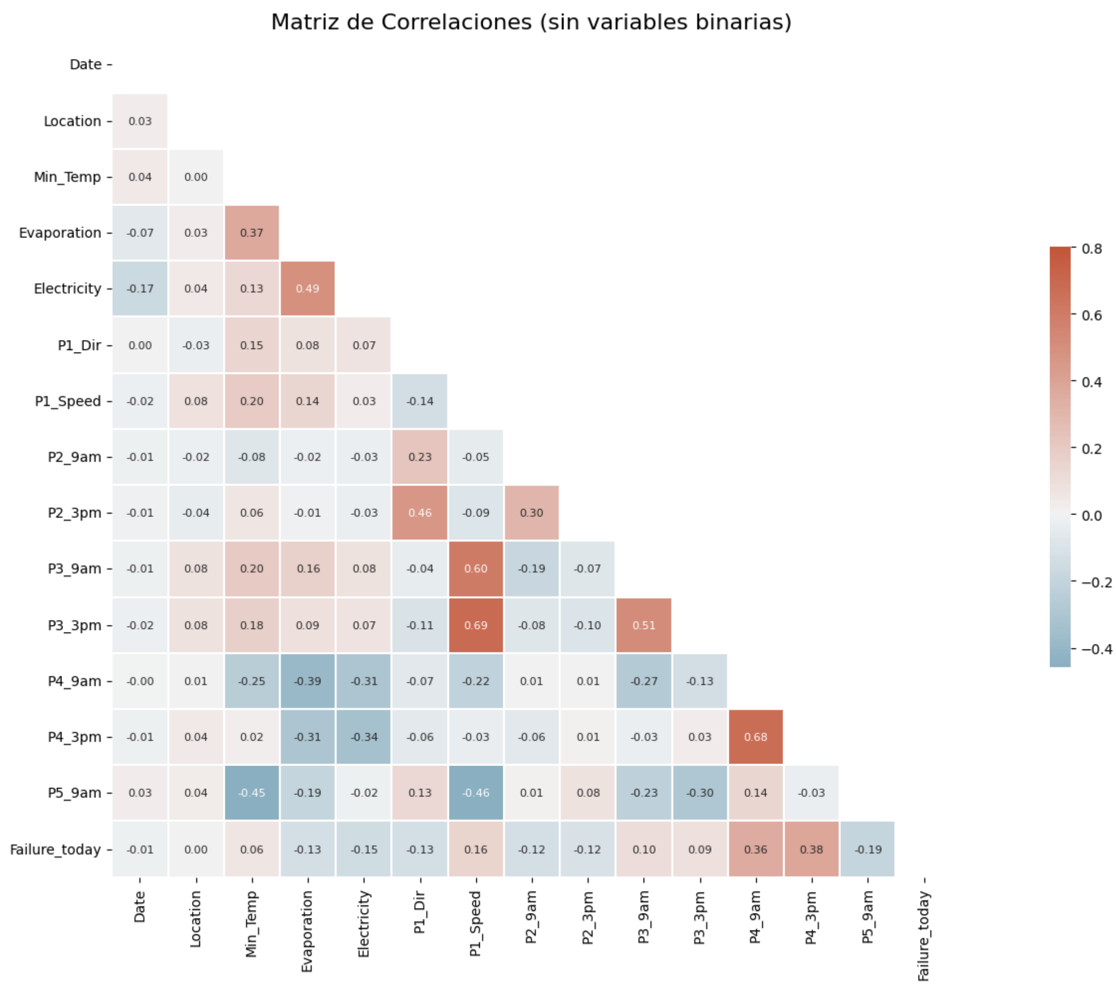
[257]:
```
cols_non_binary = [col for col in df.columns if (df[col].nunique() > 2) or (col␣
↪== 'Failure_today')]
df_corr = df[[col for col in cols_non_binary if col != 'Fecha']]

corr = df_corr.corr()

mask = np.triu(np.ones_like(corr, dtype=bool))

f, ax = plt.subplots(figsize=(20, 10))
cmap = sns.diverging_palette(230, 20, as_cmap=True)

sns.heatmap(corr, mask=mask, cmap=cmap, vmax=.8, center=0,
            square=True, linewidths=.2, cbar_kws={"shrink": .5},
            annot=True, fmt=".2f", annot_kws={"size": 8},
            xticklabels=True, yticklabels=True)

labels = corr.columns
failure_index = list(labels).index('Failure_today')


plt.title("Matriz de Correlaciones (sin variables binarias)", fontsize=16)
plt.tight_layout()
plt.show()
```



[258]:
```python
# Eliminamos las variables que tienen mucha correlación entre ellas, dejando una, en este caso "Min_Temp"
df = df.drop(['Max_Temp', 'P7_9am', 'P7_3pm'], axis=1)

# En este caso ambas variables miden lo mismo en horas distintas y tienen una correlación superior a 0.9, eliminamos una
df = df.drop(['P5_3pm'], axis=1)
```

[259]:
```python
cols_non_binary = [col for col in df.columns if (df[col].nunique() > 2) or (col == 'Failure_today')]
df_corr = df[[col for col in cols_non_binary if col != 'Fecha']]

corr = df_corr.corr()

mask = np.triu(np.ones_like(corr, dtype=bool))
```

```python
f, ax = plt.subplots(figsize=(20, 10))
cmap = sns.diverging_palette(230, 20, as_cmap=True)

sns.heatmap(corr, mask=mask, cmap=cmap, vmax=.8, center=0,
            square=True, linewidths=.2, cbar_kws={"shrink": .5},
            annot=True, fmt=".2f", annot_kws={"size": 8},
            xticklabels=True, yticklabels=True)

labels = corr.columns
failure_index = list(labels).index('Failure_today')

plt.title("Matriz de Correlaciones (sin variables binarias)", fontsize=16)
plt.tight_layout()
plt.show()
```



CONVERTIMOS LAS VARIABLES CATEGÓRICAS EN COLUMNAS DUMMI PARA PODER HACER LA REGRESIÓN

[260]:
```
df2 = df.copy()
df = df.drop(['Fecha'], axis=1)

df = pd.get_dummies(df, columns=['Location', 'P1_Dir', 'P2_9am', 'P2_3pm',␣
 ↪'Date'], drop_first=True)

# Reemplazar True por 1 y False por 0 en todo el DataFrame
df = df.replace({True: 1, False: 0})
```

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.

R: El modelo muestra que variables como Min_Temp, P1_Speed, P3_9am, P4_9am y P4_3pm tienen una asociación positiva con la probabilidad de fallo, mientras que Evaporation, Electricity, P3_3pm y P5_9am presentan efectos negativos. Las variables indicadoras de datos faltantes (Evaporation_NaN, Electricity_NaN) también muestran asociaciones negativas significativas, lo cual sugiere que la ausencia de estas mediciones se relaciona con menor probabilidad de fallo.

En el caso de las ubicaciones, la mayoría de los coeficientes son negativos, lo que indica que varias zonas presentan menor probabilidad de falla respecto del baseline.

REGRESIÓN MCO PARA LA VARIABLE "Failure_Today"

[261]:
```
y=df['Failure_today']
X=df.drop(['Failure_today'], 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:          Failure_today   R-squared:                       0.275
Model:                            OLS   Adj. R-squared:                  0.275
Method:                 Least Squares   F-statistic:                     614.6
Date:               jue, 24 abr. 2025   Prob (F-statistic):               0.00
Time:                        22:42:09   Log-Likelihood:                -44586.
No. Observations:              117793   AIC:                         8.932e+04
Df Residuals:                  117718   BIC:                         9.005e+04
Df Model:                          74
Covariance Type:                  HC0
================================================================================
===
                      coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
---
const               8.4277      0.219     38.563      0.000       7.999
8.856
```

Min_Temp	0.0029	0.000	11.725	0.000	0.002	0.003
Evaporation	-0.0062	0.000	-14.063	0.000	-0.007	-0.005
Electricity	-0.0060	0.000	-13.314	0.000	-0.007	-0.005
P1_Speed	0.0050	0.000	35.243	0.000	0.005	0.005
P3_9am	0.0049	0.000	28.263	0.000	0.005	0.005
P3_3pm	-0.0034	0.000	-18.435	0.000	-0.004	-0.003
P4_9am	0.0077	9.09e-05	85.052	0.000	0.008	0.008
P4_3pm	0.0029	9.07e-05	32.071	0.000	0.003	0.003
P5_9am	-0.0087	0.000	-40.941	0.000	-0.009	-0.008
Evaporation_NaN	-0.0253	0.006	-4.568	0.000	-0.036	-0.014
Electricity_NaN	-0.0396	0.006	-6.153	0.000	-0.052	-0.027
Location_3	-0.1212	0.009	-13.068	0.000	-0.139	-0.103
Location_4	0.0654	0.008	8.016	0.000	0.049	0.081
Location_5	-0.1628	0.010	-16.768	0.000	-0.182	-0.144
Location_6	-0.2455	0.010	-24.149	0.000	-0.265	-0.226
Location_7	-0.1535	0.009	-16.770	0.000	-0.171	-0.136
Location_8	-0.0317	0.010	-3.292	0.001	-0.051	-0.013
Location_9	-0.1176	0.010	-11.659	0.000	-0.137	-0.098
Location_10	-0.1411	0.009	-15.050	0.000	-0.159	-0.123
Location_11	-0.0504	0.009	-5.680	0.000	-0.068	-0.033
Location_12	-0.0792	0.010	-7.839	0.000	-0.099	-0.059
Location_13	-0.1717	0.010	-16.956	0.000	-0.192	-0.152
Location_14	-0.1321	0.010	-13.613	0.000	-0.151	-0.113
Location_15	-0.1379	0.010	-13.744	0.000	-0.158	-0.118

Location_16	-0.1389	0.010	-13.817	0.000	-0.159	-0.119
Location_17	-0.1729	0.014	-11.985	0.000	-0.201	-0.145
Location_18	-0.1461	0.011	-13.053	0.000	-0.168	-0.124
Location_19	-0.1179	0.011	-10.545	0.000	-0.140	-0.096
Location_20	-0.1875	0.010	-19.100	0.000	-0.207	-0.168
Location_21	-0.1282	0.009	-14.926	0.000	-0.145	-0.111
Location_22	-0.0879	0.009	-10.105	0.000	-0.105	-0.071
Location_23	-0.1324	0.010	-13.374	0.000	-0.152	-0.113
Location_26	-0.2214	0.011	-20.669	0.000	-0.242	-0.200
Location_27	-0.1807	0.010	-17.528	0.000	-0.201	-0.160
Location_28	-0.1511	0.010	-14.647	0.000	-0.171	-0.131
Location_29	-0.1051	0.009	-11.528	0.000	-0.123	-0.087
Location_30	-0.0821	0.010	-8.072	0.000	-0.102	-0.062
Location_32	-0.0481	0.009	-5.343	0.000	-0.066	-0.030
Location_33	-0.0659	0.009	-7.275	0.000	-0.084	-0.048
Location_34	-0.1318	0.010	-12.585	0.000	-0.152	-0.111
Location_35	-0.1446	0.010	-15.199	0.000	-0.163	-0.126
Location_36	-0.2466	0.010	-25.100	0.000	-0.266	-0.227
Location_38	-0.1191	0.011	-11.042	0.000	-0.140	-0.098
Location_39	-0.1048	0.010	-10.576	0.000	-0.124	-0.085
Location_40	-0.1414	0.009	-15.426	0.000	-0.159	-0.123
Location_41	-0.1062	0.009	-11.223	0.000	-0.125	-0.088
Location_42	-0.0157	0.009	-1.669	0.095	-0.034	0.003
Location_43	-0.0909	0.009	-10.001	0.000	-0.109	-0.073

Location_44	-0.1269	0.011	-11.866	0.000	-0.148	-0.106
Location_45	-0.1626	0.010	-16.622	0.000	-0.182	-0.143
Location_46	-0.1118	0.011	-10.428	0.000	-0.133	-0.091
Location_47	-0.0988	0.011	-9.385	0.000	-0.119	-0.078
Location_48	-0.1832	0.010	-18.075	0.000	-0.203	-0.163
Location_49	-0.0987	0.008	-11.641	0.000	-0.115	-0.082
P1_Dir_1	-0.0059	0.004	-1.550	0.121	-0.013	0.002
P1_Dir_2	-0.0313	0.004	-8.274	0.000	-0.039	-0.024
P1_Dir_3	-0.0141	0.004	-3.384	0.001	-0.022	-0.006
P1_Dir_4	0.1205	0.123	0.981	0.326	-0.120	0.361
P2_9am_1	-0.0256	0.004	-6.352	0.000	-0.034	-0.018
P2_9am_2	-0.0752	0.004	-17.936	0.000	-0.083	-0.067
P2_9am_3	-0.0745	0.004	-19.287	0.000	-0.082	-0.067
P2_9am_4	-0.0869	0.006	-15.028	0.000	-0.098	-0.076
P2_3pm_1	-0.0082	0.004	-2.198	0.028	-0.016	-0.001
P2_3pm_2	-0.0391	0.004	-9.425	0.000	-0.047	-0.031
P2_3pm_3	-0.0533	0.004	-14.541	0.000	-0.061	-0.046
P2_3pm_4	-0.0959	0.015	-6.199	0.000	-0.126	-0.066
Date_2010	0.0037	0.004	0.856	0.392	-0.005	0.012
Date_2011	-0.0048	0.004	-1.097	0.273	-0.013	0.004
Date_2012	0.0025	0.004	0.577	0.564	-0.006	0.011
Date_2013	0.0020	0.004	0.478	0.633	-0.006	0.010
Date_2014	-0.0010	0.004	-0.248	0.804	-0.009	0.007
Date_2015	0.0074	0.004	1.718	0.086	-0.001	0.016

```
Date_2016           -0.0005      0.004     -0.105      0.917      -0.009
0.008
Date_2017           -0.0265      0.005     -4.837      0.000      -0.037
-0.016
==============================================================================
Omnibus:                    10304.819   Durbin-Watson:                   1.801
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            13284.897
Skew:                           0.822   Prob(JB):                         0.00
Kurtosis:                       2.922   Cond. No.                     2.04e+05
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 2.04e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

0.1 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.

R: El modelo probit refuerza los resultados del MCO y entrega una mejor representación de la variable dependiente. Las variables Min_Temp, P1_Speed, P3_9am, P4_9am y P4_3pm aumentan significativamente la probabilidad de que ocurra un fallo, mientras que Evaporation, P3_3pm, P5_9am y ciertas direcciones del viento se asocian negativamente. Se observaron diferencias de magnitud respecto a OLS, con efectos marginales más grandes en torno al promedio. Por ejemplo, un cambio unitario en P4_9am implica un aumento marginal importante en la probabilidad de fallo.

[262]:
```python
X1 = X.copy()
X1 = sm.add_constant(X1)
model = sm.Probit(y, X1)
probit_model = model.fit(cov_type='HC0')
print(probit_model.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.367778
         Iterations 7
                          Probit Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:               117793
Model:                         Probit   Df Residuals:                   117718
Method:                           MLE   Df Model:                           74
Date:                jue, 24 abr. 2025   Pseudo R-squ.:                  0.3037
Time:                        22:42:12   Log-Likelihood:                -43322.
converged:                       True   LL-Null:                       -62216.
Covariance Type:                  HC0   LLR p-value:                     0.000
================================================================================
===
                       coef    std err          z      P>|z|      [0.025
```

```
0.975]
--------------------------------------------------------------------------------
---
const              29.4562      0.941     31.306      0.000      27.612
31.300
Min_Temp            0.0156      0.001     11.764      0.000       0.013
0.018
Evaporation        -0.0451      0.004    -11.787      0.000      -0.053
-0.038
Electricity        -0.0101      0.002     -4.751      0.000      -0.014
-0.006
P1_Speed            0.0188      0.001     31.939      0.000       0.018
0.020
P3_9am              0.0217      0.001     26.530      0.000       0.020
0.023
P3_3pm             -0.0104      0.001    -12.378      0.000      -0.012
-0.009
P4_9am              0.0403      0.000     82.052      0.000       0.039
0.041
P4_3pm              0.0116      0.000     31.070      0.000       0.011
0.012
P5_9am             -0.0332      0.001    -36.385      0.000      -0.035
-0.031
Evaporation_NaN    -0.2076      0.028     -7.291      0.000      -0.263
-0.152
Electricity_NaN     0.0025      0.029      0.085      0.932      -0.054
0.059
Location_3         -0.5003      0.045    -11.105      0.000      -0.589
-0.412
Location_4         -0.0979      0.060     -1.633      0.102      -0.215
0.020
Location_5         -0.6621      0.046    -14.411      0.000      -0.752
-0.572
Location_6         -1.1728      0.048    -24.603      0.000      -1.266
-1.079
Location_7         -0.6964      0.045    -15.315      0.000      -0.785
-0.607
Location_8          0.0287      0.045      0.645      0.519      -0.059
0.116
Location_9         -0.4810      0.043    -11.143      0.000      -0.566
-0.396
Location_10        -0.5281      0.046    -11.520      0.000      -0.618
-0.438
Location_11        -0.3918      0.053     -7.380      0.000      -0.496
-0.288
Location_12        -0.3051      0.044     -6.977      0.000      -0.391
-0.219
Location_13        -0.7798      0.044    -17.787      0.000      -0.866
```

						-0.694
Location_14	-0.5151	0.046	-11.304	0.000	-0.604	-0.426
Location_15	-0.5820	0.047	-12.484	0.000	-0.673	-0.491
Location_16	-0.3561	0.046	-7.666	0.000	-0.447	-0.265
Location_17	-0.7811	0.078	-9.977	0.000	-0.935	-0.628
Location_18	-0.5449	0.051	-10.773	0.000	-0.644	-0.446
Location_19	-0.3207	0.049	-6.611	0.000	-0.416	-0.226
Location_20	-0.7013	0.046	-15.248	0.000	-0.791	-0.611
Location_21	-0.8204	0.051	-16.161	0.000	-0.920	-0.721
Location_22	-0.3284	0.050	-6.596	0.000	-0.426	-0.231
Location_23	-0.5834	0.044	-13.387	0.000	-0.669	-0.498
Location_26	-1.1031	0.057	-19.256	0.000	-1.215	-0.991
Location_27	-0.7557	0.046	-16.323	0.000	-0.846	-0.665
Location_28	-0.5685	0.045	-12.721	0.000	-0.656	-0.481
Location_29	-0.6338	0.049	-12.976	0.000	-0.730	-0.538
Location_30	-0.3018	0.052	-5.798	0.000	-0.404	-0.200
Location_32	-0.1624	0.045	-3.648	0.000	-0.250	-0.075
Location_33	-0.2308	0.046	-5.033	0.000	-0.321	-0.141
Location_34	-0.5799	0.044	-13.249	0.000	-0.666	-0.494
Location_35	-0.6038	0.046	-13.063	0.000	-0.694	-0.513
Location_36	-0.9768	0.045	-21.501	0.000	-1.066	-0.888
Location_38	-0.3730	0.047	-7.878	0.000	-0.466	-0.280
Location_39	-0.3620	0.047	-7.775	0.000	-0.453	-0.271
Location_40	-0.5188	0.046	-11.228	0.000	-0.609	-0.428
Location_41	-0.3430	0.045	-7.568	0.000	-0.432	

						-0.254
Location_42	-0.3324	0.070	-4.722	0.000	-0.470	-0.194
Location_43	-0.4060	0.048	-8.467	0.000	-0.500	-0.312
Location_44	-0.5164	0.045	-11.488	0.000	-0.605	-0.428
Location_45	-0.6296	0.045	-13.968	0.000	-0.718	-0.541
Location_46	-0.3906	0.047	-8.305	0.000	-0.483	-0.298
Location_47	-0.3794	0.046	-8.299	0.000	-0.469	-0.290
Location_48	-0.7271	0.047	-15.513	0.000	-0.819	-0.635
Location_49	-0.8885	0.060	-14.925	0.000	-1.005	-0.772
P1_Dir_1	-0.0358	0.017	-2.135	0.033	-0.069	-0.003
P1_Dir_2	-0.1833	0.018	-10.429	0.000	-0.218	-0.149
P1_Dir_3	-0.0769	0.020	-3.759	0.000	-0.117	-0.037
P1_Dir_4	0.5224	0.467	1.120	0.263	-0.392	1.437
P2_9am_1	-0.0522	0.016	-3.254	0.001	-0.084	-0.021
P2_9am_2	-0.3650	0.020	-18.527	0.000	-0.404	-0.326
P2_9am_3	-0.3484	0.016	-21.712	0.000	-0.380	-0.317
P2_9am_4	-0.3978	0.027	-14.834	0.000	-0.450	-0.345
P2_3pm_1	-0.0237	0.017	-1.408	0.159	-0.057	0.009
P2_3pm_2	-0.1303	0.021	-6.351	0.000	-0.170	-0.090
P2_3pm_3	-0.2237	0.018	-12.739	0.000	-0.258	-0.189
P2_3pm_4	-0.3766	0.066	-5.672	0.000	-0.507	-0.246
Date_2010	0.0444	0.020	2.248	0.025	0.006	0.083
Date_2011	0.0037	0.020	0.183	0.855	-0.036	0.044
Date_2012	0.0220	0.021	1.070	0.285	-0.018	0.062
Date_2013	0.0120	0.021	0.583	0.560	-0.028	

```
0.052
Date_2014            0.0044      0.021      0.211      0.833      -0.036
0.045
Date_2015            0.0288      0.021      1.378      0.168      -0.012
0.070
Date_2016            0.0290      0.021      1.380      0.168      -0.012
0.070
Date_2017           -0.1200      0.026     -4.569      0.000      -0.171
-0.069
================================================================================
===
```

0.2 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.

R: Los coeficientes marginales del modelo logit son muy similares a los obtenidos en el modelo probit, tanto en signo como en magnitud, lo cual es esperable. Sin embargo, logit permite interpretar los resultados también en términos de razón de odds (odds ratio), lo que entrega una visión más completa del efecto relativo de cada variable sobre la probabilidad de fallo. El modelo mantiene la significancia de las variables clave y muestra un leve aumento en el pseudo R^2 respecto al modelo probit.

[263]:
```python
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: 0.366616
         Iterations 8
                           Logit Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:               117793
Model:                          Logit   Df Residuals:                   117718
Method:                           MLE   Df Model:                           74
Date:               jue, 24 abr. 2025   Pseudo R-squ.:                  0.3059
Time:                        22:42:14   Log-Likelihood:                -43185.
converged:                       True   LL-Null:                       -62216.
Covariance Type:                  HC0   LLR p-value:                     0.000
```

	coef	std err	z	P>|z|	[0.025	0.975]
const	51.4770	1.662	30.965	0.000	48.219	54.735
Min_Temp	0.0294	0.002	12.776	0.000	0.025	0.034
Evaporation	-0.1022	0.007	-15.428	0.000	-0.115	-0.089
Electricity	-0.0128	0.004	-3.414	0.001	-0.020	-0.005
P1_Speed	0.0331	0.001	31.850	0.000	0.031	0.035
P3_9am	0.0386	0.001	26.460	0.000	0.036	0.041
P3_3pm	-0.0175	0.001	-11.714	0.000	-0.020	-0.015
P4_9am	0.0731	0.001	84.119	0.000	0.071	0.075
P4_3pm	0.0200	0.001	30.659	0.000	0.019	0.021
P5_9am	-0.0582	0.002	-36.078	0.000	-0.061	-0.055
Evaporation_NaN	-0.4375	0.048	-9.034	0.000	-0.532	-0.343
Electricity_NaN	0.0352	0.051	0.694	0.488	-0.064	0.135
Location_3	-0.9024	0.080	-11.307	0.000	-1.059	-0.746
Location_4	-0.2214	0.108	-2.042	0.041	-0.434	-0.009
Location_5	-1.1916	0.082	-14.530	0.000	-1.352	-1.031
Location_6	-2.1377	0.084	-25.424	0.000	-2.303	-1.973
Location_7	-1.2382	0.081	-15.334	0.000	-1.396	-1.080
Location_8	0.0716	0.079	0.902	0.367	-0.084	0.227
Location_9	-0.8261	0.076	-10.832	0.000	-0.976	-0.677
Location_10	-0.9445	0.082	-11.544	0.000	-1.105	-0.784
Location_11	-0.7566	0.095	-7.943	0.000	-0.943	-0.570

Location_12	-0.5572	0.077	-7.198	0.000	-0.709	-0.405
Location_13	-1.4182	0.077	-18.301	0.000	-1.570	-1.266
Location_14	-0.8985	0.081	-11.026	0.000	-1.058	-0.739
Location_15	-1.0672	0.083	-12.902	0.000	-1.229	-0.905
Location_16	-0.6944	0.084	-8.309	0.000	-0.858	-0.531
Location_17	-1.3368	0.140	-9.527	0.000	-1.612	-1.062
Location_18	-0.9834	0.089	-11.038	0.000	-1.158	-0.809
Location_19	-0.5865	0.087	-6.772	0.000	-0.756	-0.417
Location_20	-1.2543	0.082	-15.233	0.000	-1.416	-1.093
Location_21	-1.4946	0.091	-16.375	0.000	-1.673	-1.316
Location_22	-0.6534	0.090	-7.300	0.000	-0.829	-0.478
Location_23	-1.0717	0.077	-13.833	0.000	-1.224	-0.920
Location_26	-1.9686	0.103	-19.170	0.000	-2.170	-1.767
Location_27	-1.3969	0.083	-16.819	0.000	-1.560	-1.234
Location_28	-1.0330	0.080	-12.909	0.000	-1.190	-0.876
Location_29	-1.1666	0.088	-13.284	0.000	-1.339	-0.994
Location_30	-0.5840	0.092	-6.327	0.000	-0.765	-0.403
Location_32	-0.2692	0.079	-3.400	0.001	-0.424	-0.114
Location_33	-0.4002	0.082	-4.879	0.000	-0.561	-0.239
Location_34	-1.0759	0.078	-13.736	0.000	-1.229	-0.922
Location_35	-1.0803	0.082	-13.101	0.000	-1.242	-0.919
Location_36	-1.7830	0.081	-22.028	0.000	-1.942	-1.624
Location_38	-0.6611	0.084	-7.864	0.000	-0.826	-0.496
Location_39	-0.6617	0.084	-7.834	0.000	-0.827	-0.496

Location_40	-0.8479	0.083	-10.275	0.000	-1.010	-0.686
Location_41	-0.6071	0.081	-7.537	0.000	-0.765	-0.449
Location_42	-0.6461	0.128	-5.036	0.000	-0.898	-0.395
Location_43	-0.7838	0.086	-9.161	0.000	-0.951	-0.616
Location_44	-0.9580	0.081	-11.891	0.000	-1.116	-0.800
Location_45	-1.1412	0.080	-14.220	0.000	-1.298	-0.984
Location_46	-0.7197	0.084	-8.567	0.000	-0.884	-0.555
Location_47	-0.7162	0.081	-8.842	0.000	-0.875	-0.557
Location_48	-1.3367	0.084	-15.826	0.000	-1.502	-1.171
Location_49	-1.6390	0.106	-15.475	0.000	-1.847	-1.431
P1_Dir_1	-0.0580	0.030	-1.958	0.050	-0.116	6.98e-05
P1_Dir_2	-0.3119	0.031	-10.079	0.000	-0.372	-0.251
P1_Dir_3	-0.1166	0.036	-3.237	0.001	-0.187	-0.046
P1_Dir_4	1.0584	0.923	1.146	0.252	-0.752	2.868
P2_9am_1	-0.0917	0.028	-3.246	0.001	-0.147	-0.036
P2_9am_2	-0.6484	0.035	-18.611	0.000	-0.717	-0.580
P2_9am_3	-0.6191	0.028	-21.873	0.000	-0.675	-0.564
P2_9am_4	-0.7005	0.047	-14.899	0.000	-0.793	-0.608
P2_3pm_1	-0.0404	0.030	-1.353	0.176	-0.099	0.018
P2_3pm_2	-0.2229	0.036	-6.152	0.000	-0.294	-0.152
P2_3pm_3	-0.3964	0.031	-12.789	0.000	-0.457	-0.336
P2_3pm_4	-0.6472	0.116	-5.589	0.000	-0.874	-0.420
Date_2010	0.0968	0.035	2.769	0.006	0.028	0.165
Date_2011	0.0171	0.036	0.473	0.636	-0.054	0.088

```
Date_2012            0.0538      0.036      1.475      0.140      -0.018
0.125
Date_2013            0.0212      0.036      0.580      0.562      -0.050
0.093
Date_2014            0.0151      0.037      0.413      0.680      -0.057
0.087
Date_2015            0.0594      0.037      1.598      0.110      -0.013
0.132
Date_2016            0.0636      0.037      1.704      0.088      -0.010
0.137
Date_2017           -0.2044      0.047     -4.380      0.000      -0.296
-0.113
================================================================================
===
        Logit Marginal Effects
=====================================
Dep. Variable:          Failure_today
Method:                          dydx
At:                           overall
================================================================================
===
                     dy/dx    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
---
Min_Temp             0.0034      0.000     12.813      0.000       0.003
0.004
Evaporation         -0.0118      0.001    -15.589      0.000      -0.013
-0.010
Electricity         -0.0015      0.000     -3.413      0.001      -0.002
-0.001
P1_Speed             0.0038      0.000     32.491      0.000       0.004
0.004
P3_9am               0.0045      0.000     26.761      0.000       0.004
0.005
P3_3pm              -0.0020      0.000    -11.744      0.000      -0.002
-0.002
P4_9am               0.0085   8.83e-05     95.833      0.000       0.008
0.009
P4_3pm               0.0023   7.48e-05     30.968      0.000       0.002
0.002
P5_9am              -0.0067      0.000    -36.857      0.000      -0.007
-0.006
Evaporation_NaN     -0.0506      0.006     -9.056      0.000      -0.062
-0.040
Electricity_NaN      0.0041      0.006      0.694      0.488      -0.007
0.016
Location_3          -0.1044      0.009    -11.334      0.000      -0.123
```

						-0.086
Location_4	-0.0256	0.013	-2.042	0.041	-0.050	-0.001
Location_5	-0.1379	0.009	-14.590	0.000	-0.156	-0.119
Location_6	-0.2474	0.010	-25.757	0.000	-0.266	-0.229
Location_7	-0.1433	0.009	-15.391	0.000	-0.162	-0.125
Location_8	0.0083	0.009	0.902	0.367	-0.010	0.026
Location_9	-0.0956	0.009	-10.848	0.000	-0.113	-0.078
Location_10	-0.1093	0.009	-11.577	0.000	-0.128	-0.091
Location_11	-0.0876	0.011	-7.953	0.000	-0.109	-0.066
Location_12	-0.0645	0.009	-7.206	0.000	-0.082	-0.047
Location_13	-0.1641	0.009	-18.430	0.000	-0.182	-0.147
Location_14	-0.1040	0.009	-11.047	0.000	-0.122	-0.086
Location_15	-0.1235	0.010	-12.946	0.000	-0.142	-0.105
Location_16	-0.0804	0.010	-8.337	0.000	-0.099	-0.061
Location_17	-0.1547	0.016	-9.534	0.000	-0.187	-0.123
Location_18	-0.1138	0.010	-11.070	0.000	-0.134	-0.094
Location_19	-0.0679	0.010	-6.784	0.000	-0.087	-0.048
Location_20	-0.1452	0.009	-15.314	0.000	-0.164	-0.127
Location_21	-0.1730	0.011	-16.451	0.000	-0.194	-0.152
Location_22	-0.0756	0.010	-7.311	0.000	-0.096	-0.055
Location_23	-0.1240	0.009	-13.895	0.000	-0.142	-0.107
Location_26	-0.2278	0.012	-19.285	0.000	-0.251	-0.205
Location_27	-0.1617	0.010	-16.931	0.000	-0.180	-0.143
Location_28	-0.1196	0.009	-12.963	0.000	-0.138	-0.101
Location_29	-0.1350	0.010	-13.328	0.000	-0.155	

						-0.115
Location_30	-0.0676	0.011	-6.336	0.000	-0.088	-0.047
Location_32	-0.0312	0.009	-3.401	0.001	-0.049	-0.013
Location_33	-0.0463	0.009	-4.880	0.000	-0.065	-0.028
Location_34	-0.1245	0.009	-13.804	0.000	-0.142	-0.107
Location_35	-0.1250	0.010	-13.141	0.000	-0.144	-0.106
Location_36	-0.2064	0.009	-22.288	0.000	-0.225	-0.188
Location_38	-0.0765	0.010	-7.877	0.000	-0.096	-0.057
Location_39	-0.0766	0.010	-7.847	0.000	-0.096	-0.057
Location_40	-0.0981	0.010	-10.281	0.000	-0.117	-0.079
Location_41	-0.0703	0.009	-7.547	0.000	-0.089	-0.052
Location_42	-0.0748	0.015	-5.037	0.000	-0.104	-0.046
Location_43	-0.0907	0.010	-9.183	0.000	-0.110	-0.071
Location_44	-0.1109	0.009	-11.937	0.000	-0.129	-0.093
Location_45	-0.1321	0.009	-14.291	0.000	-0.150	-0.114
Location_46	-0.0833	0.010	-8.586	0.000	-0.102	-0.064
Location_47	-0.0829	0.009	-8.863	0.000	-0.101	-0.065
Location_48	-0.1547	0.010	-15.921	0.000	-0.174	-0.136
Location_49	-0.1897	0.012	-15.533	0.000	-0.214	-0.166
P1_Dir_1	-0.0067	0.003	-1.958	0.050	-0.013	8.21e-06
P1_Dir_2	-0.0361	0.004	-10.082	0.000	-0.043	-0.029
P1_Dir_3	-0.0135	0.004	-3.237	0.001	-0.022	-0.005
P1_Dir_4	0.1225	0.107	1.146	0.252	-0.087	0.332
P2_9am_1	-0.0106	0.003	-3.248	0.001	-0.017	-0.004
P2_9am_2	-0.0750	0.004	-18.669	0.000	-0.083	

```
-0.067
P2_9am_3          -0.0716      0.003    -22.023      0.000      -0.078
-0.065
P2_9am_4          -0.0811      0.005    -14.947      0.000      -0.092
-0.070
P2_3pm_1          -0.0047      0.003     -1.353      0.176      -0.011
0.002
P2_3pm_2          -0.0258      0.004     -6.156      0.000      -0.034
-0.018
P2_3pm_3          -0.0459      0.004    -12.824      0.000      -0.053
-0.039
P2_3pm_4          -0.0749      0.013     -5.591      0.000      -0.101
-0.049
Date_2010          0.0112      0.004      2.770      0.006       0.003
0.019
Date_2011          0.0020      0.004      0.473      0.636      -0.006
0.010
Date_2012          0.0062      0.004      1.476      0.140      -0.002
0.014
Date_2013          0.0024      0.004      0.580      0.562      -0.006
0.011
Date_2014          0.0018      0.004      0.413      0.680      -0.007
0.010
Date_2015          0.0069      0.004      1.598      0.110      -0.002
0.015
Date_2016          0.0074      0.004      1.705      0.088      -0.001
0.016
Date_2017         -0.0237      0.005     -4.380      0.000      -0.034
-0.013
================================================================================
===
Odds Ratios
                 Odds Ratio        5%       95%
Min_Temp           1.025190  1.034475  1.029822
Evaporation        0.891168  0.914620  0.902818
Electricity        0.980059  0.994566  0.987286
P1_Speed           1.031586  1.035801  1.033691
P3_9am             1.036419  1.042368  1.039389
P3_3pm             0.979832  0.985571  0.982697
P4_9am             1.074040  1.077706  1.075872
P4_3pm             1.018899  1.021508  1.020203
P5_9am             0.940462  0.946430  0.943441
Evaporation_NaN    0.587214  0.709959  0.645677
Electricity_NaN    0.937759  1.144212  1.035854
Location_3         0.346859  0.474265  0.405590
Location_4         0.647949  0.991148  0.801382
Location_5         0.258624  0.356688  0.303724
Location_6         0.100008  0.139053  0.117926
```

```
Location_7        0.247473  0.339619  0.289908
```

0.3 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: En base a los resultados, el modelo MCO entrega una aproximación inicial, pero no modela adecuadamente una variable binaria, por lo que Probit o Logit son más apropiados. Ambos muestran resultados similares, pero el modelo Logit permite una mejor interpretación de los efectos, por lo que se considera más adecuado. Variables como Min_Temp, P1_Speed, P3_9am, P4_9am, P4_3pm y P5_9am resultaron ser robustas a la especificación.

0.3.1 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.

R: El modelo Poisson, aplicado a datos mensuales por ubicación, presenta buen ajuste (pseudo R^2 = 0.855). Se observa que Min_Temp, P1_Speed, P4_9am y P4_3pm están positivamente asociados al número de fallos reportados por sensores, lo que sugiere que temperaturas mínimas más altas y ciertas condiciones de viento aumentan la frecuencia esperada de fallas. En contraste, Evaporation, Electricity, P3_3pm y P5_9am se asocian negativamente, lo que podría indicar condiciones más estables o menor carga operacional.

[264]:
```python
# Mismos ajustes de la parte 1 (Sin indicadores)
df1['Failure_today'] = df1['Failure_today'].map({'No': 0, 'Yes': 1})

cols_a_convertir = [
    'Min_Temp', 'Evaporation', 'Electricity',
    'P7_9am', 'P7_3pm', 'P5_9am', 'P5_3pm', 'Failure_today'
]

for col in cols_a_convertir:
    df1[col] = pd.to_numeric(df1[col], errors='coerce')

for col in ['P1_Dir', 'P2_9am', 'P2_3pm']:
    df1[col], _ = pd.factorize(df1[col])

df1 = df1.drop(['Max_Temp', 'P7_9am', 'P7_3pm'], axis=1)

df1 = df1.drop(['P5_3pm'], axis=1)
df2 = df1.copy()
```

AGRUPAMOS POR MES Y AÑADIMOS LA CANTIDAD DE FALLOS POR MES

[265]:
```
# Convertir 'Fecha' a datetime y extraer año y mes
df2['Fecha'] = pd.to_datetime(df2['Fecha'])
df2['Year'] = df2['Fecha'].dt.year
df2['Month'] = df2['Fecha'].dt.month

categoricas = ['Location', 'P1_Dir', 'P2_9am', 'P2_3pm', 'Fecha']
df_numerico = df2.drop(columns=categoricas)

df_mensual = df2.groupby(['Year', 'Month', 'Location']).agg({
    col: 'mean' for col in df_numerico.columns if col != 'Failure_today'
})
fallos_mensuales = df2.groupby(['Year', 'Month', 'Location'])['Failure_today'].
 ↪sum().rename('Monthly_Failures')
df_final = df_mensual.join(fallos_mensuales)
```

[266]:
```
# Crear columnas indicadoras donde había NaN (antes de reemplazarlos)
df_final['Evaporation_NaN'] = df_final['Evaporation'].isna().astype(int)
df_final['Electricity_NaN'] = df_final['Electricity'].isna().astype(int)

# Reemplazar NaN por 0 en las columnas originales
df_final['Evaporation'] = df_final['Evaporation'].fillna(0)
df_final['Electricity'] = df_final['Electricity'].fillna(0)

df_final = df_final.dropna()
```

[267]:
```
y = df_final['Monthly_Failures']
X2 = df_final.drop(columns=['Monthly_Failures', 'Year', 'Month', 'Date'],␣
 ↪axis=1)

X2 = sm.add_constant(X2)
poisson_model_final=sm.GLM(y,X2,family=sm.families.Poisson()).fit()
print(poisson_model_final.summary())
```

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:       Monthly_Failures   No. Observations:                 4076
Model:                            GLM   Df Residuals:                     4064
Model Family:                 Poisson   Df Model:                           11
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -9541.0
Date:                jue, 24 abr. 2025   Deviance:                       5272.3
Time:                        22:42:44   Pearson chi2:                 4.91e+03
No. Iterations:                     5   Pseudo R-squ. (CS):             0.8550
Covariance Type:            nonrobust
================================================================================
===
                      coef    std err          z      P>|z|      [0.025
0.975]
```

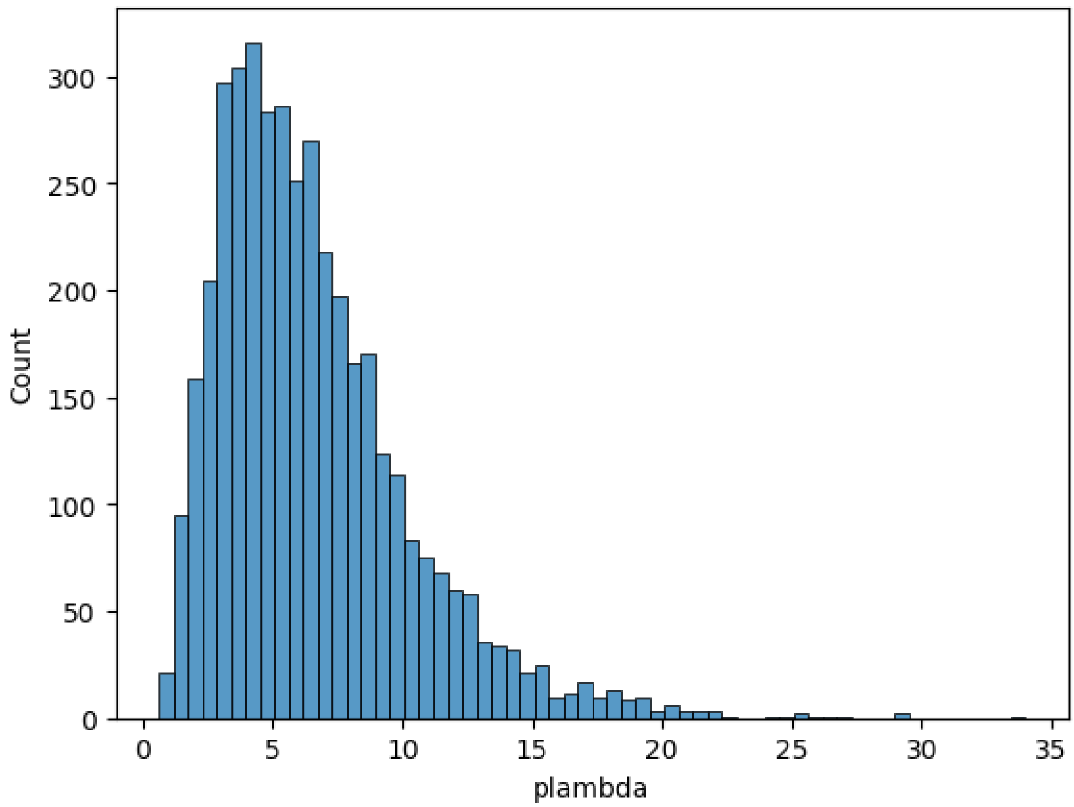
```
--------------------------------------------------------------------------------
---
const              22.4848      2.272      9.897      0.000      18.032
26.938
Min_Temp            0.0071      0.002      4.077      0.000       0.004
0.010
Evaporation        -0.0165      0.004     -3.699      0.000      -0.025
-0.008
Electricity        -0.0421      0.006     -6.759      0.000      -0.054
-0.030
P1_Speed            0.0394      0.002     19.620      0.000       0.035
0.043
P3_9am             -0.0027      0.003     -1.088      0.277      -0.008
0.002
P3_3pm             -0.0465      0.003    -17.342      0.000      -0.052
-0.041
P4_9am              0.0091      0.001      8.065      0.000       0.007
0.011
P4_3pm              0.0302      0.001     32.514      0.000       0.028
0.032
P5_9am             -0.0228      0.002    -10.444      0.000      -0.027
-0.019
Evaporation_NaN    -0.1155      0.033     -3.516      0.000      -0.180
-0.051
Electricity_NaN    -0.3688      0.051     -7.281      0.000      -0.468
-0.270
================================================================================
===
```

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

R: El análisis muestra que hay sobredispersión en los datos, ya que el valor de alpha estimado es positivo y estadísticamente significativo. Esto indica que la varianza supera a la media, lo que sugiere que un modelo Binomial Negativa puede ser más apropiado que el Poisson para explicar la cantidad de fallos mensuales.

[268]:
```
df_final['plambda'] = poisson_model_final.mu
sns.histplot(data=df_final, x="plambda")
```

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



[269]:
```
aux=((y-poisson_model_final.mu)**2 - poisson_model_final.mu) /␣
 ↪poisson_model_final.mu
auxr=sm.OLS(aux, poisson_model_final.mu).fit()
print(auxr.summary())
```

```
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:       Monthly_Failures   R-squared (uncentered):
0.001
Model:                            OLS   Adj. R-squared (uncentered):
0.001
Method:                 Least Squares   F-statistic:
5.401
Date:                jue, 24 abr. 2025   Prob (F-statistic):
0.0202
Time:                        22:42:45   Log-Likelihood:
-11541.
No. Observations:                4076   AIC:
2.308e+04
Df Residuals:                    4075   BIC:
```

```
2.309e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0198      0.009      2.324      0.020       0.003       0.037
==============================================================================
Omnibus:                    12885.645   Durbin-Watson:                   1.973
Prob(Omnibus):                  0.000   Jarque-Bera (JB):       1430803010.337
Skew:                          49.692   Prob(JB):                         0.00
Kurtosis:                    2903.840   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.
```

0.5 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.

R: El modelo Binomial Negativa, estimado con el valor de alpha obtenido en la regresión auxiliar, entrega un peor ajuste que el modelo Poisson (menor log-likelihood). Sin embargo, los coeficientes estimados son muy similares, lo que sugiere que las asociaciones identificadas se mantienen.

Variables como P1_Speed, P3_3pm, P4_3pm, P5_9am y Electricity continúan mostrando efectos significativos, manteniéndose robustas en ambas especificaciones. La interpretación de los coeficientes sigue siendo en términos del log del número esperado de fallos, por lo que no cambia respecto al modelo Poisson.

[270]:
```
negbin=sm.GLM(y,X2,family=sm.families.NegativeBinomial(alpha= np.exp(0.0197))).
  ↪fit()
print(negbin.summary())
```

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:       Monthly_Failures   No. Observations:                 4076
Model:                            GLM   Df Residuals:                     4064
Model Family:        NegativeBinomial   Df Model:                           11
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -11475.
Date:               jue, 24 abr. 2025   Deviance:                       1177.1
Time:                        22:42:45   Pearson chi2:                     959.
No. Iterations:                     8   Pseudo R-squ. (CS):             0.2466
Covariance Type:            nonrobust
================================================================================
```

```
===
                        coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------
---
const             24.2265      6.776      3.575      0.000      10.946
37.507
Min_Temp           0.0050      0.005      0.988      0.323      -0.005
0.015
Evaporation       -0.0120      0.010     -1.157      0.247      -0.032
0.008
Electricity       -0.0703      0.016     -4.296      0.000      -0.102
-0.038
P1_Speed           0.0471      0.006      7.937      0.000       0.035
0.059
P3_9am            -0.0012      0.007     -0.174      0.862      -0.014
0.012
P3_3pm            -0.0588      0.007     -7.939      0.000      -0.073
-0.044
P4_9am             0.0078      0.003      2.665      0.008       0.002
0.014
P4_3pm             0.0339      0.002     13.576      0.000       0.029
0.039
P5_9am            -0.0245      0.007     -3.760      0.000      -0.037
-0.012
Evaporation_NaN   -0.1071      0.085     -1.255      0.210      -0.274
0.060
Electricity_NaN   -0.5699      0.142     -4.009      0.000      -0.849
-0.291
==============================================================================
===
```

0.6 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?

R: El modelo Poisson (pregunta 6) presentó buen ajuste y permitió identificar relaciones significativas entre varias variables y el número de fallos. Sin embargo, en la pregunta 7 se evidenció una leve sobre-dispersión, lo que motivó el uso del modelo Binomial Negativa en la pregunta 8.

Aunque el Binomial Negativa entregó un peor ajuste (menor log-likelihood), los coeficientes fueron muy similares a los del modelo Poisson, lo que indica estabilidad en los resultados. Dado que la sobre-dispersión es leve, el modelo Poisson es más parsimonioso y, por tanto, más conveniente en este caso.

Las variables P1_Speed, P3_3pm, P4_3pm, P5_9am y Electricity fueron significativas y consistentes en todas las especificaciones, por lo que se consideran robustas.