Tarea1_Meza_Núñez

May 5, 2025

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

import warnings
warnings.filterwarnings("ignore")

%matplotlib inline


{"tags": ["hide_input"]}
```

[1]: `{'tags': ['hide_input']}`

0.0.1 Descripción general de la data

[2]:
```python
df = pd.read_csv('../../data/machine_failure_data.csv')
df.describe()
{"tags": ["hide_input"]}
```

[2]: `{'tags': ['hide_input']}`

1 Pregunta 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.

Cambiamos los "Yes" y "No" de la columa "Failure_today" por unos y ceros. También simplificamos las direcciones en solo 4

[3]:
```
for i in range(len(df)):
    if df.loc[i,"Failure_today"]=="Yes":
        df.loc[i,"Failure_today"]=1
    elif df.loc[i,"Failure_today"]=="No":
        df.loc[i,"Failure_today"]=0

    #Simplificación de direcciones
    if str(df.loc[i,"Parameter1_Dir"]).startswith("N"):
        df.loc[i,"Parameter1_Dir"]="N"
    elif str(df.loc[i,"Parameter1_Dir"]).startswith("S"):
        df.loc[i,"Parameter1_Dir"]="S"
    elif str(df.loc[i,"Parameter1_Dir"]).startswith("E"):
        df.loc[i,"Parameter1_Dir"]="E"
    elif str(df.loc[i,"Parameter1_Dir"]).startswith("W"):
        df.loc[i,"Parameter1_Dir"]="W"

    if str(df.loc[i,"Parameter2_9am"]).startswith("N"):
        df.loc[i,"Parameter2_9am"]="N"
    elif str(df.loc[i,"Parameter2_9am"]).startswith("S"):
        df.loc[i,"Parameter2_9am"]="S"
    elif str(df.loc[i,"Parameter2_9am"]).startswith("E"):
        df.loc[i,"Parameter2_9am"]="E"
    elif str(df.loc[i,"Parameter2_9am"]).startswith("W"):
        df.loc[i,"Parameter2_9am"]="W"

    if str(df.loc[i,"Parameter2_3pm"]).startswith("N"):
        df.loc[i,"Parameter2_3pm"]="N"
    elif str(df.loc[i,"Parameter2_3pm"]).startswith("S"):
        df.loc[i,"Parameter2_3pm"]="S"
    elif str(df.loc[i,"Parameter2_3pm"]).startswith("E"):
        df.loc[i,"Parameter2_3pm"]="E"
    elif str(df.loc[i,"Parameter2_3pm"]).startswith("W"):
        df.loc[i,"Parameter2_3pm"]="W"


{"tags": ["hide_input"]}
```

[3]:
```
{'tags': ['hide_input']}
```

Las filas que no tengan un valor en la variable Failure_today deben ser eliminadas. Hay muchas columnas con un gran porcentaje de valores nulos, eliminaremos las que tengan mas de un 30% de valores nulos. Luego de eliminar estas columnas, eliminaremos las filas que tengan valores nulos

[4]:
```
eliminar=[]
nulos=df.isnull().sum()
for i in range(len(nulos)):
    if nulos[i]/len(df)>0.3:
        eliminar.append(df.columns[i])
```

```
for i in eliminar:
    print(f'Se ha eliminado {i}')
    df.drop(i,axis=1,inplace=True)

c=len(df)
df.dropna(inplace=True)
c=c-len(df)
print(f'Se han eliminado {c} filas por tener un valor nulo')

df['Failure_today'] = df['Failure_today'].astype(int)
df
{"tags": ["hide_input"]}
```

```
Se ha eliminado Evaporation
Se ha eliminado Electricity
Se ha eliminado Parameter6_9am
Se ha eliminado Parameter6_3pm
Se han eliminado 29268 filas por tener un valor nulo
```

[4]: {'tags': ['hide_input']}

Transformaremos la columna "Date" en varias columnas que nos permitan analizar el tiempo de mejor manera. También eliminaremos los datos del 2008 por su inconsistencia.

[5]:
```
df['Date'] = pd.to_datetime(df['Date'], format='%m/%d/%Y')
df['Day'] = df['Date'].dt.day
df['Month'] = df['Date'].dt.month
df['Year'] = df['Date'].dt.year
df4=df["Date"]
df.drop("Date",axis=1, inplace=True)
df=df[df["Year"]>2008]
df

{"tags": ["hide_input"]}
```

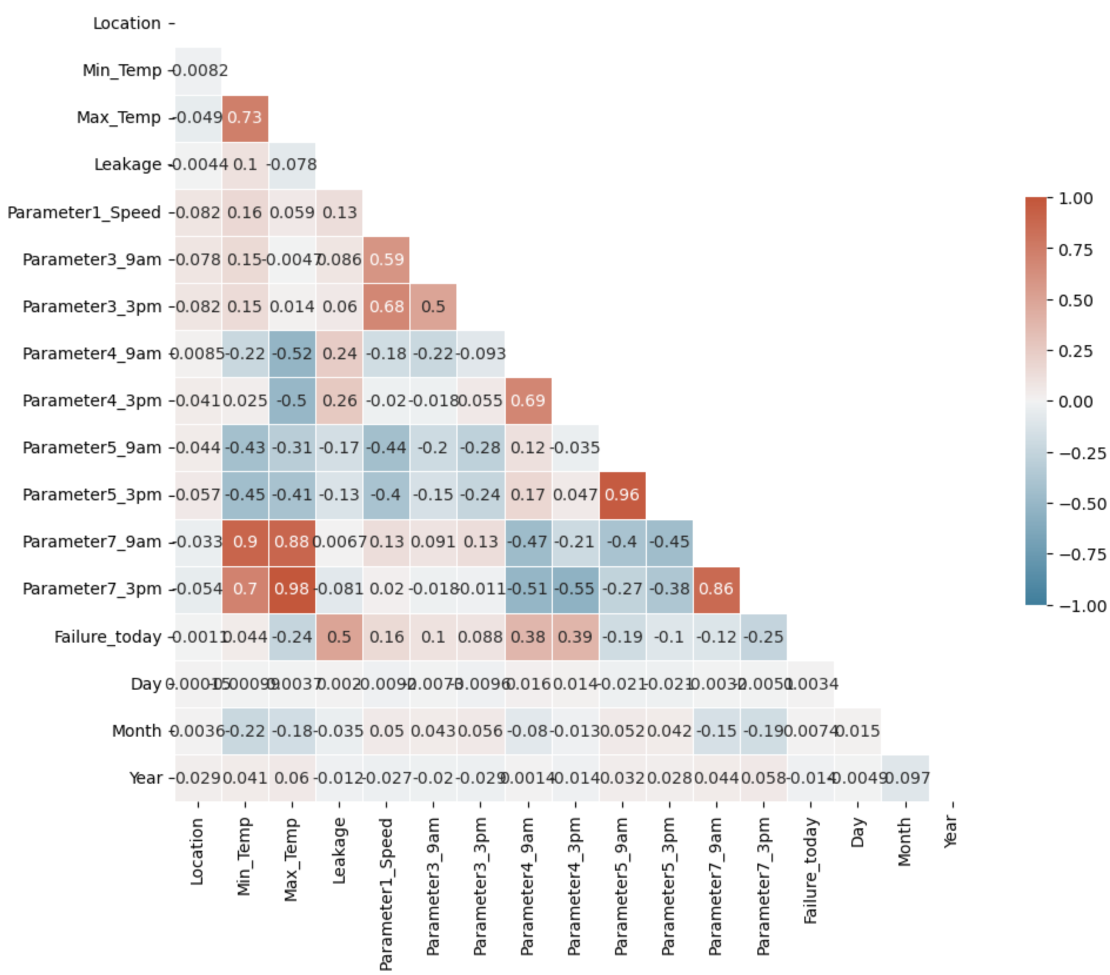
[5]: {'tags': ['hide_input']}

Haremos un analisis de correlación de las variables

[6]:
```
corr = df.corr()
mask = np.triu(np.ones_like(corr, dtype=bool))
f, ax = plt.subplots(figsize=(11, 9))
cmap = sns.diverging_palette(230, 20, as_cmap=True)
sns.heatmap(corr, mask=mask, cmap=cmap, vmax=1, vmin=-1, annot=True, center=0,
            square=True, linewidths=.5, cbar_kws={"shrink": .5})

{"tags": ["hide_input"]}
```

[6]: {'tags': ['hide_input']}



Los paramétros 4, 5 y 7 tienen correlaciones por sobre 0.65 entre sus datos a las 9 am y las 3 pm. Así mismo, las temperaturas maximas y minimas tienen una correlación de 0.72. Reemplazaremos estas columnas por columnas que representen sus promedios. El parámetro 7 parece estar muy correlacionado con la temperatura, por lo que será eliminado

[7]:
```
df["Parameter4"]=((df["Parameter4_9am"]+df["Parameter4_3pm"])/2)
df.drop("Parameter4_9am",axis=1, inplace=True)
df.drop("Parameter4_3pm",axis=1, inplace=True)

df["Parameter5"]=((df["Parameter5_9am"]+df["Parameter5_3pm"])/2)
df.drop("Parameter5_9am",axis=1, inplace=True)
df.drop("Parameter5_3pm",axis=1, inplace=True)

df["Parameter7"]=((df["Parameter7_9am"]+df["Parameter7_3pm"])/2)
df.drop("Parameter7_9am",axis=1, inplace=True)
```

```
df.drop("Parameter7_3pm",axis=1, inplace=True)

df["Temperature"]=((df["Min_Temp"]+df["Max_Temp"])/2)
df.drop("Min_Temp",axis=1, inplace=True)
df.drop("Max_Temp",axis=1, inplace=True)

df.drop("Parameter7", axis=1, inplace=True)
df

{"tags": ["hide_input"]}
```

[7]: {'tags': ['hide_input']}

Ya que el parámetro 5 se mueve en una escala mucho más grande que nuestra variable dependiente, lo estandarizaremos.

[8]:
```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
df[['Parameter5']] = scaler.fit_transform(df[['Parameter5']])

{"tags": ["hide_input"]}
```

[8]: {'tags': ['hide_input']}

Creamos un gráfico con las fallas totales a nivel mensual para ver alguna posible estacionalidad en los datos. Podemos ver que en los meses iniciales y finales del año hay menos fallas en comparación con los meses centrales

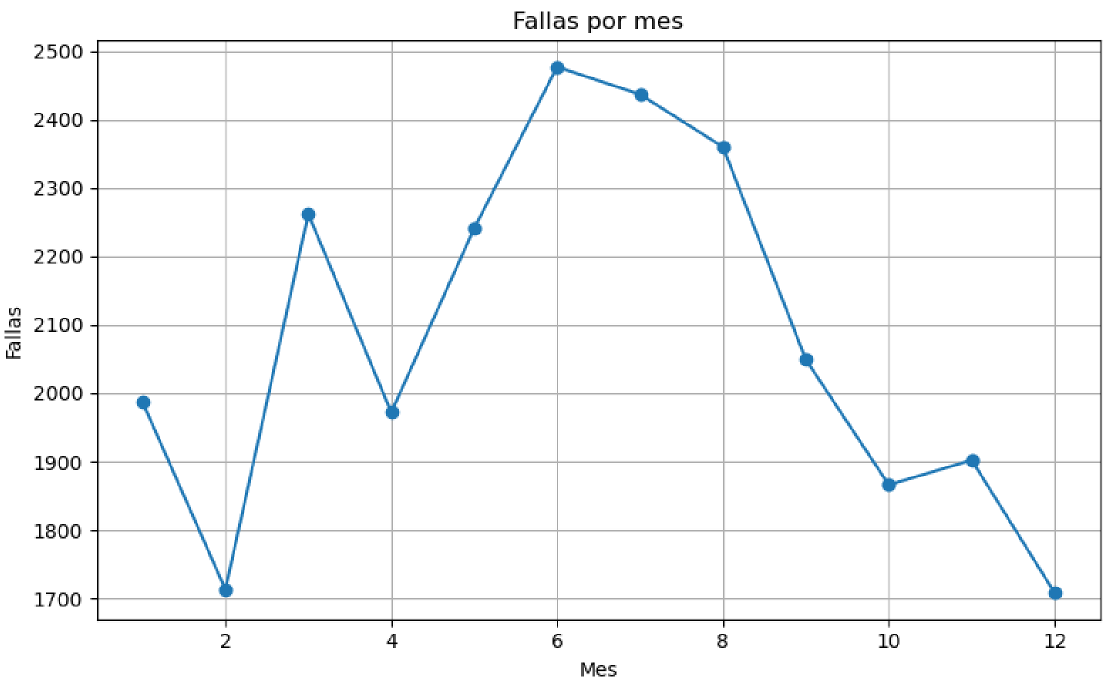
[9]:
```
Mes=[i for i in range(1,13)]
Fallas=[0,0,0,0,0,0,0,0,0,0,0,0]
for k in range(2009,2018):
    df3=df[df["Year"]==k]

    D=df3["Failure_today"].sum()
    Cc=0
    for j in range(0,11):
        c=0
        for i in range(len(df3)):
            if df3.iloc[i,10]==j+2 and df3.iloc[i,8]==1:
                c=c+1
        Cc=Cc+c
        Fallas[j+1]+=c
    Fallas[0]+=D-Cc

dfm={"Mes":Mes,"Fallas":Fallas}
plt.figure(figsize=(8, 5))  # tamaño del gráfico
plt.plot(dfm['Mes'], dfm['Fallas'], marker='o', linestyle='-')
```

```
plt.title('Fallas por mes')
plt.xlabel('Mes')
plt.ylabel('Fallas')
plt.grid(True)
plt.tight_layout()
plt.show()

{"tags": ["hide_input"]}
```



[9]: {'tags': ['hide_input']}

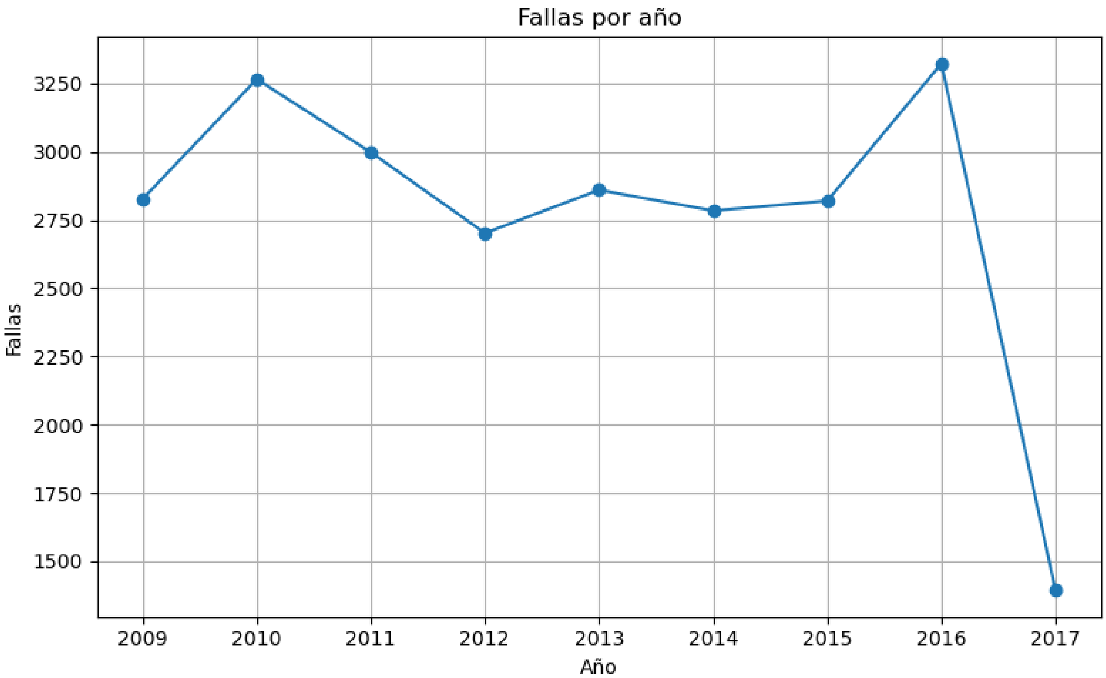
Creamos un gráfico para ver la cantidad de fallas por año. No parece que las fallas aumenten o disminuyan con los años significativamente. Notar que el año 2017 solo esta registrado hasta la mitad, por lo que se entiende que tenga menos fallas.

[10]:
```
Año=[i for i in range(2009,2018)]
Fallas=[]
for k in range(2009,2018):
    dfm=df[df["Year"]==k]
    F=dfm["Failure_today"].sum()
    Fallas.append(F)
print(Fallas)
dfm={"Año":Año,"Fallas":Fallas}
plt.figure(figsize=(8, 5))  # tamaño del gráfico
plt.plot(dfm['Año'], dfm['Fallas'], marker='o', linestyle='-')
```

```
plt.title('Fallas por año')
plt.xlabel('Año')
plt.ylabel('Fallas')
plt.grid(True)
plt.tight_layout()
plt.show()

{"tags": ["hide_input"]}
```

```
[2827, 3266, 2998, 2702, 2860, 2784, 2820, 3323, 1394]
```




```
[10]: {'tags': ['hide_input']}
```

2 Pregunta 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.

Creación de modelo MCO

```
[11]: model = smf.ols("Failure_today ~    Leakage+C(Parameter1_Dir)+␣
↪Parameter1_Speed+ C(Parameter1_Dir)*Parameter1_Speed + C(Parameter2_9am) +␣
↪C(Parameter2_3pm) + Parameter3_9am + Parameter3_3pm + C(Month)+ Day+ Year +␣
↪Parameter4 + Parameter5+ C(Location)+ Temperature", data=df).fit()
```

```
print(model.summary())

{"tags": ["hide_input"]}
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Failure_today   R-squared:                       0.382
Model:                            OLS   Adj. R-squared:                  0.382
Method:                 Least Squares   F-statistic:                     916.6
Date:                Thu, 24 Apr 2025   Prob (F-statistic):               0.00
Time:                        22:08:40   Log-Likelihood:                -33826.
No. Observations:              111179   AIC:                         6.780e+04
Df Residuals:                  111103   BIC:                         6.854e+04
Df Model:                          75
Covariance Type:            nonrobust
================================================================================
===========================
                                              coef    std err          t
P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
---------------------------
Intercept                                  -1.2402      0.820     -1.513
0.130      -2.847       0.367
C(Parameter1_Dir)[T.N]                      0.0178      0.010      1.738
0.082      -0.002       0.038
C(Parameter1_Dir)[T.S]                     -0.0077      0.010     -0.744
0.457      -0.028       0.013
C(Parameter1_Dir)[T.W]                     -0.0041      0.011     -0.380
0.704      -0.025       0.017
C(Parameter2_9am)[T.N]                      0.0063      0.003      1.848
0.065      -0.000       0.013
C(Parameter2_9am)[T.S]                      0.0488      0.003     14.822
0.000       0.042       0.055
C(Parameter2_9am)[T.W]                      0.0714      0.004     18.119
0.000       0.064       0.079
C(Parameter2_3pm)[T.N]                     -0.0142      0.004     -3.858
0.000      -0.021      -0.007
C(Parameter2_3pm)[T.S]                      0.0195      0.003      5.617
0.000       0.013       0.026
C(Parameter2_3pm)[T.W]                      0.0321      0.004      7.862
0.000       0.024       0.040
C(Month)[T.2]                              -0.0066      0.005     -1.373
0.170      -0.016       0.003
C(Month)[T.3]                               0.0167      0.005      3.591
0.000       0.008       0.026
C(Month)[T.4]                               0.0418      0.005      8.196
0.000       0.032       0.052
C(Month)[T.5]                               0.0368      0.006      6.625
```

				0.000	0.026	0.048
C(Month)[T.6]	0.0216	0.006	3.568	0.000	0.010	0.034
C(Month)[T.7]	0.0583	0.006	9.250	0.000	0.046	0.071
C(Month)[T.8]	0.0735	0.006	12.101	0.000	0.062	0.085
C(Month)[T.9]	0.0640	0.006	11.348	0.000	0.053	0.075
C(Month)[T.10]	0.0612	0.005	11.817	0.000	0.051	0.071
C(Month)[T.11]	0.0454	0.005	9.401	0.000	0.036	0.055
C(Month)[T.12]	0.0323	0.005	6.681	0.000	0.023	0.042
C(Location)[T.3]	-0.0789	0.009	-8.426	0.000	-0.097	-0.061
C(Location)[T.4]	0.0098	0.009	1.057	0.290	-0.008	0.028
C(Location)[T.5]	-0.1209	0.009	-12.771	0.000	-0.139	-0.102
C(Location)[T.6]	-0.1649	0.009	-17.737	0.000	-0.183	-0.147
C(Location)[T.7]	-0.1055	0.009	-11.527	0.000	-0.123	-0.088
C(Location)[T.8]	-0.0724	0.009	-7.820	0.000	-0.091	-0.054
C(Location)[T.9]	-0.1515	0.010	-15.480	0.000	-0.171	-0.132
C(Location)[T.10]	-0.1001	0.010	-10.474	0.000	-0.119	-0.081
C(Location)[T.11]	-0.0412	0.009	-4.564	0.000	-0.059	-0.023
C(Location)[T.12]	-0.1138	0.009	-12.123	0.000	-0.132	-0.095
C(Location)[T.13]	-0.1043	0.010	-10.865	0.000	-0.123	-0.085
C(Location)[T.14]	-0.1441	0.010	-14.697	0.000	-0.163	-0.125
C(Location)[T.15]	-0.1456	0.009	-15.479	0.000	-0.164	-0.127
C(Location)[T.16]	-0.0870	0.009	-9.471	0.000	-0.105	-0.069
C(Location)[T.17]	-0.1488	0.015	-10.080	0.000	-0.178	-0.120
C(Location)[T.18]	-0.0916	0.011	-8.408	0.000	-0.113	-0.070
C(Location)[T.19]	-0.0725	0.010	-7.355			

```
0.000      -0.092       -0.053
C(Location)[T.20]                          -0.1224      0.009    -13.452
0.000      -0.140       -0.105
C(Location)[T.21]                          -0.0994      0.009    -11.149
0.000      -0.117       -0.082
C(Location)[T.22]                          -0.0775      0.009     -8.424
0.000      -0.096       -0.059
C(Location)[T.23]                          -0.0768      0.009     -8.432
0.000      -0.095       -0.059
C(Location)[T.26]                          -0.1505      0.011    -14.024
0.000      -0.172       -0.129
C(Location)[T.27]                          -0.1649      0.009    -17.924
0.000      -0.183       -0.147
C(Location)[T.28]                          -0.1290      0.009    -13.983
0.000      -0.147       -0.111
C(Location)[T.29]                          -0.0783      0.009     -8.673
0.000      -0.096       -0.061
C(Location)[T.30]                          -0.0655      0.009     -6.986
0.000      -0.084       -0.047
C(Location)[T.32]                          -0.0623      0.009     -6.946
0.000      -0.080       -0.045
C(Location)[T.33]                          -0.0622      0.009     -6.914
0.000      -0.080       -0.045
C(Location)[T.34]                          -0.0825      0.009     -8.994
0.000      -0.100       -0.065
C(Location)[T.35]                          -0.1071      0.010    -10.950
0.000      -0.126       -0.088
C(Location)[T.36]                          -0.1799      0.009    -19.415
0.000      -0.198       -0.162
C(Location)[T.38]                          -0.1190      0.010    -12.266
0.000      -0.138       -0.100
C(Location)[T.39]                          -0.1009      0.009    -11.111
0.000      -0.119       -0.083
C(Location)[T.40]                          -0.1760      0.010    -18.049
0.000      -0.195       -0.157
C(Location)[T.41]                          -0.0851      0.010     -8.903
0.000      -0.104       -0.066
C(Location)[T.42]                          -0.0079      0.011     -0.707
0.480      -0.030        0.014
C(Location)[T.43]                          -0.0705      0.009     -7.768
0.000      -0.088       -0.053
C(Location)[T.44]                          -0.1011      0.009    -10.815
0.000      -0.119       -0.083
C(Location)[T.45]                          -0.1029      0.009    -11.288
0.000      -0.121       -0.085
C(Location)[T.46]                          -0.1153      0.010    -11.888
0.000      -0.134       -0.096
C(Location)[T.47]                          -0.0821      0.009     -8.659
```

```
0.000      -0.101      -0.064
C(Location)[T.48]                          -0.1839      0.009    -20.030
0.000      -0.202      -0.166
C(Location)[T.49]                          -0.0896      0.009     -9.886
0.000      -0.107      -0.072
Leakage                                     0.0177      0.000    143.400
0.000       0.017       0.018
Parameter1_Speed                            0.0031      0.000     13.063
0.000       0.003       0.004
C(Parameter1_Dir)[T.N]:Parameter1_Speed    -0.0009      0.000     -3.301
0.001      -0.001      -0.000
C(Parameter1_Dir)[T.S]:Parameter1_Speed     0.0004      0.000      1.657
0.097   -7.84e-05       0.001
C(Parameter1_Dir)[T.W]:Parameter1_Speed     0.0005      0.000      1.961
0.050    3.12e-07       0.001
Parameter3_9am                              0.0031      0.000     19.114
0.000       0.003       0.003
Parameter3_3pm                             -0.0019      0.000    -10.983
0.000      -0.002      -0.002
Day                                        -0.0002      0.000     -1.621
0.105      -0.000     3.8e-05
Year                                        0.0004      0.000      0.982
0.326      -0.000       0.001
Parameter4                                  0.0084   7.43e-05    113.216
0.000       0.008       0.009
Parameter5                                 -0.0405      0.001    -29.486
0.000      -0.043      -0.038
Temperature                                 0.0006      0.000      1.652
0.099      -0.000       0.001
==============================================================================
Omnibus:                    16135.505   Durbin-Watson:                   1.859
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            86702.031
Skew:                           0.595   Prob(JB):                         0.00
Kurtosis:                       7.160   Cond. No.                     1.68e+06
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.68e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[11]: {'tags': ['hide_input']}

El día, el año, la temperatura y la interacción del parámetro 1 Dirección-Velocidad tienen valores p>0.05 y al eliminarlos del modelo el R cuadrado no disminuye, por lo que los eliminaremos. Finalmente nos quedaremos con los parámetros 1 (Dirección y velocidad), 2 (En sus 2 horarios), 3 (En sus 2 horarios), 4 y 5. También con las variables Location, Month y Leakage. La interpretación de los resultados es de la siguiente forma: Si la variable leakage aumenta en una unidad, la probabilidad de que ese día ocurra una falla aumentará en 0.01777, a excepción del parámetro 5 que está estandarizado, en ese caso un cambio de una desviación estándar del parámetro disminuye la probabilidad de falla en 0.04 . Factores como las fugas, la velocidad del viento y el parámetro 4 aumentan la probabilidad de fallas cuando aumentan. Las direcciones del viento S y W parecen aumentar la probabilidad de fallas con respecto a la dirección E, mientras que la dirección N parece disminuirla. El aumento del parámetro 3 a las 9am parece aumentar la probabilidad de falla, mientras que a las 3 pm parece disminuirla.

[12]:
```
model = smf.ols("Failure_today ~   Leakage+ C(Parameter1_Dir)+ Parameter1_Speed␣
 ↪+ C(Parameter2_9am) + C(Parameter2_3pm) + Parameter3_9am + Parameter3_3pm +␣
 ↪C(Month) + Parameter4 + Parameter5 +C(Location)", data=df).fit()
print(model.summary())

{"tags": ["hide_input"]}
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Failure_today   R-squared:                       0.382
Model:                            OLS   Adj. R-squared:                  0.381
Method:                 Least Squares   F-statistic:                     994.8
Date:                Thu, 24 Apr 2025   Prob (F-statistic):               0.00
Time:                        22:08:42   Log-Likelihood:                -33859.
No. Observations:              111179   AIC:                         6.786e+04
Df Residuals:                  111109   BIC:                         6.853e+04
Df Model:                          69
Covariance Type:            nonrobust
================================================================================
==========
                             coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
----------
Intercept                 -0.4242      0.009    -46.250      0.000      -0.442
-0.406
C(Parameter1_Dir)[T.N]    -0.0156      0.004     -4.141      0.000      -0.023
-0.008
C(Parameter1_Dir)[T.S]     0.0096      0.004      2.708      0.007       0.003
0.016
C(Parameter1_Dir)[T.W]     0.0188      0.004      4.580      0.000       0.011
0.027
C(Parameter2_9am)[T.N]     0.0050      0.003      1.496      0.135      -0.002
0.012
C(Parameter2_9am)[T.S]     0.0484      0.003     14.770      0.000       0.042
0.055
C(Parameter2_9am)[T.W]     0.0714      0.004     18.199      0.000       0.064
0.079
C(Parameter2_3pm)[T.N]    -0.0141      0.004     -3.851      0.000      -0.021
```

```
-0.007
C(Parameter2_3pm)[T.S]      0.0196      0.003      5.672      0.000       0.013
0.026
C(Parameter2_3pm)[T.W]      0.0325      0.004      7.971      0.000       0.025
0.040
C(Month)[T.2]              -0.0064      0.005     -1.344      0.179      -0.016
0.003
C(Month)[T.3]               0.0155      0.005      3.350      0.001       0.006
0.025
C(Month)[T.4]               0.0389      0.005      7.952      0.000       0.029
0.048
C(Month)[T.5]               0.0317      0.005      6.473      0.000       0.022
0.041
C(Month)[T.6]               0.0151      0.005      2.980      0.003       0.005
0.025
C(Month)[T.7]               0.0510      0.005      9.885      0.000       0.041
0.061
C(Month)[T.8]               0.0666      0.005     13.330      0.000       0.057
0.076
C(Month)[T.9]               0.0582      0.005     11.931      0.000       0.049
0.068
C(Month)[T.10]              0.0570      0.005     11.931      0.000       0.048
0.066
C(Month)[T.11]              0.0428      0.005      9.126      0.000       0.034
0.052
C(Month)[T.12]              0.0309      0.005      6.440      0.000       0.021
0.040
C(Location)[T.3]           -0.0798      0.009     -8.538      0.000      -0.098
-0.062
C(Location)[T.4]            0.0124      0.009      1.334      0.182      -0.006
0.031
C(Location)[T.5]           -0.1192      0.009    -12.598      0.000      -0.138
-0.101
C(Location)[T.6]           -0.1684      0.009    -18.341      0.000      -0.186
-0.150
C(Location)[T.7]           -0.1060      0.009    -11.629      0.000      -0.124
-0.088
C(Location)[T.8]           -0.0691      0.009     -7.586      0.000      -0.087
-0.051
C(Location)[T.9]           -0.1435      0.009    -15.464      0.000      -0.162
-0.125
C(Location)[T.10]          -0.1016      0.009    -10.751      0.000      -0.120
-0.083
C(Location)[T.11]          -0.0399      0.009     -4.434      0.000      -0.058
-0.022
C(Location)[T.12]          -0.1125      0.009    -12.033      0.000      -0.131
-0.094
C(Location)[T.13]          -0.1053      0.010    -11.014      0.000      -0.124
```

```
-0.087
C(Location)[T.14]         -0.1359      0.009    -14.887      0.000      -0.154
-0.118
C(Location)[T.15]         -0.1395      0.009    -15.114      0.000      -0.158
-0.121
C(Location)[T.16]         -0.0897      0.009     -9.971      0.000      -0.107
-0.072
C(Location)[T.17]         -0.1414      0.014     -9.818      0.000      -0.170
-0.113
C(Location)[T.18]         -0.0925      0.011     -8.592      0.000      -0.114
-0.071
C(Location)[T.19]         -0.0761      0.010     -7.732      0.000      -0.095
-0.057
C(Location)[T.20]         -0.1270      0.009    -14.042      0.000      -0.145
-0.109
C(Location)[T.21]         -0.0990      0.009    -11.106      0.000      -0.116
-0.082
C(Location)[T.22]         -0.0751      0.009     -8.181      0.000      -0.093
-0.057
C(Location)[T.23]         -0.0775      0.009     -8.556      0.000      -0.095
-0.060
C(Location)[T.26]         -0.1491      0.011    -13.963      0.000      -0.170
-0.128
C(Location)[T.27]         -0.1621      0.009    -17.707      0.000      -0.180
-0.144
C(Location)[T.28]         -0.1270      0.009    -13.851      0.000      -0.145
-0.109
C(Location)[T.29]         -0.0786      0.009     -8.727      0.000      -0.096
-0.061
C(Location)[T.30]         -0.0638      0.009     -6.833      0.000      -0.082
-0.045
C(Location)[T.32]         -0.0609      0.009     -6.802      0.000      -0.078
-0.043
C(Location)[T.33]         -0.0608      0.009     -6.779      0.000      -0.078
-0.043
C(Location)[T.34]         -0.0830      0.009     -9.115      0.000      -0.101
-0.065
C(Location)[T.35]         -0.1048      0.010    -10.734      0.000      -0.124
-0.086
C(Location)[T.36]         -0.1802      0.009    -19.535      0.000      -0.198
-0.162
C(Location)[T.38]         -0.1170      0.010    -12.116      0.000      -0.136
-0.098
C(Location)[T.39]         -0.0977      0.009    -10.795      0.000      -0.115
-0.080
C(Location)[T.40]         -0.1688      0.009    -18.022      0.000      -0.187
-0.150
C(Location)[T.41]         -0.0850      0.009     -8.971      0.000      -0.104
```

```
-0.066
C(Location)[T.42]          -0.0035      0.011     -0.316      0.752      -0.025
0.018
C(Location)[T.43]          -0.0707      0.009     -7.799      0.000      -0.088
-0.053
C(Location)[T.44]          -0.1014      0.009    -10.860      0.000      -0.120
-0.083
C(Location)[T.45]          -0.1049      0.009    -11.542      0.000      -0.123
-0.087
C(Location)[T.46]          -0.1131      0.010    -11.672      0.000      -0.132
-0.094
C(Location)[T.47]          -0.0824      0.009     -8.693      0.000      -0.101
-0.064
C(Location)[T.48]          -0.1811      0.009    -19.757      0.000      -0.199
-0.163
C(Location)[T.49]          -0.0870      0.009     -9.623      0.000      -0.105
-0.069
Leakage                     0.0177      0.000    143.649      0.000       0.018
0.018
Parameter1_Speed            0.0031      0.000     25.169      0.000       0.003
0.003
Parameter3_9am              0.0030      0.000     18.968      0.000       0.003
0.003
Parameter3_3pm             -0.0019      0.000    -11.420      0.000      -0.002
-0.002
Parameter4                  0.0084   7.07e-05    118.766      0.000       0.008
0.009
Parameter5                 -0.0403      0.001    -30.480      0.000      -0.043
-0.038
==============================================================================
Omnibus:                    16091.859   Durbin-Watson:                   1.858
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            86446.958
Skew:                           0.593   Prob(JB):                         0.00
Kurtosis:                       7.154   Cond. No.                     3.47e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 3.47e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[12]: {'tags': ['hide_input']}

3 Pregunta 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.

Podemos ver que ha habido un fallo al calcular los coeficientes de la regresión. Estos fallos pueden ocurrir por correlaciones entre las variables, por lo que revisaremos si hay alguna correlación muy alta

[13]:
```
probit_model1 = smf.probit("Failure_today ~    Leakage+C(Parameter1_Dir)+␣
↪Parameter1_Speed + C(Parameter2_9am) + C(Parameter2_3pm) + Parameter3_9am +␣
↪Parameter3_3pm + C(Month) + Parameter4 + Parameter5 + C(Location)", data=df).
↪fit()
print(probit_model1.summary())

mfx = probit_model1.get_margeff()
print(mfx.summary())

{"tags": ["hide_input"]}
```

```
Optimization terminated successfully.
         Current function value: nan
         Iterations 24
                          Probit Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:               111179
Model:                         Probit   Df Residuals:                   111109
Method:                           MLE   Df Model:                           69
Date:                Thu, 24 Apr 2025   Pseudo R-squ.:                     nan
Time:                        22:08:46   Log-Likelihood:                    nan
converged:                       True   LL-Null:                       -59226.
Covariance Type:            nonrobust   LLR p-value:                       nan
================================================================================
==========
                             coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----------
Intercept                     nan        nan        nan        nan         nan
nan
C(Parameter1_Dir)[T.N]        nan        nan        nan        nan         nan
nan
C(Parameter1_Dir)[T.S]        nan        nan        nan        nan         nan
nan
C(Parameter1_Dir)[T.W]        nan        nan        nan        nan         nan
nan
C(Parameter2_9am)[T.N]        nan        nan        nan        nan         nan
nan
C(Parameter2_9am)[T.S]        nan        nan        nan        nan         nan
```

```
nan
C(Parameter2_9am)[T.W]         nan        nan        nan        nan        nan
nan
C(Parameter2_3pm)[T.N]         nan        nan        nan        nan        nan
nan
C(Parameter2_3pm)[T.S]         nan        nan        nan        nan        nan
nan
C(Parameter2_3pm)[T.W]         nan        nan        nan        nan        nan
nan
C(Month)[T.2]                  nan        nan        nan        nan        nan
nan
C(Month)[T.3]                  nan        nan        nan        nan        nan
nan
C(Month)[T.4]                  nan        nan        nan        nan        nan
nan
C(Month)[T.5]                  nan        nan        nan        nan        nan
nan
C(Month)[T.6]                  nan        nan        nan        nan        nan
nan
C(Month)[T.7]                  nan        nan        nan        nan        nan
nan
C(Month)[T.8]                  nan        nan        nan        nan        nan
nan
C(Month)[T.9]                  nan        nan        nan        nan        nan
nan
C(Month)[T.10]                 nan        nan        nan        nan        nan
nan
C(Month)[T.11]                 nan        nan        nan        nan        nan
nan
C(Month)[T.12]                 nan        nan        nan        nan        nan
nan
C(Location)[T.3]               nan        nan        nan        nan        nan
nan
C(Location)[T.4]               nan        nan        nan        nan        nan
nan
C(Location)[T.5]               nan        nan        nan        nan        nan
nan
C(Location)[T.6]               nan        nan        nan        nan        nan
nan
C(Location)[T.7]               nan        nan        nan        nan        nan
nan
C(Location)[T.8]               nan        nan        nan        nan        nan
nan
C(Location)[T.9]               nan        nan        nan        nan        nan
nan
C(Location)[T.10]              nan        nan        nan        nan        nan
nan
C(Location)[T.11]              nan        nan        nan        nan        nan
```

```
nan
C(Location)[T.12]          nan        nan        nan        nan         nan
nan
C(Location)[T.13]          nan        nan        nan        nan         nan
nan
C(Location)[T.14]          nan        nan        nan        nan         nan
nan
C(Location)[T.15]          nan        nan        nan        nan         nan
nan
C(Location)[T.16]          nan        nan        nan        nan         nan
nan
C(Location)[T.17]          nan        nan        nan        nan         nan
nan
C(Location)[T.18]          nan        nan        nan        nan         nan
nan
C(Location)[T.19]          nan        nan        nan        nan         nan
nan
C(Location)[T.20]          nan        nan        nan        nan         nan
nan
C(Location)[T.21]          nan        nan        nan        nan         nan
nan
C(Location)[T.22]          nan        nan        nan        nan         nan
nan
C(Location)[T.23]          nan        nan        nan        nan         nan
nan
C(Location)[T.26]          nan        nan        nan        nan         nan
nan
C(Location)[T.27]          nan        nan        nan        nan         nan
nan
C(Location)[T.28]          nan        nan        nan        nan         nan
nan
C(Location)[T.29]          nan        nan        nan        nan         nan
nan
C(Location)[T.30]          nan        nan        nan        nan         nan
nan
C(Location)[T.32]          nan        nan        nan        nan         nan
nan
C(Location)[T.33]          nan        nan        nan        nan         nan
nan
C(Location)[T.34]          nan        nan        nan        nan         nan
nan
C(Location)[T.35]          nan        nan        nan        nan         nan
nan
C(Location)[T.36]          nan        nan        nan        nan         nan
nan
C(Location)[T.38]          nan        nan        nan        nan         nan
nan
C(Location)[T.39]          nan        nan        nan        nan         nan
```

```
nan
C(Location)[T.40]           nan        nan        nan        nan          nan
nan
C(Location)[T.41]           nan        nan        nan        nan          nan
nan
C(Location)[T.42]           nan        nan        nan        nan          nan
nan
C(Location)[T.43]           nan        nan        nan        nan          nan
nan
C(Location)[T.44]           nan        nan        nan        nan          nan
nan
C(Location)[T.45]           nan        nan        nan        nan          nan
nan
C(Location)[T.46]           nan        nan        nan        nan          nan
nan
C(Location)[T.47]           nan        nan        nan        nan          nan
nan
C(Location)[T.48]           nan        nan        nan        nan          nan
nan
C(Location)[T.49]           nan        nan        nan        nan          nan
nan
Leakage                     nan        nan        nan        nan          nan
nan
Parameter1_Speed            nan        nan        nan        nan          nan
nan
Parameter3_9am              nan        nan        nan        nan          nan
nan
Parameter3_3pm              nan        nan        nan        nan          nan
nan
Parameter4                  nan        nan        nan        nan          nan
nan
Parameter5                  nan        nan        nan        nan          nan
nan
================================================================================
==========
       Probit Marginal Effects
=====================================
Dep. Variable:          Failure_today
Method:                          dydx
At:                           overall
================================================================================
==========
                          dy/dx    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----------
C(Parameter1_Dir)[T.N]      nan        nan        nan        nan          nan
nan
```

```
C(Parameter1_Dir)[T.S]          nan        nan        nan        nan         nan
nan
C(Parameter1_Dir)[T.W]          nan        nan        nan        nan         nan
nan
C(Parameter2_9am)[T.N]          nan        nan        nan        nan         nan
nan
C(Parameter2_9am)[T.S]          nan        nan        nan        nan         nan
nan
C(Parameter2_9am)[T.W]          nan        nan        nan        nan         nan
nan
C(Parameter2_3pm)[T.N]          nan        nan        nan        nan         nan
nan
C(Parameter2_3pm)[T.S]          nan        nan        nan        nan         nan
nan
C(Parameter2_3pm)[T.W]          nan        nan        nan        nan         nan
nan
C(Month)[T.2]                   nan        nan        nan        nan         nan
nan
C(Month)[T.3]                   nan        nan        nan        nan         nan
nan
C(Month)[T.4]                   nan        nan        nan        nan         nan
nan
C(Month)[T.5]                   nan        nan        nan        nan         nan
nan
C(Month)[T.6]                   nan        nan        nan        nan         nan
nan
C(Month)[T.7]                   nan        nan        nan        nan         nan
nan
C(Month)[T.8]                   nan        nan        nan        nan         nan
nan
C(Month)[T.9]                   nan        nan        nan        nan         nan
nan
C(Month)[T.10]                  nan        nan        nan        nan         nan
nan
C(Month)[T.11]                  nan        nan        nan        nan         nan
nan
C(Month)[T.12]                  nan        nan        nan        nan         nan
nan
C(Location)[T.3]                nan        nan        nan        nan         nan
nan
C(Location)[T.4]                nan        nan        nan        nan         nan
nan
C(Location)[T.5]                nan        nan        nan        nan         nan
nan
C(Location)[T.6]                nan        nan        nan        nan         nan
nan
C(Location)[T.7]                nan        nan        nan        nan         nan
nan
```

C(Location)[T.8]	nan	nan	nan	nan	nan	nan
C(Location)[T.9]	nan	nan	nan	nan	nan	nan
C(Location)[T.10]	nan	nan	nan	nan	nan	nan
C(Location)[T.11]	nan	nan	nan	nan	nan	nan
C(Location)[T.12]	nan	nan	nan	nan	nan	nan
C(Location)[T.13]	nan	nan	nan	nan	nan	nan
C(Location)[T.14]	nan	nan	nan	nan	nan	nan
C(Location)[T.15]	nan	nan	nan	nan	nan	nan
C(Location)[T.16]	nan	nan	nan	nan	nan	nan
C(Location)[T.17]	nan	nan	nan	nan	nan	nan
C(Location)[T.18]	nan	nan	nan	nan	nan	nan
C(Location)[T.19]	nan	nan	nan	nan	nan	nan
C(Location)[T.20]	nan	nan	nan	nan	nan	nan
C(Location)[T.21]	nan	nan	nan	nan	nan	nan
C(Location)[T.22]	nan	nan	nan	nan	nan	nan
C(Location)[T.23]	nan	nan	nan	nan	nan	nan
C(Location)[T.26]	nan	nan	nan	nan	nan	nan
C(Location)[T.27]	nan	nan	nan	nan	nan	nan
C(Location)[T.28]	nan	nan	nan	nan	nan	nan
C(Location)[T.29]	nan	nan	nan	nan	nan	nan
C(Location)[T.30]	nan	nan	nan	nan	nan	nan
C(Location)[T.32]	nan	nan	nan	nan	nan	nan
C(Location)[T.33]	nan	nan	nan	nan	nan	nan
C(Location)[T.34]	nan	nan	nan	nan	nan	nan

```
C(Location)[T.35]                  nan         nan         nan         nan         nan
nan
C(Location)[T.36]                  nan         nan         nan         nan         nan
nan
C(Location)[T.38]                  nan         nan         nan         nan         nan
nan
C(Location)[T.39]                  nan         nan         nan         nan         nan
nan
C(Location)[T.40]                  nan         nan         nan         nan         nan
nan
C(Location)[T.41]                  nan         nan         nan         nan         nan
nan
C(Location)[T.42]                  nan         nan         nan         nan         nan
nan
C(Location)[T.43]                  nan         nan         nan         nan         nan
nan
C(Location)[T.44]                  nan         nan         nan         nan         nan
nan
C(Location)[T.45]                  nan         nan         nan         nan         nan
nan
C(Location)[T.46]                  nan         nan         nan         nan         nan
nan
C(Location)[T.47]                  nan         nan         nan         nan         nan
nan
C(Location)[T.48]                  nan         nan         nan         nan         nan
nan
C(Location)[T.49]                  nan         nan         nan         nan         nan
nan
Leakage                            nan         nan         nan         nan         nan
nan
Parameter1_Speed                   nan         nan         nan         nan         nan
nan
Parameter3_9am                     nan         nan         nan         nan         nan
nan
Parameter3_3pm                     nan         nan         nan         nan         nan
nan
Parameter4                         nan         nan         nan         nan         nan
nan
Parameter5                         nan         nan         nan         nan         nan
nan
================================================================================
==========
```

[13]: {'tags': ['hide_input']}

[14]:
```
X=df.drop("Failure_today",axis=1)
X.corr()
```

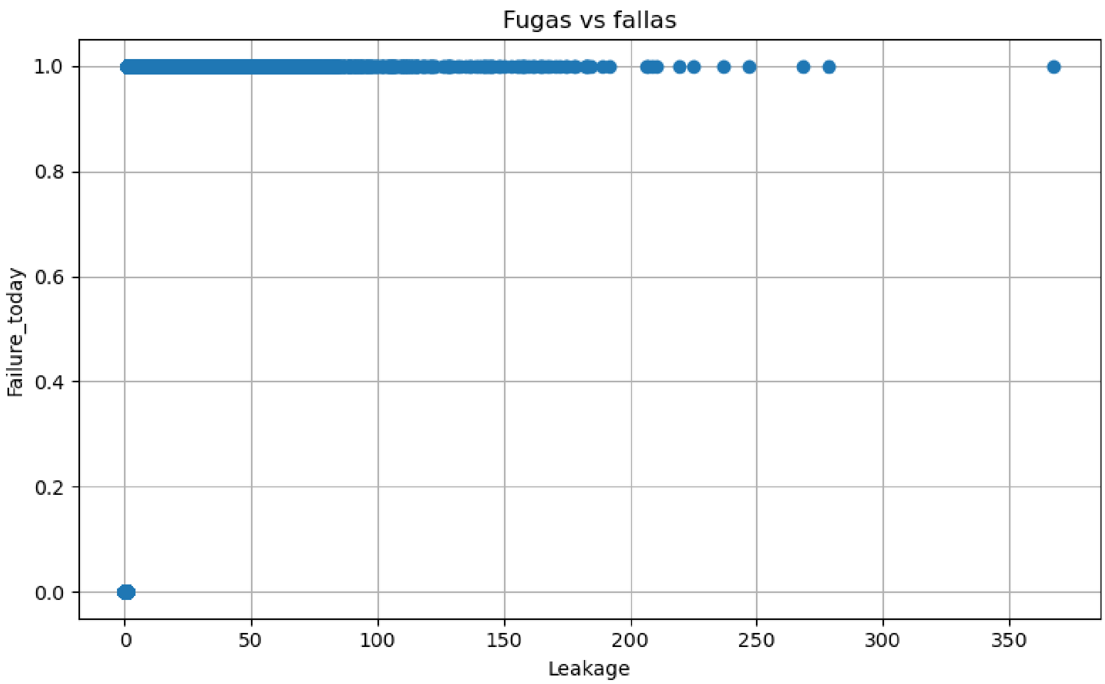
[14]:
```
                 Location   Leakage  Parameter1_Speed  Parameter3_9am  \
Location         1.000000 -0.004391          0.082421        0.078436
Leakage         -0.004391  1.000000          0.130412        0.085633
Parameter1_Speed  0.082421  0.130412          1.000000        0.589986
Parameter3_9am   0.078436  0.085633          0.589986        1.000000
Parameter3_3pm   0.081880  0.060184          0.679434        0.499260
Day              0.000145  0.001954         -0.009190       -0.007321
Month            0.003646 -0.035109          0.049568        0.042794
Year             0.029313 -0.011593         -0.026627       -0.019739
Parameter4       0.027951  0.271737         -0.106684       -0.124955
Parameter5       0.050594 -0.148029         -0.425855       -0.176197
Temperature     -0.032194  0.008449          0.115364        0.074475

                 Parameter3_3pm       Day     Month      Year  Parameter4  \
Location               0.081880  0.000145  0.003646  0.029313    0.027951
Leakage                0.060184  0.001954 -0.035109 -0.011593    0.271737
Parameter1_Speed       0.679434 -0.009190  0.049568 -0.026627   -0.106684
Parameter3_9am         0.499260 -0.007321  0.042794 -0.019739   -0.124955
Parameter3_3pm         1.000000 -0.009575  0.056126 -0.029215   -0.017074
Day                   -0.009575  1.000000  0.014984 -0.004875    0.016577
Month                  0.056126  0.014984  1.000000 -0.097206   -0.048878
Year                  -0.029215 -0.004875 -0.097206  1.000000   -0.007488
Parameter4            -0.017074  0.016577 -0.048878 -0.007488    1.000000
Parameter5            -0.265455 -0.021383  0.047244  0.030608    0.077695
Temperature            0.082626 -0.002593 -0.213105  0.054961   -0.366543

                 Parameter5  Temperature
Location           0.050594    -0.032194
Leakage           -0.148029     0.008449
Parameter1_Speed  -0.425855     0.115364
Parameter3_9am    -0.176197     0.074475
Parameter3_3pm    -0.265455     0.082626
Day               -0.021383    -0.002593
Month              0.047244    -0.213105
Year               0.030608     0.054961
Parameter4         0.077695    -0.366543
Parameter5         1.000000    -0.434422
Temperature       -0.434422     1.000000
```

[15]:
```python
plt.figure(figsize=(8, 5))  # tamaño del gráfico
plt.plot(df['Leakage'], df['Failure_today'], marker='o', linestyle='')
plt.title('Fugas vs fallas')
plt.xlabel('Leakage')
plt.ylabel('Failure_today')
plt.grid(True)
plt.tight_layout()
plt.show()
```

```
{"tags": ["hide_input"]}
```



[15]:
```
{'tags': ['hide_input']}
```

Podemos ver que Parameter3_3pm tiene una correlación de mas de 0.65 con la variable Parameter1_Speed, por lo que la eliminamos. También la variable Leakage parece estar generando problemas en el modelo probablemente debido a que explica perfectamente a la variable dependiente, por lo que la eliminaremos. Además en este modelo, la variable Temperature sí es significativa, por lo que la añadiremos.

[16]:
```
probit_model = smf.probit("Failure_today ~   Temperature+ C(Parameter1_Dir)+␣
↪Parameter1_Speed + C(Parameter2_9am) + C(Parameter2_3pm) + Parameter3_9am +␣
↪C(Month) + Parameter4 + Parameter5 + C(Location)", data=df).fit()
print(probit_model.summary())

mfx = probit_model.get_margeff()
print(mfx.summary())

{"tags": ["hide_input"]}
```

```
Optimization terminated successfully.
         Current function value: 0.380021
         Iterations 7
                          Probit Regression Results
==============================================================================
```

```
Dep. Variable:          Failure_today   No. Observations:               111179
Model:                         Probit   Df Residuals:                   111110
Method:                           MLE   Df Model:                           68
Date:                Thu, 24 Apr 2025   Pseudo R-squ.:                  0.2866
Time:                        22:09:12   Log-Likelihood:                -42250.
converged:                       True   LL-Null:                       -59226.
Covariance Type:            nonrobust   LLR p-value:                     0.000
================================================================================
==========
                           coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----------
Intercept                 -4.1898      0.069    -60.303      0.000      -4.326
-4.054
C(Parameter1_Dir)[T.N]    -0.1180      0.020     -5.811      0.000      -0.158
-0.078
C(Parameter1_Dir)[T.S]     0.0269      0.019      1.448      0.148      -0.010
0.063
C(Parameter1_Dir)[T.W]     0.0622      0.021      2.944      0.003       0.021
0.104
C(Parameter2_9am)[T.N]     0.0078      0.019      0.415      0.678      -0.029
0.045
C(Parameter2_9am)[T.S]     0.2950      0.018     16.656      0.000       0.260
0.330
C(Parameter2_9am)[T.W]     0.3338      0.020     16.662      0.000       0.295
0.373
C(Parameter2_3pm)[T.N]    -0.0859      0.020     -4.374      0.000      -0.124
-0.047
C(Parameter2_3pm)[T.S]     0.0929      0.018      5.185      0.000       0.058
0.128
C(Parameter2_3pm)[T.W]     0.1557      0.021      7.424      0.000       0.115
0.197
C(Month)[T.2]              0.0006      0.025      0.025      0.980      -0.048
0.050
C(Month)[T.3]              0.0490      0.024      2.069      0.039       0.003
0.095
C(Month)[T.4]              0.0515      0.026      2.001      0.045       0.001
0.102
C(Month)[T.5]             -0.0737      0.028     -2.654      0.008      -0.128
-0.019
C(Month)[T.6]             -0.2003      0.030     -6.667      0.000      -0.259
-0.141
C(Month)[T.7]             -0.0977      0.031     -3.127      0.002      -0.159
-0.036
C(Month)[T.8]             -0.0017      0.030     -0.056      0.956      -0.061
0.058
C(Month)[T.9]              0.0305      0.029      1.067      0.286      -0.025
```

```
0.086
C(Month)[T.10]             0.0688      0.027      2.570      0.010       0.016
0.121
C(Month)[T.11]             0.1196      0.025      4.800      0.000       0.071
0.168
C(Month)[T.12]             0.1160      0.025      4.633      0.000       0.067
0.165
C(Location)[T.3]          -0.3981      0.048     -8.331      0.000      -0.492
-0.304
C(Location)[T.4]           0.0183      0.059      0.309      0.757      -0.098
0.135
C(Location)[T.5]          -0.4558      0.047     -9.608      0.000      -0.549
-0.363
C(Location)[T.6]          -1.2169      0.047    -25.860      0.000      -1.309
-1.125
C(Location)[T.7]          -0.6196      0.048    -13.002      0.000      -0.713
-0.526
C(Location)[T.8]           0.0721      0.046      1.565      0.118      -0.018
0.162
C(Location)[T.9]          -0.3032      0.048     -6.371      0.000      -0.396
-0.210
C(Location)[T.10]         -0.5021      0.048    -10.362      0.000      -0.597
-0.407
C(Location)[T.11]         -0.1959      0.051     -3.843      0.000      -0.296
-0.096
C(Location)[T.12]         -0.2929      0.045     -6.455      0.000      -0.382
-0.204
C(Location)[T.13]         -0.7886      0.047    -16.858      0.000      -0.880
-0.697
C(Location)[T.14]         -0.1883      0.050     -3.798      0.000      -0.285
-0.091
C(Location)[T.15]         -0.5074      0.046    -10.913      0.000      -0.599
-0.416
C(Location)[T.16]         -0.4341      0.045     -9.579      0.000      -0.523
-0.345
C(Location)[T.17]         -0.3324      0.081     -4.095      0.000      -0.491
-0.173
C(Location)[T.18]         -0.5083      0.053     -9.594      0.000      -0.612
-0.404
C(Location)[T.19]         -0.3142      0.048     -6.610      0.000      -0.407
-0.221
C(Location)[T.20]         -0.7010      0.045    -15.416      0.000      -0.790
-0.612
C(Location)[T.21]         -0.6474      0.051    -12.711      0.000      -0.747
-0.548
C(Location)[T.22]         -0.1822      0.051     -3.603      0.000      -0.281
-0.083
C(Location)[T.23]         -0.6107      0.045    -13.661      0.000      -0.698
```

```
-0.523
C(Location)[T.26]          -1.0142      0.058    -17.392      0.000      -1.128
-0.900
C(Location)[T.27]          -0.7126      0.045    -15.852      0.000      -0.801
-0.624
C(Location)[T.28]          -0.5693      0.044    -12.837      0.000      -0.656
-0.482
C(Location)[T.29]          -0.6496      0.049    -13.355      0.000      -0.745
-0.554
C(Location)[T.30]          -0.2741      0.049     -5.569      0.000      -0.371
-0.178
C(Location)[T.32]          -0.1877      0.047     -4.031      0.000      -0.279
-0.096
C(Location)[T.33]          -0.2126      0.047     -4.547      0.000      -0.304
-0.121
C(Location)[T.34]          -0.6776      0.044    -15.446      0.000      -0.764
-0.592
C(Location)[T.35]          -0.3606      0.049     -7.347      0.000      -0.457
-0.264
C(Location)[T.36]          -0.9657      0.046    -21.101      0.000      -1.055
-0.876
C(Location)[T.38]          -0.3137      0.047     -6.669      0.000      -0.406
-0.221
C(Location)[T.39]          -0.3685      0.045     -8.182      0.000      -0.457
-0.280
C(Location)[T.40]          -0.4700      0.050     -9.392      0.000      -0.568
-0.372
C(Location)[T.41]          -0.3178      0.048     -6.623      0.000      -0.412
-0.224
C(Location)[T.42]          -0.0011      0.074     -0.015      0.988      -0.146
0.144
C(Location)[T.43]          -0.3643      0.048     -7.596      0.000      -0.458
-0.270
C(Location)[T.44]          -0.5795      0.045    -12.959      0.000      -0.667
-0.492
C(Location)[T.45]          -0.5674      0.045    -12.521      0.000      -0.656
-0.479
C(Location)[T.46]          -0.3751      0.047     -7.980      0.000      -0.467
-0.283
C(Location)[T.47]          -0.4042      0.046     -8.819      0.000      -0.494
-0.314
C(Location)[T.48]          -0.7452      0.045    -16.393      0.000      -0.834
-0.656
C(Location)[T.49]          -0.7399      0.057    -12.977      0.000      -0.852
-0.628
Temperature                -0.0249      0.002    -12.380      0.000      -0.029
-0.021
Parameter1_Speed            0.0147      0.001     28.841      0.000       0.014
```

```
0.016
Parameter3_9am             0.0145      0.001     18.376      0.000       0.013
0.016
Parameter4                 0.0511      0.000    125.244      0.000       0.050
0.052
Parameter5                -0.2085      0.006    -32.119      0.000      -0.221
-0.196
================================================================================
==========
       Probit Marginal Effects
=====================================
Dep. Variable:          Failure_today
Method:                          dydx
At:                           overall
================================================================================
==========
                            dy/dx    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----------
C(Parameter1_Dir)[T.N]    -0.0250      0.004     -5.813      0.000      -0.033
-0.017
C(Parameter1_Dir)[T.S]     0.0057      0.004      1.448      0.148      -0.002
0.013
C(Parameter1_Dir)[T.W]     0.0132      0.004      2.944      0.003       0.004
0.022
C(Parameter2_9am)[T.N]     0.0017      0.004      0.415      0.678      -0.006
0.009
C(Parameter2_9am)[T.S]     0.0626      0.004     16.703      0.000       0.055
0.070
C(Parameter2_9am)[T.W]     0.0708      0.004     16.711      0.000       0.063
0.079
C(Parameter2_3pm)[T.N]    -0.0182      0.004     -4.375      0.000      -0.026
-0.010
C(Parameter2_3pm)[T.S]     0.0197      0.004      5.187      0.000       0.012
0.027
C(Parameter2_3pm)[T.W]     0.0330      0.004      7.429      0.000       0.024
0.042
C(Month)[T.2]              0.0001      0.005      0.025      0.980      -0.010
0.011
C(Month)[T.3]              0.0104      0.005      2.069      0.039       0.001
0.020
C(Month)[T.4]              0.0109      0.005      2.001      0.045       0.000
0.022
C(Month)[T.5]             -0.0156      0.006     -2.654      0.008      -0.027
-0.004
C(Month)[T.6]             -0.0425      0.006     -6.671      0.000      -0.055
-0.030
```

C(Month)[T.7]	-0.0207	0.007	-3.127	0.002	-0.034	-0.008
C(Month)[T.8]	-0.0004	0.006	-0.056	0.956	-0.013	0.012
C(Month)[T.9]	0.0065	0.006	1.068	0.286	-0.005	0.018
C(Month)[T.10]	0.0146	0.006	2.570	0.010	0.003	0.026
C(Month)[T.11]	0.0254	0.005	4.801	0.000	0.015	0.036
C(Month)[T.12]	0.0246	0.005	4.634	0.000	0.014	0.035
C(Location)[T.3]	-0.0845	0.010	-8.337	0.000	-0.104	-0.065
C(Location)[T.4]	0.0039	0.013	0.309	0.757	-0.021	0.029
C(Location)[T.5]	-0.0967	0.010	-9.619	0.000	-0.116	-0.077
C(Location)[T.6]	-0.2582	0.010	-26.058	0.000	-0.278	-0.239
C(Location)[T.7]	-0.1315	0.010	-13.027	0.000	-0.151	-0.112
C(Location)[T.8]	0.0153	0.010	1.565	0.118	-0.004	0.034
C(Location)[T.9]	-0.0643	0.010	-6.376	0.000	-0.084	-0.045
C(Location)[T.10]	-0.1065	0.010	-10.374	0.000	-0.127	-0.086
C(Location)[T.11]	-0.0416	0.011	-3.843	0.000	-0.063	-0.020
C(Location)[T.12]	-0.0621	0.010	-6.459	0.000	-0.081	-0.043
C(Location)[T.13]	-0.1674	0.010	-16.914	0.000	-0.187	-0.148
C(Location)[T.14]	-0.0400	0.011	-3.799	0.000	-0.061	-0.019
C(Location)[T.15]	-0.1077	0.010	-10.932	0.000	-0.127	-0.088
C(Location)[T.16]	-0.0921	0.010	-9.589	0.000	-0.111	-0.073
C(Location)[T.17]	-0.0705	0.017	-4.096	0.000	-0.104	-0.037
C(Location)[T.18]	-0.1079	0.011	-9.604	0.000	-0.130	-0.086
C(Location)[T.19]	-0.0667	0.010	-6.614	0.000	-0.086	-0.047
C(Location)[T.20]	-0.1488	0.010	-15.461	0.000	-0.168	-0.130

C(Location)[T.21]	-0.1374	0.011	-12.734	0.000	-0.159	-0.116
C(Location)[T.22]	-0.0387	0.011	-3.603	0.000	-0.060	-0.018
C(Location)[T.23]	-0.1296	0.009	-13.691	0.000	-0.148	-0.111
C(Location)[T.26]	-0.2152	0.012	-17.453	0.000	-0.239	-0.191
C(Location)[T.27]	-0.1512	0.010	-15.902	0.000	-0.170	-0.133
C(Location)[T.28]	-0.1208	0.009	-12.865	0.000	-0.139	-0.102
C(Location)[T.29]	-0.1378	0.010	-13.380	0.000	-0.158	-0.118
C(Location)[T.30]	-0.0582	0.010	-5.571	0.000	-0.079	-0.038
C(Location)[T.32]	-0.0398	0.010	-4.032	0.000	-0.059	-0.020
C(Location)[T.33]	-0.0451	0.010	-4.548	0.000	-0.065	-0.026
C(Location)[T.34]	-0.1438	0.009	-15.488	0.000	-0.162	-0.126
C(Location)[T.35]	-0.0765	0.010	-7.352	0.000	-0.097	-0.056
C(Location)[T.36]	-0.2049	0.010	-21.213	0.000	-0.224	-0.186
C(Location)[T.38]	-0.0666	0.010	-6.674	0.000	-0.086	-0.047
C(Location)[T.39]	-0.0782	0.010	-8.188	0.000	-0.097	-0.059
C(Location)[T.40]	-0.0997	0.011	-9.407	0.000	-0.121	-0.079
C(Location)[T.41]	-0.0674	0.010	-6.626	0.000	-0.087	-0.047
C(Location)[T.42]	-0.0002	0.016	-0.015	0.988	-0.031	0.031
C(Location)[T.43]	-0.0773	0.010	-7.600	0.000	-0.097	-0.057
C(Location)[T.44]	-0.1230	0.009	-12.985	0.000	-0.142	-0.104
C(Location)[T.45]	-0.1204	0.010	-12.545	0.000	-0.139	-0.102
C(Location)[T.46]	-0.0796	0.010	-7.987	0.000	-0.099	-0.060
C(Location)[T.47]	-0.0858	0.010	-8.828	0.000	-0.105	-0.067
C(Location)[T.48]	-0.1581	0.010	-16.448	0.000	-0.177	-0.139

```
C(Location)[T.49]        -0.1570      0.012    -12.999      0.000      -0.181
-0.133
Temperature              -0.0053      0.000    -12.394      0.000      -0.006
-0.004
Parameter1_Speed          0.0031      0.000     29.147      0.000       0.003
0.003
Parameter3_9am            0.0031      0.000     18.461      0.000       0.003
0.003
Parameter4                0.0109   6.86e-05    158.233      0.000       0.011
0.011
Parameter5               -0.0442      0.001    -32.559      0.000      -0.047
-0.042
================================================================================
==========
```

[16]: {'tags': ['hide_input']}

Una vez obtenidos los resultados debemos fijarnos en los coeficientes dy/dx. En este caso una variación de una unidad de la variable Parameter4 aumentará la probabilidad de fallos en un 0.011 . Para el caso del parámetro 5, al estar estandarizado se interpreta como que al aumentar en una desviación estándar, la probabilidad de falla disminuye en 0.041 . Nuevamente la dirección norte en 2 de los 3 casos parece disminuir la probabilidad de falla (Con respecto a la dirección base), mientras que las otras direcciones parecen aumentarla. En general los parámetros no han cambiado mucho con respecto del modelo MCO a excepción del parámetro 4, que pasó de 0.0084 a 0.011

4 Pregunta 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.

[17]:
```python
logit_model = smf.logit("Failure_today ~   Temperature+ Parameter5 +Parameter4␣
 ↪+C(Month)+ C(Location)+ Parameter3_9am + Parameter2_9am + Parameter2_3pm +␣
 ↪Parameter1_Dir + Parameter1_Speed", data=df).fit(maxiter=100)
print(logit_model.summary())

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

{"tags": ["hide_input"]}
```

```
Optimization terminated successfully.
         Current function value: 0.380086
         Iterations 7
                           Logit Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:               111179
Model:                          Logit   Df Residuals:                   111110
Method:                           MLE   Df Model:                           68
```

```
Date:                Thu, 24 Apr 2025   Pseudo R-squ.:                  0.2865
Time:                        22:09:40   Log-Likelihood:                -42258.
converged:                       True   LL-Null:                       -59226.
Covariance Type:            nonrobust   LLR p-value:                     0.000
================================================================================
=======
                          coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
-------
Intercept              -7.2262      0.125    -57.918      0.000      -7.471
-6.982
C(Month)[T.2]           0.0147      0.045      0.329      0.742      -0.073
0.102
C(Month)[T.3]           0.1026      0.042      2.435      0.015       0.020
0.185
C(Month)[T.4]           0.1004      0.046      2.206      0.027       0.011
0.190
C(Month)[T.5]          -0.1292      0.049     -2.629      0.009      -0.226
-0.033
C(Month)[T.6]          -0.3641      0.053     -6.851      0.000      -0.468
-0.260
C(Month)[T.7]          -0.1818      0.055     -3.290      0.001      -0.290
-0.074
C(Month)[T.8]          -0.0124      0.054     -0.231      0.818      -0.118
0.093
C(Month)[T.9]           0.0394      0.051      0.777      0.437      -0.060
0.139
C(Month)[T.10]          0.0899      0.048      1.874      0.061      -0.004
0.184
C(Month)[T.11]          0.1967      0.045      4.405      0.000       0.109
0.284
C(Month)[T.12]          0.1953      0.045      4.349      0.000       0.107
0.283
C(Location)[T.3]       -0.7425      0.084     -8.833      0.000      -0.907
-0.578
C(Location)[T.4]       -0.0179      0.109     -0.164      0.869      -0.231
0.196
C(Location)[T.5]       -0.7924      0.084     -9.484      0.000      -0.956
-0.629
C(Location)[T.6]       -2.1973      0.083    -26.498      0.000      -2.360
-2.035
C(Location)[T.7]       -1.1219      0.084    -13.396      0.000      -1.286
-0.958
C(Location)[T.8]        0.1727      0.081      2.135      0.033       0.014
0.331
C(Location)[T.9]       -0.4517      0.083     -5.466      0.000      -0.614
-0.290
```

C(Location)[T.10]	-0.9218	0.086	-10.702	0.000	-1.091	-0.753
C(Location)[T.11]	-0.4235	0.092	-4.605	0.000	-0.604	-0.243
C(Location)[T.12]	-0.4901	0.079	-6.220	0.000	-0.645	-0.336
C(Location)[T.13]	-1.4105	0.081	-17.386	0.000	-1.570	-1.251
C(Location)[T.14]	-0.2499	0.087	-2.873	0.004	-0.420	-0.079
C(Location)[T.15]	-0.8488	0.081	-10.417	0.000	-1.008	-0.689
C(Location)[T.16]	-0.8103	0.080	-10.097	0.000	-0.968	-0.653
C(Location)[T.17]	-0.5101	0.145	-3.517	0.000	-0.794	-0.226
C(Location)[T.18]	-0.9212	0.093	-9.915	0.000	-1.103	-0.739
C(Location)[T.19]	-0.5655	0.084	-6.773	0.000	-0.729	-0.402
C(Location)[T.20]	-1.2498	0.080	-15.594	0.000	-1.407	-1.093
C(Location)[T.21]	-1.1764	0.091	-12.890	0.000	-1.355	-0.998
C(Location)[T.22]	-0.3585	0.092	-3.914	0.000	-0.538	-0.179
C(Location)[T.23]	-1.0895	0.078	-13.996	0.000	-1.242	-0.937
C(Location)[T.26]	-1.8169	0.104	-17.475	0.000	-2.021	-1.613
C(Location)[T.27]	-1.2540	0.079	-15.891	0.000	-1.409	-1.099
C(Location)[T.28]	-0.9893	0.077	-12.802	0.000	-1.141	-0.838
C(Location)[T.29]	-1.1938	0.086	-13.862	0.000	-1.363	-1.025
C(Location)[T.30]	-0.4808	0.087	-5.546	0.000	-0.651	-0.311
C(Location)[T.32]	-0.3035	0.081	-3.734	0.000	-0.463	-0.144
C(Location)[T.33]	-0.3486	0.082	-4.243	0.000	-0.510	-0.188
C(Location)[T.34]	-1.2239	0.077	-15.982	0.000	-1.374	-1.074
C(Location)[T.35]	-0.6211	0.086	-7.185	0.000	-0.791	-0.452
C(Location)[T.36]	-1.7332	0.081	-21.431	0.000	-1.892	-1.575

C(Location)[T.38]	-0.5334	0.082	-6.492	0.000	-0.694	-0.372
C(Location)[T.39]	-0.6674	0.080	-8.351	0.000	-0.824	-0.511
C(Location)[T.40]	-0.7220	0.088	-8.184	0.000	-0.895	-0.549
C(Location)[T.41]	-0.5882	0.085	-6.931	0.000	-0.755	-0.422
C(Location)[T.42]	-0.0349	0.137	-0.255	0.799	-0.303	0.233
C(Location)[T.43]	-0.7184	0.086	-8.386	0.000	-0.886	-0.550
C(Location)[T.44]	-1.0283	0.078	-13.195	0.000	-1.181	-0.876
C(Location)[T.45]	-1.0176	0.080	-12.791	0.000	-1.174	-0.862
C(Location)[T.46]	-0.6532	0.082	-7.932	0.000	-0.815	-0.492
C(Location)[T.47]	-0.7155	0.080	-8.981	0.000	-0.872	-0.559
C(Location)[T.48]	-1.3122	0.080	-16.396	0.000	-1.469	-1.155
C(Location)[T.49]	-1.3796	0.105	-13.130	0.000	-1.586	-1.174
Parameter2_9am[T.N]	0.0217	0.034	0.638	0.524	-0.045	0.088
Parameter2_9am[T.S]	0.5325	0.032	16.720	0.000	0.470	0.595
Parameter2_9am[T.W]	0.5981	0.036	16.765	0.000	0.528	0.668
Parameter2_3pm[T.N]	-0.1614	0.035	-4.580	0.000	-0.231	-0.092
Parameter2_3pm[T.S]	0.1524	0.032	4.790	0.000	0.090	0.215
Parameter2_3pm[T.W]	0.2637	0.037	7.060	0.000	0.191	0.337
Parameter1_Dir[T.N]	-0.2178	0.036	-5.971	0.000	-0.289	-0.146
Parameter1_Dir[T.S]	0.0361	0.033	1.093	0.274	-0.029	0.101
Parameter1_Dir[T.W]	0.0977	0.038	2.597	0.009	0.024	0.172
Temperature	-0.0496	0.004	-13.625	0.000	-0.057	-0.042
Parameter5	-0.3609	0.011	-31.561	0.000	-0.383	-0.338
Parameter4	0.0902	0.001	119.354	0.000	0.089	0.092

```
Parameter3_9am          0.0247      0.001     17.763      0.000       0.022
0.027
Parameter1_Speed        0.0258      0.001     28.701      0.000       0.024
0.028
================================================================================
=======
        Logit Marginal Effects
=====================================
Dep. Variable:          Failure_today
Method:                          dydx
At:                           overall
================================================================================
=======
                         dy/dx    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
-------
C(Month)[T.2]           0.0018      0.005      0.329      0.742      -0.009
0.012
C(Month)[T.3]           0.0123      0.005      2.435      0.015       0.002
0.022
C(Month)[T.4]           0.0121      0.005      2.206      0.027       0.001
0.023
C(Month)[T.5]          -0.0155      0.006     -2.629      0.009      -0.027
-0.004
C(Month)[T.6]          -0.0438      0.006     -6.855      0.000      -0.056
-0.031
C(Month)[T.7]          -0.0219      0.007     -3.290      0.001      -0.035
-0.009
C(Month)[T.8]          -0.0015      0.006     -0.231      0.818      -0.014
0.011
C(Month)[T.9]           0.0047      0.006      0.777      0.437      -0.007
0.017
C(Month)[T.10]          0.0108      0.006      1.874      0.061      -0.000
0.022
C(Month)[T.11]          0.0237      0.005      4.406      0.000       0.013
0.034
C(Month)[T.12]          0.0235      0.005      4.350      0.000       0.013
0.034
C(Location)[T.3]       -0.0893      0.010     -8.843      0.000      -0.109
-0.070
C(Location)[T.4]       -0.0022      0.013     -0.164      0.869      -0.028
0.024
C(Location)[T.5]       -0.0953      0.010     -9.499      0.000      -0.115
-0.076
C(Location)[T.6]       -0.2644      0.010    -26.785      0.000      -0.284
-0.245
C(Location)[T.7]       -0.1350      0.010    -13.432      0.000      -0.155
```

						-0.115
C(Location)[T.8]	0.0208	0.010	2.135	0.033	0.002	0.040
C(Location)[T.9]	-0.0543	0.010	-5.470	0.000	-0.074	-0.035
C(Location)[T.10]	-0.1109	0.010	-10.721	0.000	-0.131	-0.091
C(Location)[T.11]	-0.0510	0.011	-4.606	0.000	-0.073	-0.029
C(Location)[T.12]	-0.0590	0.009	-6.225	0.000	-0.078	-0.040
C(Location)[T.13]	-0.1697	0.010	-17.468	0.000	-0.189	-0.151
C(Location)[T.14]	-0.0301	0.010	-2.874	0.004	-0.051	-0.010
C(Location)[T.15]	-0.1021	0.010	-10.438	0.000	-0.121	-0.083
C(Location)[T.16]	-0.0975	0.010	-10.116	0.000	-0.116	-0.079
C(Location)[T.17]	-0.0614	0.017	-3.518	0.000	-0.096	-0.027
C(Location)[T.18]	-0.1108	0.011	-9.932	0.000	-0.133	-0.089
C(Location)[T.19]	-0.0680	0.010	-6.779	0.000	-0.088	-0.048
C(Location)[T.20]	-0.1504	0.010	-15.658	0.000	-0.169	-0.132
C(Location)[T.21]	-0.1415	0.011	-12.920	0.000	-0.163	-0.120
C(Location)[T.22]	-0.0431	0.011	-3.915	0.000	-0.065	-0.022
C(Location)[T.23]	-0.1311	0.009	-14.040	0.000	-0.149	-0.113
C(Location)[T.26]	-0.2186	0.012	-17.557	0.000	-0.243	-0.194
C(Location)[T.27]	-0.1509	0.009	-15.959	0.000	-0.169	-0.132
C(Location)[T.28]	-0.1190	0.009	-12.842	0.000	-0.137	-0.101
C(Location)[T.29]	-0.1436	0.010	-13.899	0.000	-0.164	-0.123
C(Location)[T.30]	-0.0578	0.010	-5.549	0.000	-0.078	-0.037
C(Location)[T.32]	-0.0365	0.010	-3.735	0.000	-0.056	-0.017
C(Location)[T.33]	-0.0419	0.010	-4.244	0.000	-0.061	-0.023
C(Location)[T.34]	-0.1472	0.009	-16.047	0.000	-0.165	

						-0.129
C(Location)[T.35]	-0.0747	0.010	-7.191	0.000	-0.095	-0.054
C(Location)[T.36]	-0.2085	0.010	-21.595	0.000	-0.227	-0.190
C(Location)[T.38]	-0.0642	0.010	-6.498	0.000	-0.084	-0.045
C(Location)[T.39]	-0.0803	0.010	-8.362	0.000	-0.099	-0.061
C(Location)[T.40]	-0.0869	0.011	-8.196	0.000	-0.108	-0.066
C(Location)[T.41]	-0.0708	0.010	-6.936	0.000	-0.091	-0.051
C(Location)[T.42]	-0.0042	0.016	-0.255	0.799	-0.036	0.028
C(Location)[T.43]	-0.0864	0.010	-8.394	0.000	-0.107	-0.066
C(Location)[T.44]	-0.1237	0.009	-13.234	0.000	-0.142	-0.105
C(Location)[T.45]	-0.1224	0.010	-12.826	0.000	-0.141	-0.104
C(Location)[T.46]	-0.0786	0.010	-7.942	0.000	-0.098	-0.059
C(Location)[T.47]	-0.0861	0.010	-8.994	0.000	-0.105	-0.067
C(Location)[T.48]	-0.1579	0.010	-16.471	0.000	-0.177	-0.139
C(Location)[T.49]	-0.1660	0.013	-13.160	0.000	-0.191	-0.141
Parameter2_9am[T.N]	0.0026	0.004	0.638	0.524	-0.005	0.011
Parameter2_9am[T.S]	0.0641	0.004	16.773	0.000	0.057	0.072
Parameter2_9am[T.W]	0.0720	0.004	16.824	0.000	0.064	0.080
Parameter2_3pm[T.N]	-0.0194	0.004	-4.581	0.000	-0.028	-0.011
Parameter2_3pm[T.S]	0.0183	0.004	4.792	0.000	0.011	0.026
Parameter2_3pm[T.W]	0.0317	0.004	7.067	0.000	0.023	0.041
Parameter1_Dir[T.N]	-0.0262	0.004	-5.975	0.000	-0.035	-0.018
Parameter1_Dir[T.S]	0.0043	0.004	1.093	0.274	-0.003	0.012
Parameter1_Dir[T.W]	0.0118	0.005	2.597	0.009	0.003	0.021
Temperature	-0.0060	0.000	-13.652	0.000	-0.007	

```
-0.005
Parameter5             -0.0434      0.001    -32.094      0.000      -0.046
-0.041
Parameter4              0.0108   6.87e-05    157.918      0.000       0.011
0.011
Parameter3_9am          0.0030      0.000     17.857      0.000       0.003
0.003
Parameter1_Speed        0.0031      0.000     29.104      0.000       0.003
0.003
================================================================================
=======
```

[17]: {'tags': ['hide_input']}

[18]:
```
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[57:68 , ])

{"tags": ["hide_input"]}
```

```
Odds Ratios
                     Odds Ratio        5%       95%
Parameter2_9am[T.W]    1.695810  1.950345  1.818630
Parameter2_3pm[T.N]    0.794104  0.911787  0.850913
Parameter2_3pm[T.S]    1.094233  1.239606  1.164654
Parameter2_3pm[T.W]    1.209857  1.400624  1.301751
Parameter1_Dir[T.N]    0.748795  0.863886  0.804284
Parameter1_Dir[T.S]    0.971744  1.106229  1.036809
Parameter1_Dir[T.W]    1.024261  1.187095  1.102676
Temperature            0.944845  0.958425  0.951610
Parameter5             0.681624  0.712869  0.697071
Parameter4             1.092747  1.095988  1.094366
Parameter3_9am         1.022256  1.027853  1.025051
```

[18]: {'tags': ['hide_input']}

No se observan grandes variaciones con respecto al modelo probit y la forma en la que se interpretan los resultados es la misma. Los parámetros 4, 3 y 1 parecen aumentar la probabilidad de fallos, mientras que el parámetro 5 la disminuye. Con respecto a los Odds ratios la interpretación es como sigue: Si el parámetro 4 aumenta en una unidad, las probabilidades de que se produzca un fallo se multiplican por 1.09 . Por otro lado, si el parámetro 5 aumenta una desviación estándar, las probabilidades de que ocurra una falla se multiplican por 0.7 .

5 PREGUNTA 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?
Como se puede apreciar en la tabla siguiente, los 3 modelos ofrecen resuestas muy similares, aunque podemos descartar el método MCO como el más adecuado porque está hecho para predecir variables continuas, mientras que logit y probit se especializan en variables binarias. Entre logit y probit no hay mucha diferencia en los resultados, pero elegiría el modelo logit por la posibilidad de analizar los odds ratios.
El parámetro 1 de velocidad, el parámetro 3 a las 9am, el parámetro 4 y el parámetro 5, resultaron ser robustos a la especificación, pues sus magnitudes y signos se mantuvieron similares a traves de los 3 modelos. Pasando a las variables categóricas, los meses y las locaciones también mantuvieron valores similares.

[19]:
```
results_df = pd.DataFrame({
    'Modelo MCO': model.params,
    'Modelo Probit': probit_model.get_margeff().summary_frame()["dy/dx"],
    'Modelo Logit': logit_model.get_margeff().summary_frame()["dy/dx"]
})

results_df = results_df.round(3)

results_df.tail(12)

{"tags": ["hide_input"]}
```

[19]:
```
{'tags': ['hide_input']}
```

6 PREGUNTA 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.

[20]:
```
df5=df.
 ↪drop(["Parameter1_Dir","Parameter2_9am","Parameter2_3pm","Day","Month","Year"],axis=1)
df5["Date"]=df4
df5['Date'] = pd.to_datetime(df5['Date'])
df5['mes'] = df5['Date'].dt.to_period('M')


df_m = df5.groupby(["mes","Location"]).agg({
    'Leakage': 'mean',
    'Parameter1_Speed': 'mean',
```

```
    'Parameter3_9am': 'mean',
    'Parameter3_3pm': 'mean',
    'Parameter4': 'mean',
    'Parameter5': 'mean',
    'Temperature': 'mean',
    'Failure_today': 'sum'
}).reset_index()

df_m['Month'] = df_m['mes'].dt.month
df_m

{"tags": ["hide_input"]}
```

[20]: {'tags': ['hide_input']}

[21]:
```
poisson = smf.glm(formula='Failure_today ~ C(Location) + Parameter3_9am+␣
↪Parameter3_3pm + Parameter4 + Parameter5 + C(Month)', data=df_m, family=sm.
↪families.Poisson()).fit()

print(poisson.summary())
print("Resultados exponenciales")
print(np.exp(poisson.params))

{"tags": ["hide_input"]}
```

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:                 4076
Model:                            GLM   Df Residuals:                     4017
Model Family:                 Poisson   Df Model:                           58
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -9191.2
Date:                Thu, 24 Apr 2025   Deviance:                       4857.4
Time:                        22:11:02   Pearson chi2:                 4.36e+03
No. Iterations:                     5   Pseudo R-squ. (CS):             0.8501
Covariance Type:            nonrobust
================================================================================
=====
                        coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
-----
Intercept            -1.1452      0.081    -14.201      0.000      -1.303
-0.987
C(Location)[T.3]     -0.6954      0.060    -11.588      0.000      -0.813
-0.578
C(Location)[T.4]     -0.2370      0.081     -2.940      0.003      -0.395
-0.079
```

C(Location)[T.5]	-0.7190	0.062	-11.655	0.000	-0.840	-0.598
C(Location)[T.6]	-1.0548	0.066	-16.100	0.000	-1.183	-0.926
C(Location)[T.7]	-0.6717	0.061	-11.016	0.000	-0.791	-0.552
C(Location)[T.8]	-0.2861	0.059	-4.832	0.000	-0.402	-0.170
C(Location)[T.9]	-0.5749	0.057	-10.138	0.000	-0.686	-0.464
C(Location)[T.10]	-0.5379	0.063	-8.483	0.000	-0.662	-0.414
C(Location)[T.11]	-0.3173	0.068	-4.650	0.000	-0.451	-0.184
C(Location)[T.12]	-0.3681	0.059	-6.291	0.000	-0.483	-0.253
C(Location)[T.13]	-0.9359	0.059	-15.967	0.000	-1.051	-0.821
C(Location)[T.14]	-0.7190	0.059	-12.168	0.000	-0.835	-0.603
C(Location)[T.15]	-0.5420	0.064	-8.408	0.000	-0.668	-0.416
C(Location)[T.16]	-0.4715	0.058	-8.106	0.000	-0.585	-0.357
C(Location)[T.17]	-1.1410	0.107	-10.671	0.000	-1.351	-0.931
C(Location)[T.18]	-0.9390	0.067	-13.989	0.000	-1.071	-0.807
C(Location)[T.19]	-0.4543	0.065	-7.039	0.000	-0.581	-0.328
C(Location)[T.20]	-0.5249	0.064	-8.176	0.000	-0.651	-0.399
C(Location)[T.21]	-0.6163	0.070	-8.839	0.000	-0.753	-0.480
C(Location)[T.22]	-0.5865	0.071	-8.254	0.000	-0.726	-0.447
C(Location)[T.23]	-0.5027	0.059	-8.571	0.000	-0.618	-0.388
C(Location)[T.26]	-0.8073	0.079	-10.157	0.000	-0.963	-0.652
C(Location)[T.27]	-0.7010	0.059	-11.800	0.000	-0.817	-0.585
C(Location)[T.28]	-0.7000	0.062	-11.290	0.000	-0.821	-0.578
C(Location)[T.29]	-0.4117	0.061	-6.780	0.000	-0.531	-0.293
C(Location)[T.30]	-0.4029	0.064	-6.269	0.000	-0.529	-0.277

C(Location)[T.32]	-0.2848	0.059	-4.790	0.000	-0.401	-0.168
C(Location)[T.33]	-0.1312	0.061	-2.133	0.033	-0.252	-0.011
C(Location)[T.34]	-0.5795	0.057	-10.103	0.000	-0.692	-0.467
C(Location)[T.35]	-0.7565	0.063	-11.991	0.000	-0.880	-0.633
C(Location)[T.36]	-0.7237	0.061	-11.776	0.000	-0.844	-0.603
C(Location)[T.38]	-0.2632	0.060	-4.385	0.000	-0.381	-0.146
C(Location)[T.39]	-0.1236	0.062	-1.992	0.046	-0.245	-0.002
C(Location)[T.40]	-0.8367	0.065	-12.814	0.000	-0.965	-0.709
C(Location)[T.41]	-0.5707	0.062	-9.266	0.000	-0.691	-0.450
C(Location)[T.42]	-0.4654	0.107	-4.353	0.000	-0.675	-0.256
C(Location)[T.43]	-0.4869	0.061	-7.993	0.000	-0.606	-0.368
C(Location)[T.44]	-0.7834	0.058	-13.615	0.000	-0.896	-0.671
C(Location)[T.45]	-0.6204	0.057	-10.838	0.000	-0.733	-0.508
C(Location)[T.46]	-0.5418	0.062	-8.738	0.000	-0.663	-0.420
C(Location)[T.47]	-0.4680	0.058	-8.037	0.000	-0.582	-0.354
C(Location)[T.48]	-0.7412	0.062	-11.947	0.000	-0.863	-0.620
C(Location)[T.49]	-0.6871	0.085	-8.039	0.000	-0.855	-0.520
C(Month)[T.2]	-0.1008	0.033	-3.027	0.002	-0.166	-0.036
C(Month)[T.3]	0.0769	0.032	2.379	0.017	0.014	0.140
C(Month)[T.4]	0.1811	0.036	5.034	0.000	0.111	0.252
C(Month)[T.5]	0.0517	0.036	1.421	0.155	-0.020	0.123
C(Month)[T.6]	-0.1150	0.038	-3.015	0.003	-0.190	-0.040
C(Month)[T.7]	0.0879	0.038	2.344	0.019	0.014	0.161
C(Month)[T.8]	0.2472	0.035	7.135	0.000	0.179	0.315

```
C(Month)[T.9]           0.2598      0.034      7.630      0.000       0.193
0.327
C(Month)[T.10]          0.3780      0.035     10.670      0.000       0.309
0.447
C(Month)[T.11]          0.2574      0.033      7.801      0.000       0.193
0.322
C(Month)[T.12]          0.1688      0.033      5.105      0.000       0.104
0.234
Parameter3_9am          0.0415      0.004     11.832      0.000       0.035
0.048
Parameter3_3pm         -0.0312      0.003     -9.523      0.000      -0.038
-0.025
Parameter4              0.0540      0.001     61.507      0.000       0.052
0.056
Parameter5             -0.3772      0.018    -20.774      0.000      -0.413
-0.342
================================================================================
=====
Resultados exponenciales
Intercept              0.318152
C(Location)[T.3]       0.498879
C(Location)[T.4]       0.788991
C(Location)[T.5]       0.487255
C(Location)[T.6]       0.348253
C(Location)[T.7]       0.510850
C(Location)[T.8]       0.751198
C(Location)[T.9]       0.562736
C(Location)[T.10]      0.583975
C(Location)[T.11]      0.728093
C(Location)[T.12]      0.692017
C(Location)[T.13]      0.392243
C(Location)[T.14]      0.487221
C(Location)[T.15]      0.581609
C(Location)[T.16]      0.624084
C(Location)[T.17]      0.319506
C(Location)[T.18]      0.391033
C(Location)[T.19]      0.634865
C(Location)[T.20]      0.591611
C(Location)[T.21]      0.539925
C(Location)[T.22]      0.556293
C(Location)[T.23]      0.604912
C(Location)[T.26]      0.446062
C(Location)[T.27]      0.496097
C(Location)[T.28]      0.496609
C(Location)[T.29]      0.662521
C(Location)[T.30]      0.668361
C(Location)[T.32]      0.752146
C(Location)[T.33]      0.877078
```

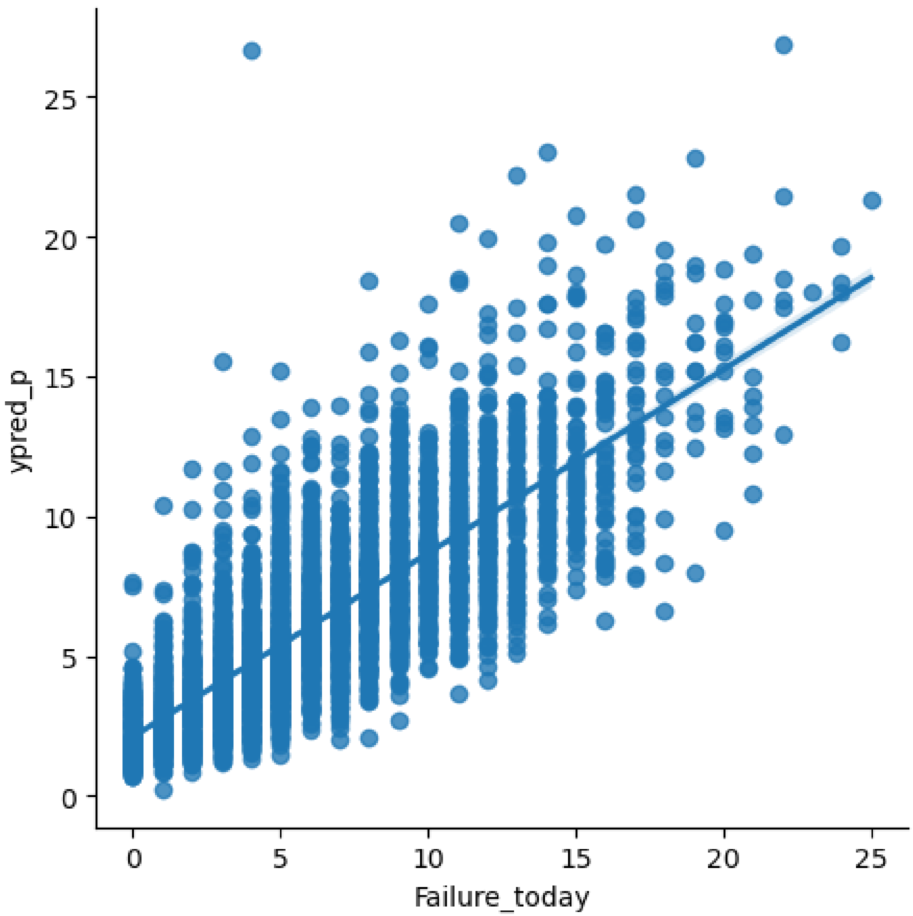
```
C(Location)[T.34]     0.560176
C(Location)[T.35]     0.469313
C(Location)[T.36]     0.484955
C(Location)[T.38]     0.768580
C(Location)[T.39]     0.883741
C(Location)[T.40]     0.433150
C(Location)[T.41]     0.565151
C(Location)[T.42]     0.627909
C(Location)[T.43]     0.614513
C(Location)[T.44]     0.456862
C(Location)[T.45]     0.537712
C(Location)[T.46]     0.581674
C(Location)[T.47]     0.626231
C(Location)[T.48]     0.476545
C(Location)[T.49]     0.503017
C(Month)[T.2]         0.904094
C(Month)[T.3]         1.079958
C(Month)[T.4]         1.198482
C(Month)[T.5]         1.053107
C(Month)[T.6]         0.891403
C(Month)[T.7]         1.091926
C(Month)[T.8]         1.280397
C(Month)[T.9]         1.296661
C(Month)[T.10]        1.459418
C(Month)[T.11]        1.293582
C(Month)[T.12]        1.183914
Parameter3_9am        1.042361
Parameter3_3pm        0.969243
Parameter4            1.055479
Parameter5            0.685780
dtype: float64
```

[21]: {'tags': ['hide_input']}

[22]:
```
predictions = poisson.predict(df_m)
df_m['ypred_p'] = predictions
sns.lmplot(data=df_m, x='Failure_today', y='ypred_p')

{"tags": ["hide_input"]}
```

[22]: {'tags': ['hide_input']}



Eliminamos el parámetro 1 de velocidad porque su valor p era demasiado alto. Los coeficientes e^B representan el multiplicador que se debe aplicar a la media esperada cuando la variable aumenta en una unidad (O una desviación estándar en el caso del Parámetro 5). Por ejemplo, si el parámetro 4 aumenta en una unidad, la media esperada de los fallos se multiplicará por 1.055 . Estos resultados se condicen con los anteriores en el sentido de si aumentan o disminuyen la probabilidad de fallas.

7 PREGUNTA 7

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

Podemos ver que el estadístico de sobredispersión es cercano a 1, por lo que el modelo Poisson puede ser adecuado. Aunque es mayor que uno, por lo que tal vez se podría considerar un poco de sobredispersión.

[23]:
```
residuos_pearson = poisson.resid_pearson

pearson_chi2 = np.sum(residuos_pearson**2)
```

```
grados_libertad = poisson.df_resid

dispersion = pearson_chi2 / grados_libertad
print("Estadístico de dispersión:", dispersion)

{"tags": ["hide_input"]}
```

```
Estadístico de dispersión: 1.0855149563214965
```

[23]:
```
{'tags': ['hide_input']}
```

Podemos ver que el estimador para ln(Alfa) es 0.0094, por lo que usaremos Alfa=1.0094

[24]:
```
aux=((df_m['Failure_today']-predictions)**2-predictions)/predictions
auxr=sm.OLS(aux,predictions).fit()
print(auxr.summary())
Alfa=np.exp(auxr.params[0])
print("")
print("Alfa es ",Alfa)

{"tags": ["hide_input"]}
```

```
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                      y   R-squared (uncentered):
0.002
Model:                            OLS   Adj. R-squared (uncentered):
0.002
Method:                 Least Squares   F-statistic:
9.670
Date:                Thu, 24 Apr 2025   Prob (F-statistic):
0.00189
Time:                        22:11:03   Log-Likelihood:
-7638.9
No. Observations:                4076   AIC:
1.528e+04
Df Residuals:                    4075   BIC:
1.529e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0109      0.003      3.110      0.002       0.004       0.018
==============================================================================
Omnibus:                     3390.860   Durbin-Watson:                   1.852
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           112809.210
Skew:                           3.811   Prob(JB):                         0.00
```

```
Kurtosis:                      27.620   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.

Alfa es  1.01093381777703
```

[24]: {'tags': ['hide_input']}

8 PREGUNTA 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.

Los resultados se interpretan de la misma forma que los del modelo poisson: Con los coeficientes exponenciales. En este caso, si el parámetro 4 aumenta una unidad, la media esperada se multiplicará por 1.063

[25]:
```python
negbin= smf.glm(formula='Failure_today ~ C(Location) + Parameter3_9am+␣
↪Parameter3_3pm + Parameter4 + Parameter5 + C(Month)',
    data=df_m,
    family=sm.families.NegativeBinomial(alpha=Alfa)
).fit()
print(negbin.summary())
print("Resultados exponenciales")
print(np.exp(negbin.params))

{"tags": ["hide_input"]}
```

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:          Failure_today   No. Observations:                 4076
Model:                            GLM   Df Residuals:                     4017
Model Family:        NegativeBinomial   Df Model:                           58
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -11168.
Date:                Thu, 24 Apr 2025   Deviance:                       1116.7
Time:                        22:11:03   Pearson chi2:                     802.
No. Iterations:                    10   Pseudo R-squ. (CS):             0.2637
Covariance Type:            nonrobust
================================================================================
=====
                        coef    std err          z      P>|z|      [0.025
0.975]
```

```
--------------------------------------------------------------------------------
-----
Intercept             -1.4438      0.213     -6.775      0.000      -1.862
-1.026
C(Location)[T.3]      -0.6393      0.159     -4.009      0.000      -0.952
-0.327
C(Location)[T.4]      -0.2915      0.178     -1.634      0.102      -0.641
0.058
C(Location)[T.5]      -0.8155      0.161     -5.068      0.000      -1.131
-0.500
C(Location)[T.6]      -1.1207      0.183     -6.114      0.000      -1.480
-0.761
C(Location)[T.7]      -0.6288      0.161     -3.903      0.000      -0.944
-0.313
C(Location)[T.8]      -0.3734      0.161     -2.313      0.021      -0.690
-0.057
C(Location)[T.9]      -0.7257      0.167     -4.341      0.000      -1.053
-0.398
C(Location)[T.10]     -0.5121      0.164     -3.118      0.002      -0.834
-0.190
C(Location)[T.11]     -0.2622      0.166     -1.575      0.115      -0.588
0.064
C(Location)[T.12]     -0.4269      0.168     -2.533      0.011      -0.757
-0.097
C(Location)[T.13]     -1.0268      0.168     -6.119      0.000      -1.356
-0.698
C(Location)[T.14]     -1.1031      0.171     -6.459      0.000      -1.438
-0.768
C(Location)[T.15]     -0.5748      0.180     -3.193      0.001      -0.928
-0.222
C(Location)[T.16]     -0.5185      0.163     -3.181      0.001      -0.838
-0.199
C(Location)[T.17]     -1.5652      0.268     -5.841      0.000      -2.090
-1.040
C(Location)[T.18]     -0.9893      0.184     -5.378      0.000      -1.350
-0.629
C(Location)[T.19]     -0.5028      0.178     -2.830      0.005      -0.851
-0.155
C(Location)[T.20]     -0.5565      0.177     -3.138      0.002      -0.904
-0.209
C(Location)[T.21]     -0.5876      0.165     -3.571      0.000      -0.910
-0.265
C(Location)[T.22]     -0.6280      0.173     -3.634      0.000      -0.967
-0.289
C(Location)[T.23]     -0.5581      0.171     -3.257      0.001      -0.894
-0.222
C(Location)[T.26]     -0.8310      0.202     -4.118      0.000      -1.227
-0.435
```

C(Location)[T.27]	-0.7180	0.167	-4.305	0.000	-1.045	-0.391
C(Location)[T.28]	-0.7453	0.179	-4.175	0.000	-1.095	-0.395
C(Location)[T.29]	-0.4000	0.162	-2.471	0.013	-0.717	-0.083
C(Location)[T.30]	-0.4886	0.166	-2.944	0.003	-0.814	-0.163
C(Location)[T.32]	-0.4341	0.159	-2.733	0.006	-0.745	-0.123
C(Location)[T.33]	-0.2393	0.167	-1.430	0.153	-0.567	0.089
C(Location)[T.34]	-0.6928	0.170	-4.078	0.000	-1.026	-0.360
C(Location)[T.35]	-0.8164	0.160	-5.091	0.000	-1.131	-0.502
C(Location)[T.36]	-0.7698	0.168	-4.574	0.000	-1.100	-0.440
C(Location)[T.38]	-0.3102	0.171	-1.815	0.069	-0.645	0.025
C(Location)[T.39]	-0.1114	0.176	-0.633	0.527	-0.456	0.233
C(Location)[T.40]	-1.0475	0.174	-6.016	0.000	-1.389	-0.706
C(Location)[T.41]	-0.5833	0.159	-3.676	0.000	-0.894	-0.272
C(Location)[T.42]	-0.5244	0.220	-2.381	0.017	-0.956	-0.093
C(Location)[T.43]	-0.4122	0.160	-2.578	0.010	-0.726	-0.099
C(Location)[T.44]	-0.9217	0.166	-5.550	0.000	-1.247	-0.596
C(Location)[T.45]	-0.6917	0.159	-4.360	0.000	-1.003	-0.381
C(Location)[T.46]	-0.5704	0.169	-3.377	0.001	-0.901	-0.239
C(Location)[T.47]	-0.6298	0.167	-3.762	0.000	-0.958	-0.302
C(Location)[T.48]	-0.7538	0.171	-4.417	0.000	-1.088	-0.419
C(Location)[T.49]	-0.6700	0.183	-3.669	0.000	-1.028	-0.312
C(Month)[T.2]	-0.1466	0.087	-1.684	0.092	-0.317	0.024
C(Month)[T.3]	0.0303	0.088	0.345	0.730	-0.142	0.202
C(Month)[T.4]	0.1468	0.098	1.500	0.134	-0.045	0.339

```
C(Month)[T.5]        -0.0126      0.100     -0.126      0.900      -0.209
0.183
C(Month)[T.6]        -0.2253      0.107     -2.114      0.035      -0.434
-0.016
C(Month)[T.7]        -0.0066      0.105     -0.062      0.950      -0.213
0.200
C(Month)[T.8]         0.1438      0.098      1.470      0.142      -0.048
0.336
C(Month)[T.9]         0.2617      0.093      2.809      0.005       0.079
0.444
C(Month)[T.10]        0.3945      0.095      4.171      0.000       0.209
0.580
C(Month)[T.11]        0.3265      0.087      3.753      0.000       0.156
0.497
C(Month)[T.12]        0.2604      0.088      2.959      0.003       0.088
0.433
Parameter3_9am        0.0503      0.009      5.528      0.000       0.032
0.068
Parameter3_3pm       -0.0416      0.009     -4.781      0.000      -0.059
-0.025
Parameter4            0.0613      0.002     27.395      0.000       0.057
0.066
Parameter5           -0.4304      0.051     -8.375      0.000      -0.531
-0.330
==============================================================================
=====
Resultados exponenciales
Intercept            0.236020
C(Location)[T.3]     0.527648
C(Location)[T.4]     0.747123
C(Location)[T.5]     0.442413
C(Location)[T.6]     0.326066
C(Location)[T.7]     0.533254
C(Location)[T.8]     0.688409
C(Location)[T.9]     0.484001
C(Location)[T.10]    0.599256
C(Location)[T.11]    0.769379
C(Location)[T.12]    0.652541
C(Location)[T.13]    0.358143
C(Location)[T.14]    0.331842
C(Location)[T.15]    0.562842
C(Location)[T.16]    0.595408
C(Location)[T.17]    0.209056
C(Location)[T.18]    0.371826
C(Location)[T.19]    0.604859
C(Location)[T.20]    0.573235
C(Location)[T.21]    0.555685
C(Location)[T.22]    0.533677
```

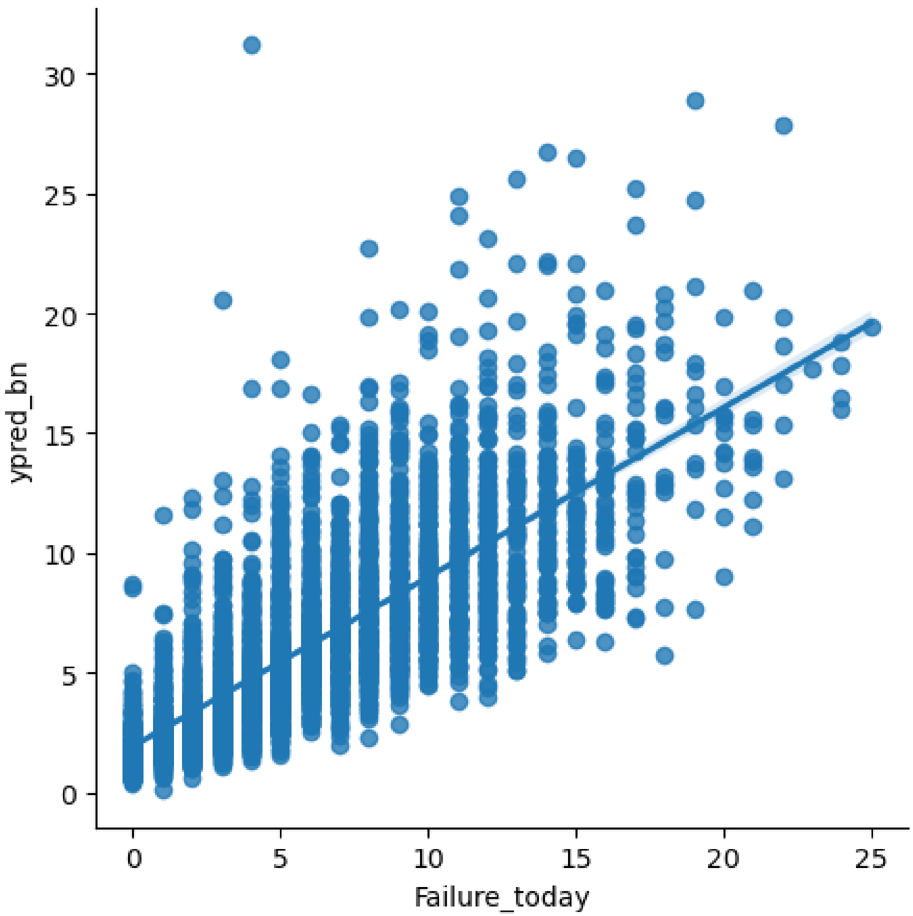
```
C(Location)[T.23]    0.572287
C(Location)[T.26]    0.435599
C(Location)[T.27]    0.487708
C(Location)[T.28]    0.474577
C(Location)[T.29]    0.670303
C(Location)[T.30]    0.613494
C(Location)[T.32]    0.647864
C(Location)[T.33]    0.787217
C(Location)[T.34]    0.500167
C(Location)[T.35]    0.442025
C(Location)[T.36]    0.463120
C(Location)[T.38]    0.733303
C(Location)[T.39]    0.894606
C(Location)[T.40]    0.350816
C(Location)[T.41]    0.558060
C(Location)[T.42]    0.591938
C(Location)[T.43]    0.662161
C(Location)[T.44]    0.397853
C(Location)[T.45]    0.500713
C(Location)[T.46]    0.565323
C(Location)[T.47]    0.532682
C(Location)[T.48]    0.470557
C(Location)[T.49]    0.511705
C(Month)[T.2]        0.863598
C(Month)[T.3]        1.030740
C(Month)[T.4]        1.158142
C(Month)[T.5]        0.987463
C(Month)[T.6]        0.798308
C(Month)[T.7]        0.993471
C(Month)[T.8]        1.154689
C(Month)[T.9]        1.299152
C(Month)[T.10]       1.483676
C(Month)[T.11]       1.386126
C(Month)[T.12]       1.297397
Parameter3_9am       1.051614
Parameter3_3pm       0.959231
Parameter4           1.063271
Parameter5           0.650257
dtype: float64
```

[25]: {'tags': ['hide_input']}

[26]:
```
df_m['ypred_bn'] = negbin.predict(df_m)
sns.lmplot(data=df_m, x='Failure_today', y='ypred_bn')

{"tags": ["hide_input"]}
```

[26]: {'tags': ['hide_input']}



[ ]:

9 PREGUNTA 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?

Como se puede apreciar en la tablas siguientes, los valores predichos son similares aunque el modelo poisson resultó ser más preciso al predecir los resultados de nuestro data set y, dado que nuestro modelo no presenta sobredispersión, nos quedaremos con el modelo poisson.

[27]:
```
df_m.loc[:,"ypred_p":"ypred_bn"]

{"tags": ["hide_input"]}
```

[27]:
```
{'tags': ['hide_input']}
```

[28]:
```
df_m["y-ypred_p"]=(df_m["Failure_today"]-df_m["ypred_p"]).abs()
df_m["y-ypred_bn"]=(df_m["Failure_today"]-df_m["ypred_bn"]).abs()
df_m.loc[:,"y-ypred_p":"y-ypred_bn"]

{"tags": ["hide_input"]}
```

[28]: {'tags': ['hide_input']}

[29]:
```
print("Distancia total al valor real del modelo Poisson: ",df_m["y-ypred_p"].
 ↪sum())
print("Distancia total al valor real del modelo Binomial Negativa:␣
 ↪",df_m["y-ypred_bn"].sum())

{"tags": ["hide_input"]}
```

```
Distancia total al valor real del modelo Poisson:  7982.864960537252
Distancia total al valor real del modelo Binomial Negativa:  8521.693187651039
```

[29]: {'tags': ['hide_input']}

En la tabla siguiente podemos ver los coeficientes asignados a cada variable por modelo. En ella podemos apreciar que las diferencias en las variables continuas es muy pequeña (La mayor diferencia es de 0.01), por lo que podríamos decir que son robustas a la especificación. Por otro lado las diferencias entre coeficientes de las ubicaciones y meses son mas grandes en general (Llegando hasta diferencias de 0.11). Diremos que estas no son robustas a la especificación.

[30]:
```
comparacion = pd.DataFrame({
    'Poisson': np.exp(poisson.params),
    'Binomial_Neg': np.exp(negbin.params)

})
comparacion["Poisson-Binomial"]=(comparacion["Poisson"]-comparacion["Binomial_Neg"]).
 ↪abs()
comparacion

{"tags": ["hide_input"]}
```

[30]: {'tags': ['hide_input']}

10 Transformación a PDF

[31]:
```
!jupyter nbconvert --to pdf --TagRemovePreprocessor.enabled=True␣
 ↪--TagRemovePreprocessor.remove_input_tags="['hide_input']" Tarea1_Meza_Núñez.
 ↪ipynb
```

```
C:\Users\franm\anaconda3\lib\site-packages\traitlets\traitlets.py:2915:
FutureWarning: --TagRemovePreprocessor.remove_input_tags=['hide_input'] for
containers is deprecated in traitlets 5.0. You can pass
```

```
`--TagRemovePreprocessor.remove_input_tags item` … multiple times to add items
to a list.
  warn(
[NbConvertApp] Converting notebook Tarea1_Meza_Núñez.ipynb to pdf
[NbConvertApp] ERROR | Error while converting 'Tarea1_Meza_Núñez.ipynb'
Traceback (most recent call last):
  File "C:\Users\franm\anaconda3\lib\site-packages\nbconvert\nbconvertapp.py",
line 488, in export_single_notebook
    output, resources = self.exporter.from_filename(
  File "C:\Users\franm\anaconda3\lib\site-
packages\nbconvert\exporters\exporter.py", line 189, in from_filename
    return self.from_file(f, resources=resources, **kw)
  File "C:\Users\franm\anaconda3\lib\site-
packages\nbconvert\exporters\exporter.py", line 206, in from_file
    return self.from_notebook_node(
  File "C:\Users\franm\anaconda3\lib\site-packages\nbconvert\exporters\pdf.py",
line 181, in from_notebook_node
    latex, resources = super().from_notebook_node(nb, resources=resources, **kw)
  File "C:\Users\franm\anaconda3\lib\site-
packages\nbconvert\exporters\latex.py", line 74, in from_notebook_node
    return super().from_notebook_node(nb, resources, **kw)
  File "C:\Users\franm\anaconda3\lib\site-
packages\nbconvert\exporters\templateexporter.py", line 413, in
from_notebook_node
    output = self.template.render(nb=nb_copy, resources=resources)
  File "C:\Users\franm\anaconda3\lib\site-packages\jinja2\environment.py", line
1301, in render
    self.environment.handle_exception()
  File "C:\Users\franm\anaconda3\lib\site-packages\jinja2\environment.py", line
936, in handle_exception
    raise rewrite_traceback_stack(source=source)
  File
"C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\index.tex.j2",
line 8, in top-level template code
    ((* extends cell_style *))
  File "C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\style_j
upyter.tex.j2", line 176, in top-level template code
    \prompt{(((prompt)))}{(((prompt_color)))}{(((execution_count)))}{(((extra_sp
ace)))}
  File
"C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\base.tex.j2",
line 7, in top-level template code
    ((*- extends 'document_contents.tex.j2' -*))
  File "C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\documen
t_contents.tex.j2", line 51, in top-level template code
    ((*- block figure scoped -*))
  File "C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\display
_priority.j2", line 5, in top-level template code
```

```
    ((*- extends 'null.j2' -*))
  File
"C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\null.j2", line
30, in top-level template code
    ((*- block body -*))
  File
"C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\base.tex.j2",
line 215, in block 'body'
    ((( super() )))
  File
"C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\null.j2", line
32, in block 'body'
    ((*- block any_cell scoped -*))
  File
"C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\null.j2", line
85, in block 'any_cell'
    ((*- block markdowncell scoped-*)) ((*- endblock markdowncell -*))
  File "C:\Users\franm\anaconda3\share\jupyter\nbconvert\templates\latex\documen
t_contents.tex.j2", line 68, in block 'markdowncell'
    ((( cell.source | citation2latex | strip_files_prefix |
convert_pandoc('markdown+tex_math_double_backslash', 'json',extra_args=[]) |
resolve_references | convert_pandoc('json','latex'))))
  File "C:\Users\franm\anaconda3\lib\site-packages\nbconvert\filters\pandoc.py",
line 24, in convert_pandoc
    return pandoc(source, from_format, to_format, extra_args=extra_args)
  File "C:\Users\franm\anaconda3\lib\site-packages\nbconvert\utils\pandoc.py",
line 51, in pandoc
    check_pandoc_version()
  File "C:\Users\franm\anaconda3\lib\site-packages\nbconvert\utils\pandoc.py",
line 99, in check_pandoc_version
    v = get_pandoc_version()
  File "C:\Users\franm\anaconda3\lib\site-packages\nbconvert\utils\pandoc.py",
line 76, in get_pandoc_version
    raise PandocMissing()
nbconvert.utils.pandoc.PandocMissing: Pandoc wasn't found.
Please check that pandoc is installed:
https://pandoc.org/installing.html
```

[ ]:

[ ]: