

Tarea1__Altamirano_Paredes

April 30, 2025

Tarea 1, Machine Learning and Data Analysis

Juan José Altamirano

2022403951

0.1 Importación de Librerías

```
[154]: 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 missingno as msn
import datetime as dt
```

0.2 Lectura de datos, exploración y limpieza de datos

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

```
[155]: df= pd.read_csv('machine_failure_data.csv', delimiter=",")
```

Las variables iniciales de estudio son las siguientes:

- Date: data medida en frecuencia diaria
- Location: ubicacion del medidor
- Min_Temp: temperatura minima observada
- Max_Temp: temperatura maxima observada
- Leakage: Filtracion medida en el area
- Evaporation: Tasa de evaporacion
- Electricity: Consumo electrico KW
- Parameter#: Diferentes sensores de reportando direccion y velocidad de viento en distintos momentos del dia, asi como otras metricas relevantes.
- Failure today: El sensor reporta fallo (o no)

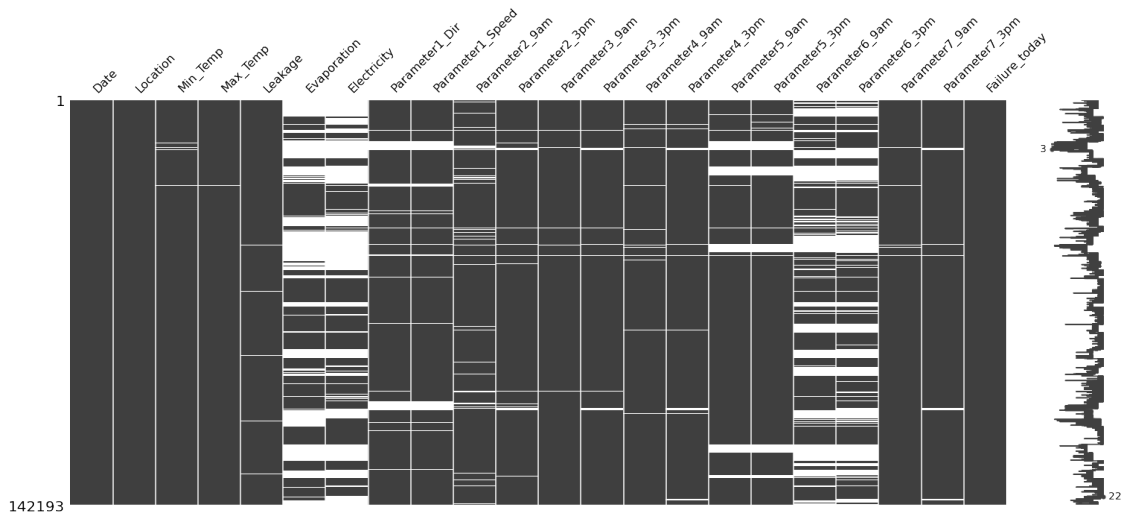
La variable “Failure_today” se transforma de manera binaria, indicando con un 1 si el sensor reporta un fallo, 0 en otro caso.

```
[156]: df['Failure_today']=df['Failure_today'].apply(lambda x: 0 if x=='No' else 1)
```

A continuación se muestran visualmente los datos vacíos en el dataframe en cada una de las variables:

```
[157]: msn.matrix(df)
```

```
[157]: <AxesSubplot: >
```



Notamos que las variables Evaporation, Electricity y el parámetro 6 (AM y PM) muestran un alto porcentaje de missing data, por lo que serán excluidos del análisis.

```
[158]: borrar= ['Parameter6_9am','Parameter6_3pm','Evaporation','Electricity',]
df.drop(borrar,axis=1, inplace=True)
df.reset_index(inplace=True, drop=True)
```

```
[159]: df2=df.copy()
```

Para mejor manejo de las direcciones, disminuirémos la cantidad, pasando de 16 posibles direcciones a 8, siguiendo la transformación del diccionario definido en la siguiente celda:

```
[160]: map_dir= {'SSE':'SE','WSW':'SW','SSW':'S','WNW':'W','ENE':'NE','ESE':'E','NNW':
    ↪ 'NW','NNE':'N'}
cols_dir = ['Parameter1_Dir','Parameter2_9am', 'Parameter2_3pm']
df2[cols_dir] = df2[cols_dir].replace(map_dir)
df2['Parameter1_Dir'].value_counts()
```

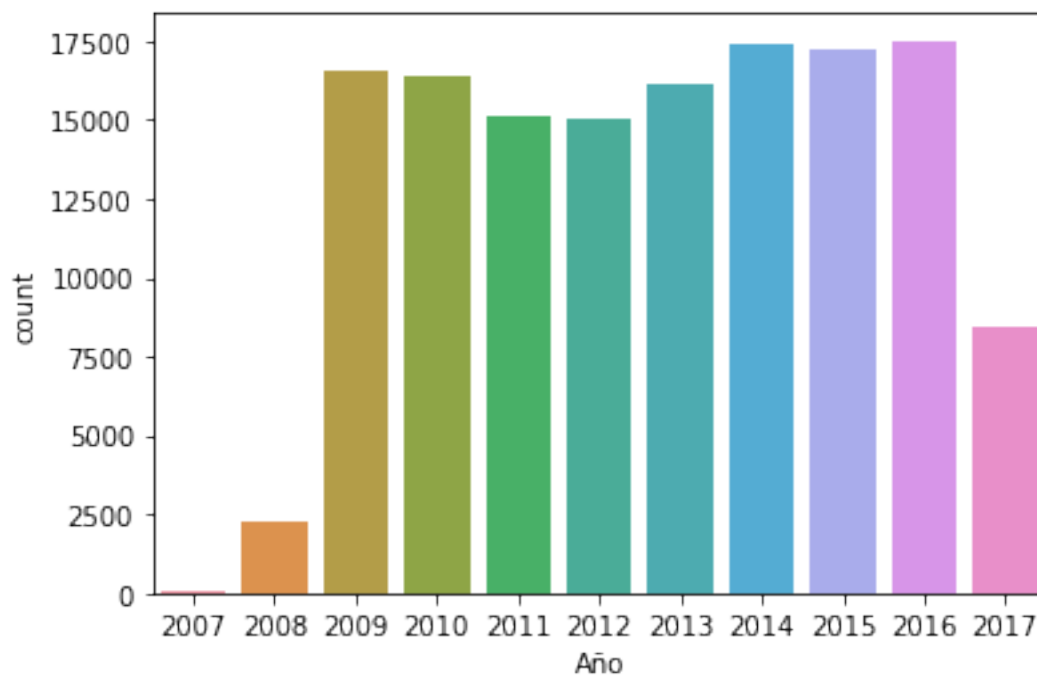
```
[160]: SE      18302  
      W       17846  
      SW      17698  
      S       17559  
      E       16376  
      N       15466  
      NE      15052  
      NW      14564  
      Name: Parameter1_Dir, dtype: int64
```

Transformamos la columna date para incorporar nuevas columnas de “Año” y “Mes” con el objetivo de facilitar el manejo temporal a posteriori.

```
[161]: df2['Date'] = pd.to_datetime(df2['Date'])  
      df2['Mes'] = df2['Date'].dt.month_name(locale='es')  
      df2['Año'] = df2['Date'].dt.year
```

```
[162]: sns.countplot(x='Año',data=df2)
```

```
[162]: <AxesSubplot: xlabel='Año', ylabel='count'>
```



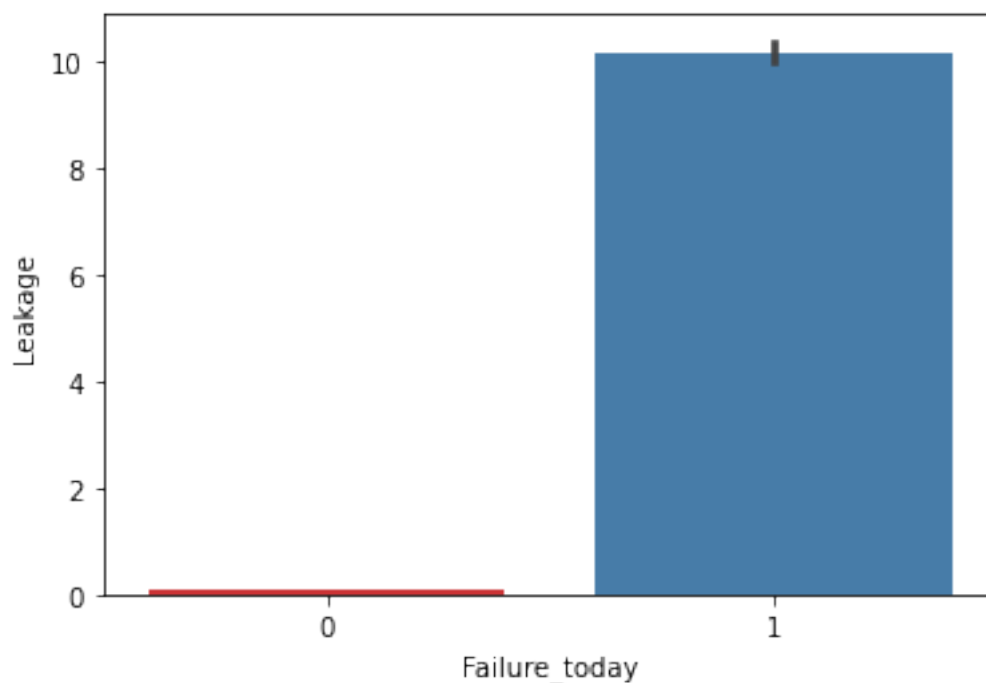
Notamos la poca cantidad de datos para los años 2007, 2008 y 2017, por lo que serán excluidos del análisis.

```
[163]: mask1 = df2[(df2['Año']==2007) |  
                (df2['Año']==2008) |  
                (df2['Año']==2017)]  
df2.drop(mask1.index,inplace=True)  
df2.reset_index(inplace=True, drop=True)
```

Por último, la variable Leakage está directamente relacionada a la variable “Failure_Today”, ya que si no hay filtraciones, no se reportan fallos y si hay filtraciones, se reportan fallos. Por lo tanto, se considera como una variable que sobreexplica el modelo, así que se excluirá del análisis.

```
[164]: sns.barplot(data=df2, y='Leakage', x='Failure_today', palette='Set1')
```

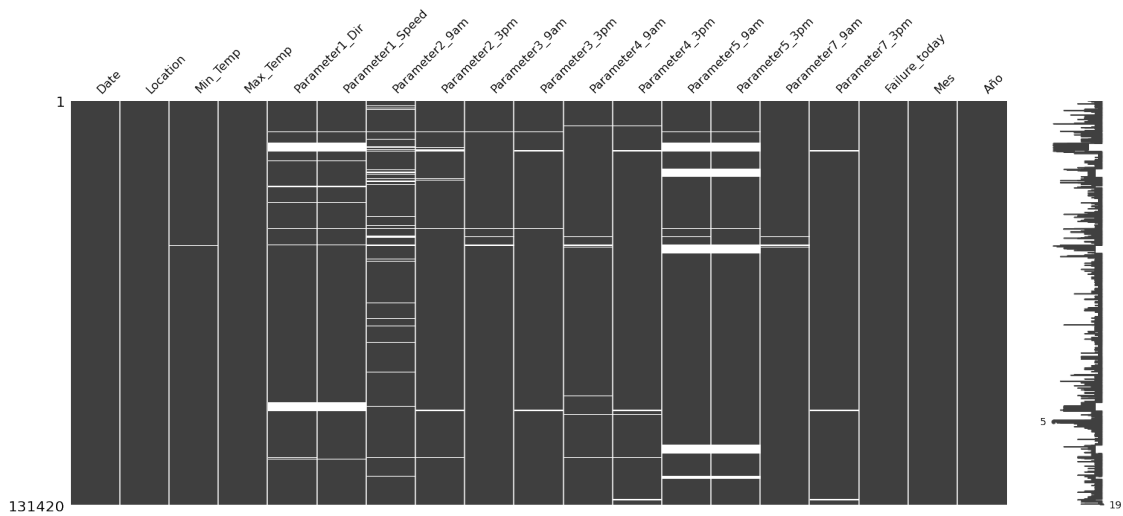
```
[164]: <AxesSubplot: xlabel='Failure_today', ylabel='Leakage'>
```



```
[165]: df2.drop("Leakage",axis=1, inplace=True)  
df2.reset_index(inplace=True, drop=True)
```

```
[166]: msn.matrix(df2)
```

```
[166]: <AxesSubplot: >
```



A continuación, se excluirán del análisis todas las observaciones que cuenten con algún dato vacío.

```
[167]: df2.dropna(inplace=True)
df2.reset_index(inplace=True, drop=True)
```

El siguiente gráfico radial muestra en qué dirección del viento del parámetro 1 se muestra la mayor cantidad de fallas:

```
[168]: # Direcciones ordenadas en sentido horario
directions = ['N',
              #'NNE',
              'NE',
              #'ENE',
              'E',
              #'ESE',
              'SE',
              #'SSE',
              'S',
              #'SSW',
              'SW',
              #'WSW',
              'W',
              #'WNW',
              'NW',
              #'NNW'
            ]

# Filtrar datos con fallo
failures = df2[df2['Failure_today'] == 1]
```

```

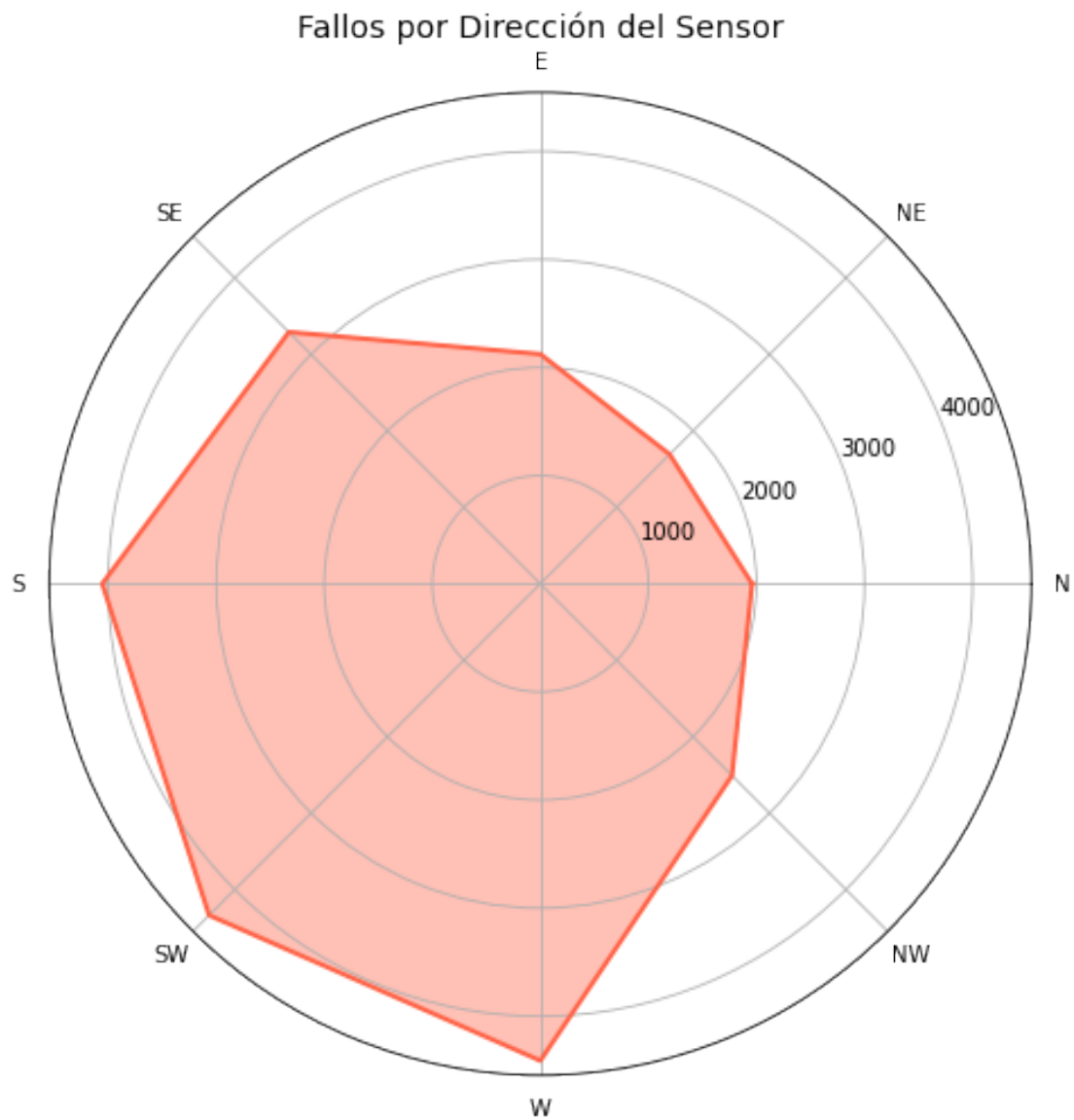
# Contar ocurrencias por dirección
counts = failures['Parameter1_Dir'].value_counts()
counts = counts.reindex(directions, fill_value=0)

# Convertir a radianes para el gráfico
angles = np.deg2rad(np.linspace(0, 360, len(directions), endpoint=False))

# Repetir el primer valor al final para cerrar el círculo
values = counts.values.tolist()
values += values[:1]
angles = np.append(angles, angles[0])

# Graficar
plt.figure(figsize=(8, 8))
ax = plt.subplot(111, polar=True)
ax.plot(angles, values, linewidth=2, linestyle='solid', color='tomato')
ax.fill(angles, values, alpha=0.4, color='tomato')
ax.set_xticks(np.deg2rad(np.linspace(0, 360, len(directions), endpoint=False)))
ax.set_xticklabels(directions)
ax.set_title('Fallos por Dirección del Sensor', size=14)
plt.show()

```



```
[169]: df2
```

```
[169]:
```

	Date	Location	Min_Temp	Max_Temp	Parameter1_Dir	\
0	2009-01-01	3	11.30	26.50	W	
1	2009-01-02	3	9.60	23.90	W	
2	2009-01-03	3	10.50	28.80	SE	
3	2009-01-04	3	12.30	34.60	W	
4	2009-01-05	3	12.90	35.80	W	
...	
105378	2016-12-27	42	22.10	35.80	W	
105379	2016-12-28	42	22.60	36.80	NW	

105380	2016-12-29	42	23.20	38.00	S
105381	2016-12-30	42	19.70	37.00	E
105382	2016-12-31	42	23.70	33.00	NE

	Parameter1_Speed	Parameter2_9am	Parameter2_3pm	Parameter3_9am	\
0	56.00	W	W	19.00	
1	41.00	SW	S	19.00	
2	26.00	SE	E	11.00	
3	37.00	SE	NW	6.00	
4	41.00	NE	NW	6.00	
...	
105378	43.00	W	W	17.00	
105379	50.00	NW	W	30.00	
105380	33.00	S	SW	17.00	
105381	37.00	E	W	22.00	
105382	46.00	NE	NE	24.00	

	Parameter3_3pm	Parameter4_9am	Parameter4_3pm	Parameter5_9am	\
0	31.00	46.00	26.00	1004.50	
1	11.00	44.00	22.00	1014.40	
2	7.00	43.00	22.00	1018.70	
3	17.00	41.00	12.00	1015.10	
4	26.00	41.00	9.00	1012.60	
...	
105378	31.00	77.00	41.00	997.80	
105379	15.00	63.00	39.00	1000.00	
105380	17.00	25.00	14.00	1004.40	
105381	6.00	30.00	23.00	1004.60	
105382	17.00	40.00	38.00	1005.10	

	Parameter5_3pm	Parameter7_9am	Parameter7_3pm	Failure_today	\
0	1003.20	19.70	25.70	0	
1	1013.10	14.90	22.10	0	
2	1014.80	17.10	26.50	0	
3	1010.30	20.70	33.90	0	
4	1009.20	22.40	34.40	0	
...	
105378	995.20	26.00	33.50	1	
105379	998.80	29.70	34.00	0	
105380	1001.00	28.70	36.40	0	
105381	1000.90	28.20	35.10	0	
105382	1002.70	30.10	31.50	0	

	Mes	Año
0	Enero	2009
1	Enero	2009
2	Enero	2009


```

3          Enero  2009
4          Enero  2009
...
105378  Diciembre  2016
105379  Diciembre  2016
105380  Diciembre  2016
105381  Diciembre  2016
105382  Diciembre  2016

```

```
[105383 rows x 19 columns]
```

El siguiente gráfico de correlación nos permite ver qué variables están altamente correlacionadas con el fin de evaluar su exclusión del modelo.

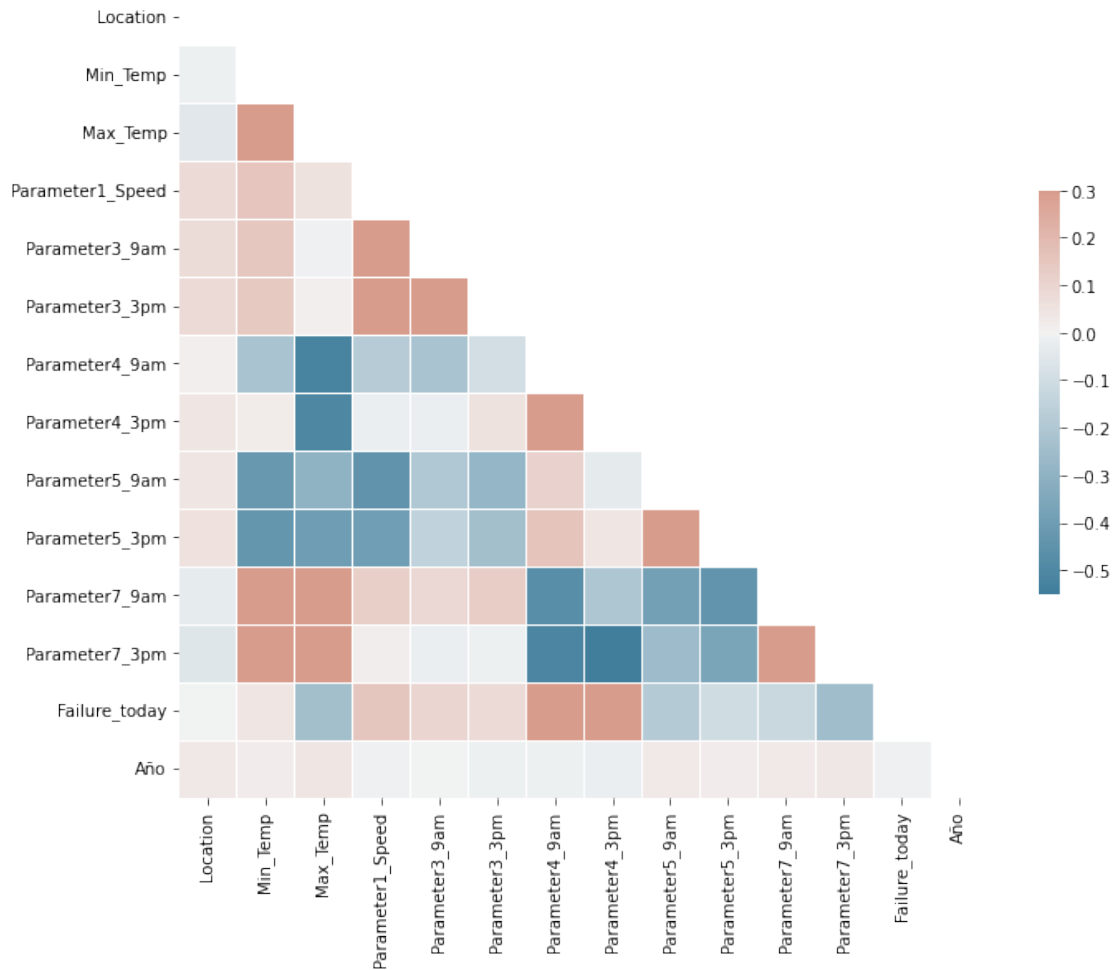
```

[171]: corr = df2.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=.3, center=0,
            square=True, linewidths=.5, cbar_kws={"shrink": .5})

```

```
[171]: <AxesSubplot: >
```



No existe ningún par de variables con una correlación significativa (cercana a 1), por lo que todas las variables presentes se incorporarán al análisis.

Con el fin de detectar supuestas variaciones estacionales, se crea la variable estación, la cual según el mes del año, se le asignará la estación a la que corresponde. Asumimos que los datos son originarios del hemisferio sur y agruparemos la estación Primavera y Otoño en la categoría “Otro”, para detectar cambios más significativos.

```
[172]: def get_estacion(mes):
    if mes in ['Diciembre', 'Enero', 'Febrero']:
        return 'Invierno'
    elif mes in ['Junio', 'Julio', 'Agosto']:
        return 'Verano'
    else:
        return 'Otro'

df2['Estacion'] = df2['Mes'].apply(get_estacion)
```

Establecemos el orden simbólico de las estaciones con el objetivo de que en nuestros modelos se tome como referencia la estación “Otro”

```
[173]: orden_estaciones = ['Otro', 'Verano', 'Invierno']
df2['Estacion'] = pd.Categorical(df2['Estacion'], categories=orden_estaciones,
↳ordered=True)
```

0.3 Modelo OLS

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

R:

En base a la formulación del modelo, se toma como referencia los siguientes valores de las variables categóricas: - Estación: Otro - Año: 2009 - Location: 2 - Parameter1_Dir: East - Parameter2_9am: East - Parameter2_3pm: East

Los principales efectos sobre la base establecida son la relación negativa de pasar a invierno o a verano. Esto significa que en comparación a la estación “Otro”, en Invierno y en Verano tienden a disminuir la probabilidad de que el sensor 2 detecte un fallo.

Al variar entre años, notamos una tendencia al aumento de probabilidad de que el sensor detecte un fallo, sin embargo, el cambio a algunos años como 2011 y 2016 no son significativos.

Si consideramos el cambio a otro sensor, notamos variabilidades en ambos sentidos, pero predominantemente negativas, lo que permite intuir que el sensor base (2) pudo haber detectado mayor cantidad de fallas en comparación a la gran cantidad de sensores que se relaciona de manera negativa en los coeficientes de la regresión.

Respecto a los parámetros direccionales, podemos concluir que una variación en el parámetro 1 no es significativa para el incremento o disminución del valor de nuestra variable explicativa. El parámetro 2 presenta coeficientes significativos y en su mayoría positivos, por lo que un cambio en las direcciones base (Este), tiende a aumentar la probabilidad de que el sensor detecte un fallo.

Por último, se destaca que a medida que aumenta la temperatura máxima detectada, también disminuye la probabilidad de tener fallas y análogamente, lo cual es un resultado que llama la atención por que en cierta parte desafía la lógica.

Dichos resultados se pueden deber a que un modelo de mínimos cuadrados no es lo suficientemente adecuado para estimar un valor binario, por lo que se continúa estudiando los datos con otros modelos.

```
[174]: resultado = smf.ols('Failure_today ~ C(Estacion) + C(Año) + C(Location) +
↳Min_Temp + Max_Temp + C(Parameter1_Dir) + Parameter1_Speed +
↳C(Parameter2_9am) + C(Parameter2_3pm) + Parameter3_9am + Parameter3_3pm +
↳Parameter4_9am + Parameter4_3pm + Parameter5_9am + Parameter5_3pm +
↳Parameter7_9am + Parameter7_3pm', data=df2).fit()
print(resultado.summary())
```

OLS Regression Results

=====

```

Dep. Variable:      Failure_today      R-squared:      0.298
Model:              OLS                Adj. R-squared: 0.297
Method:             Least Squares      F-statistic:    531.0
Date:               jue, 24 abr. 2025   Prob (F-statistic): 0.00
Time:               23:52:45           Log-Likelihood: -39926.
No. Observations:   105383            AIC:            8.002e+04
Df Residuals:       105298            BIC:            8.084e+04
Df Model:           84
Covariance Type:    nonrobust

```

	coef	std err	t	P> t	[0.025
0.975]					

Intercept	8.0374	0.228	35.239	0.000	7.590
8.484					
C(Estacion) [T.Verano]	-0.0201	0.003	-6.096	0.000	-0.027
-0.014					
C(Estacion) [T.Invierno]	-0.0240	0.003	-7.471	0.000	-0.030
-0.018					
C(Año) [T.2010]	0.0087	0.004	1.973	0.048	5.79e-05
0.017					
C(Año) [T.2011]	0.0033	0.004	0.737	0.461	-0.005
0.012					
C(Año) [T.2012]	0.0116	0.004	2.602	0.009	0.003
0.020					
C(Año) [T.2013]	0.0122	0.004	2.778	0.005	0.004
0.021					
C(Año) [T.2014]	0.0100	0.004	2.294	0.022	0.001
0.018					
C(Año) [T.2015]	0.0179	0.004	4.071	0.000	0.009
0.026					
C(Año) [T.2016]	0.0063	0.004	1.460	0.144	-0.002
0.015					
C(Location) [T.3]	-0.0600	0.011	-5.703	0.000	-0.081
-0.039					
C(Location) [T.4]	0.1028	0.010	9.844	0.000	0.082
0.123					
C(Location) [T.5]	-0.0991	0.011	-9.254	0.000	-0.120
-0.078					
C(Location) [T.6]	-0.2198	0.010	-21.007	0.000	-0.240
-0.199					
C(Location) [T.7]	-0.1113	0.010	-10.831	0.000	-0.131
-0.091					
C(Location) [T.8]	0.0095	0.010	0.908	0.364	-0.011
0.030					
C(Location) [T.9]	-0.0286	0.011	-2.590	0.010	-0.050

-0.007					
C(Location) [T.10]	-0.1016	0.011	-9.383	0.000	-0.123
-0.080					
C(Location) [T.11]	-0.0304	0.010	-3.042	0.002	-0.050
-0.011					
C(Location) [T.12]	-0.0418	0.011	-3.874	0.000	-0.063
-0.021					
C(Location) [T.13]	-0.1578	0.011	-14.602	0.000	-0.179
-0.137					
C(Location) [T.14]	-0.0899	0.011	-8.275	0.000	-0.111
-0.069					
C(Location) [T.15]	-0.0475	0.011	-4.462	0.000	-0.068
-0.027					
C(Location) [T.16]	-0.1619	0.010	-15.759	0.000	-0.182
-0.142					
C(Location) [T.17]	-0.0732	0.016	-4.510	0.000	-0.105
-0.041					
C(Location) [T.18]	-0.1615	0.012	-13.415	0.000	-0.185
-0.138					
C(Location) [T.19]	-0.0608	0.011	-5.588	0.000	-0.082
-0.039					
C(Location) [T.20]	-0.1663	0.010	-16.339	0.000	-0.186
-0.146					
C(Location) [T.21]	-0.0962	0.010	-9.592	0.000	-0.116
-0.077					
C(Location) [T.22]	-0.0309	0.010	-2.999	0.003	-0.051
-0.011					
C(Location) [T.23]	-0.1007	0.010	-9.872	0.000	-0.121
-0.081					
C(Location) [T.26]	-0.1776	0.012	-14.413	0.000	-0.202
-0.153					
C(Location) [T.27]	-0.1392	0.010	-13.537	0.000	-0.159
-0.119					
C(Location) [T.28]	-0.1520	0.010	-14.805	0.000	-0.172
-0.132					
C(Location) [T.29]	-0.0919	0.010	-9.130	0.000	-0.112
-0.072					
C(Location) [T.30]	-0.0095	0.010	-0.917	0.359	-0.030
0.011					
C(Location) [T.32]	-0.0376	0.010	-3.768	0.000	-0.057
-0.018					
C(Location) [T.33]	-0.0400	0.010	-3.993	0.000	-0.060
-0.020					
C(Location) [T.34]	-0.1282	0.010	-12.586	0.000	-0.148
-0.108					
C(Location) [T.35]	-0.0822	0.011	-7.469	0.000	-0.104
-0.061					
C(Location) [T.36]	-0.1989	0.010	-19.060	0.000	-0.219

-0.178					
C(Location) [T.38]	-0.0880	0.011	-8.061	0.000	-0.109
-0.067					
C(Location) [T.39]	-0.0721	0.010	-7.044	0.000	-0.092
-0.052					
C(Location) [T.40]	-0.0883	0.011	-8.008	0.000	-0.110
-0.067					
C(Location) [T.41]	-0.0711	0.011	-6.645	0.000	-0.092
-0.050					
C(Location) [T.42]	0.0749	0.013	5.940	0.000	0.050
0.100					
C(Location) [T.43]	-0.0682	0.010	-6.760	0.000	-0.088
-0.048					
C(Location) [T.44]	-0.1048	0.010	-10.070	0.000	-0.125
-0.084					
C(Location) [T.45]	-0.1518	0.010	-14.972	0.000	-0.172
-0.132					
C(Location) [T.46]	-0.0152	0.011	-1.396	0.163	-0.036
0.006					
C(Location) [T.47]	-0.0499	0.011	-4.678	0.000	-0.071
-0.029					
C(Location) [T.48]	-0.1728	0.010	-16.811	0.000	-0.193
-0.153					
C(Location) [T.49]	-0.0820	0.010	-8.121	0.000	-0.102
-0.062					
C(Parameter1_Dir) [T.N]	-0.0123	0.006	-2.200	0.028	-0.023
-0.001					
C(Parameter1_Dir) [T.NE]	-0.0099	0.005	-1.971	0.049	-0.020
-5.61e-05					
C(Parameter1_Dir) [T.NW]	0.0016	0.006	0.270	0.787	-0.010
0.013					
C(Parameter1_Dir) [T.S]	-0.0004	0.005	-0.069	0.945	-0.011
0.010					
C(Parameter1_Dir) [T.SE]	-0.0066	0.005	-1.362	0.173	-0.016
0.003					
C(Parameter1_Dir) [T.SW]	0.0061	0.005	1.137	0.256	-0.004
0.017					
C(Parameter1_Dir) [T.W]	-0.0023	0.006	-0.409	0.683	-0.013
0.009					
C(Parameter2_9am) [T.N]	-0.0049	0.005	-1.001	0.317	-0.014
0.005					
C(Parameter2_9am) [T.NE]	0.0188	0.005	4.052	0.000	0.010
0.028					
C(Parameter2_9am) [T.NW]	0.0039	0.005	0.740	0.459	-0.006
0.014					
C(Parameter2_9am) [T.S]	0.0254	0.005	5.145	0.000	0.016
0.035					
C(Parameter2_9am) [T.SE]	0.0044	0.005	0.974	0.330	-0.004

0.013					
C(Parameter2_9am) [T.SW]	0.0662	0.005	12.678	0.000	0.056
0.076					
C(Parameter2_9am) [T.W]	0.0278	0.005	5.250	0.000	0.017
0.038					
C(Parameter2_3pm) [T.N]	-0.0054	0.006	-0.977	0.328	-0.016
0.005					
C(Parameter2_3pm) [T.NE]	-0.0202	0.005	-4.053	0.000	-0.030
-0.010					
C(Parameter2_3pm) [T.NW]	0.0302	0.006	5.279	0.000	0.019
0.041					
C(Parameter2_3pm) [T.S]	0.0159	0.005	3.085	0.002	0.006
0.026					
C(Parameter2_3pm) [T.SE]	0.0090	0.005	1.946	0.052	-6.35e-05
0.018					
C(Parameter2_3pm) [T.SW]	0.0148	0.005	2.733	0.006	0.004
0.025					
C(Parameter2_3pm) [T.W]	0.0288	0.006	5.195	0.000	0.018
0.040					
Min_Temp	0.0087	0.001	15.471	0.000	0.008
0.010					
Max_Temp	-0.0326	0.001	-30.327	0.000	-0.035
-0.031					
Parameter1_Speed	0.0053	0.000	38.484	0.000	0.005
0.006					
Parameter3_9am	0.0029	0.000	15.646	0.000	0.003
0.003					
Parameter3_3pm	-0.0039	0.000	-20.686	0.000	-0.004
-0.004					
Parameter4_9am	0.0072	0.000	59.147	0.000	0.007
0.007					
Parameter4_3pm	0.0023	0.000	16.685	0.000	0.002
0.003					
Parameter5_9am	-0.0362	0.001	-46.111	0.000	-0.038
-0.035					
Parameter5_3pm	0.0280	0.001	35.918	0.000	0.026
0.030					
Parameter7_9am	-0.0002	0.001	-0.257	0.797	-0.002
0.002					
Parameter7_3pm	0.0246	0.001	20.850	0.000	0.022
0.027					
=====					
Omnibus:	7736.456	Durbin-Watson:		1.786	
Prob(Omnibus):	0.000	Jarque-Bera (JB):		9465.860	
Skew:	0.728	Prob(JB):		0.00	
Kurtosis:	2.811	Cond. No.		3.02e+05	
=====					

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 3.02e+05. This might indicate that there are strong multicollinearity or other numerical problems.

0.4 Modelo Probit

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

R: Si bien los resultados en coeficientes son similares entre el modelo OLS y los efectos marginales del modelo Probit, podemos notar que existen diferencias en la significancia de ciertos coeficientes como el cambio a la estación de Invierno, que el modelo Probit ya no lo considera significativo.

Además, Probit arroja que el cambio a cualquier año distinto al base también es significativo, pero mantiene la relación positiva en torno a la estimación de nuestra variable Failure Today.

Por otro lado, disminuye la magnitud de gran parte de los coeficientes, suavizando en cierta forma el impacto de algunas variables sobre la variable failure_today.

```
[175]: probit = smf.probit('Failure_today ~ C(Estacion) + C(Año) + C(Location) +  
  ↳Min_Temp + Max_Temp + C(Parameter1_Dir) + Parameter1_Speed +  
  ↳C(Parameter2_9am) + C(Parameter2_3pm) + Parameter3_9am + Parameter3_3pm +  
  ↳Parameter4_9am + Parameter4_3pm + Parameter5_9am + Parameter5_3pm +  
  ↳Parameter7_9am + Parameter7_3pm' , data=df2).fit()  
print(probit.summary())  
  
mfx = probit.get_margeff()  
print(mfx.summary())
```

Optimization terminated successfully.

Current function value: 0.363349

Iterations 7

Probit Regression Results

```
=====
```

Dep. Variable:	Failure_today	No. Observations:	105383
Model:	Probit	Df Residuals:	105298
Method:	MLE	Df Model:	84
Date:	jue, 24 abr. 2025	Pseudo R-squ.:	0.3282
Time:	23:52:53	Log-Likelihood:	-38291.
converged:	True	LL-Null:	-57001.
Covariance Type:	nonrobust	LLR p-value:	0.000

```
=====
```

	coef	std err	z	P> z	[0.025
0.975]					

Intercept	27.6812	1.036	26.722	0.000	25.651

29.712					
C(Estacion) [T.Verano]	-0.1924	0.015	-12.502	0.000	-0.223
-0.162					
C(Estacion) [T.Invierno]	-0.0189	0.016	-1.214	0.225	-0.049
0.012					
C(Año) [T.2010]	0.0777	0.021	3.772	0.000	0.037
0.118					
C(Año) [T.2011]	0.0591	0.021	2.819	0.005	0.018
0.100					
C(Año) [T.2012]	0.0676	0.021	3.175	0.001	0.026
0.109					
C(Año) [T.2013]	0.0686	0.021	3.267	0.001	0.027
0.110					
C(Año) [T.2014]	0.0657	0.021	3.131	0.002	0.025
0.107					
C(Año) [T.2015]	0.0812	0.021	3.839	0.000	0.040
0.123					
C(Año) [T.2016]	0.0728	0.021	3.544	0.000	0.033
0.113					
C(Location) [T.3]	-0.2491	0.051	-4.868	0.000	-0.349
-0.149					
C(Location) [T.4]	0.2631	0.063	4.176	0.000	0.140
0.387					
C(Location) [T.5]	-0.2657	0.051	-5.240	0.000	-0.365
-0.166					
C(Location) [T.6]	-1.0538	0.050	-21.078	0.000	-1.152
-0.956					
C(Location) [T.7]	-0.5144	0.051	-10.072	0.000	-0.615
-0.414					
C(Location) [T.8]	0.3469	0.049	7.071	0.000	0.251
0.443					
C(Location) [T.9]	0.1580	0.051	3.089	0.002	0.058
0.258					
C(Location) [T.10]	-0.3404	0.052	-6.526	0.000	-0.443
-0.238					
C(Location) [T.11]	-0.2128	0.054	-3.976	0.000	-0.318
-0.108					
C(Location) [T.12]	0.0276	0.050	0.553	0.581	-0.070
0.125					
C(Location) [T.13]	-0.7287	0.050	-14.570	0.000	-0.827
-0.631					
C(Location) [T.14]	-0.0459	0.052	-0.882	0.378	-0.148
0.056					
C(Location) [T.15]	0.0421	0.050	0.839	0.402	-0.056
0.140					
C(Location) [T.16]	-0.5745	0.048	-11.999	0.000	-0.668
-0.481					
C(Location) [T.17]	0.0074	0.084	0.089	0.929	-0.157

0.172					
C(Location) [T.18]	-0.6033	0.056	-10.853	0.000	-0.712
-0.494					
C(Location) [T.19]	-0.1313	0.049	-2.685	0.007	-0.227
-0.035					
C(Location) [T.20]	-0.6412	0.048	-13.269	0.000	-0.736
-0.547					
C(Location) [T.21]	-0.6493	0.054	-11.969	0.000	-0.756
-0.543					
C(Location) [T.22]	0.0732	0.052	1.398	0.162	-0.029
0.176					
C(Location) [T.23]	-0.4477	0.048	-9.385	0.000	-0.541
-0.354					
C(Location) [T.26]	-0.9467	0.063	-14.951	0.000	-1.071
-0.823					
C(Location) [T.27]	-0.4661	0.047	-9.815	0.000	-0.559
-0.373					
C(Location) [T.28]	-0.4816	0.047	-10.323	0.000	-0.573
-0.390					
C(Location) [T.29]	-0.5762	0.051	-11.242	0.000	-0.677
-0.476					
C(Location) [T.30]	0.1439	0.051	2.818	0.005	0.044
0.244					
C(Location) [T.32]	-0.0562	0.050	-1.129	0.259	-0.154
0.041					
C(Location) [T.33]	-0.0211	0.050	-0.418	0.676	-0.120
0.078					
C(Location) [T.34]	-0.5700	0.046	-12.302	0.000	-0.661
-0.479					
C(Location) [T.35]	-0.2110	0.052	-4.019	0.000	-0.314
-0.108					
C(Location) [T.36]	-0.7289	0.049	-14.919	0.000	-0.825
-0.633					
C(Location) [T.38]	-0.2159	0.050	-4.309	0.000	-0.314
-0.118					
C(Location) [T.39]	-0.1812	0.048	-3.770	0.000	-0.275
-0.087					
C(Location) [T.40]	-0.0887	0.054	-1.639	0.101	-0.195
0.017					
C(Location) [T.41]	-0.2015	0.051	-3.946	0.000	-0.302
-0.101					
C(Location) [T.42]	0.2155	0.079	2.724	0.006	0.060
0.370					
C(Location) [T.43]	-0.2857	0.051	-5.635	0.000	-0.385
-0.186					
C(Location) [T.44]	-0.3779	0.048	-7.930	0.000	-0.471
-0.284					
C(Location) [T.45]	-0.6227	0.048	-12.995	0.000	-0.717

-0.529					
C(Location) [T.46]	0.1264	0.049	2.561	0.010	0.030
0.223					
C(Location) [T.47]	-0.1122	0.050	-2.262	0.024	-0.209
-0.015					
C(Location) [T.48]	-0.5918	0.048	-12.305	0.000	-0.686
-0.498					
C(Location) [T.49]	-0.7142	0.060	-11.976	0.000	-0.831
-0.597					
C(Parameter1_Dir) [T.N]	-0.0864	0.029	-3.025	0.002	-0.142
-0.030					
C(Parameter1_Dir) [T.NE]	-0.0499	0.026	-1.935	0.053	-0.100
0.001					
C(Parameter1_Dir) [T.NW]	-0.0149	0.029	-0.511	0.609	-0.072
0.042					
C(Parameter1_Dir) [T.S]	-0.0109	0.025	-0.431	0.667	-0.061
0.039					
C(Parameter1_Dir) [T.SE]	-0.0361	0.024	-1.519	0.129	-0.083
0.010					
C(Parameter1_Dir) [T.SW]	0.0427	0.026	1.617	0.106	-0.009
0.095					
C(Parameter1_Dir) [T.W]	0.0007	0.027	0.028	0.978	-0.052
0.054					
C(Parameter2_9am) [T.N]	-0.0201	0.026	-0.783	0.433	-0.070
0.030					
C(Parameter2_9am) [T.NE]	0.0627	0.026	2.457	0.014	0.013
0.113					
C(Parameter2_9am) [T.NW]	0.0560	0.026	2.151	0.031	0.005
0.107					
C(Parameter2_9am) [T.S]	0.1491	0.024	6.111	0.000	0.101
0.197					
C(Parameter2_9am) [T.SE]	0.0970	0.024	4.102	0.000	0.051
0.143					
C(Parameter2_9am) [T.SW]	0.2847	0.025	11.313	0.000	0.235
0.334					
C(Parameter2_9am) [T.W]	0.1422	0.026	5.513	0.000	0.092
0.193					
C(Parameter2_3pm) [T.N]	-0.0446	0.028	-1.599	0.110	-0.099
0.010					
C(Parameter2_3pm) [T.NE]	-0.0840	0.025	-3.367	0.001	-0.133
-0.035					
C(Parameter2_3pm) [T.NW]	0.0875	0.028	3.090	0.002	0.032
0.143					
C(Parameter2_3pm) [T.S]	0.0060	0.025	0.244	0.807	-0.042
0.054					
C(Parameter2_3pm) [T.SE]	0.0038	0.022	0.171	0.864	-0.040
0.048					
C(Parameter2_3pm) [T.SW]	-0.0066	0.026	-0.252	0.801	-0.058

0.045					
C(Parameter2_3pm) [T.W]	0.0606	0.027	2.253	0.024	0.008
0.113					
Min_Temp	0.0697	0.003	23.340	0.000	0.064
0.076					
Max_Temp	-0.1372	0.005	-25.908	0.000	-0.148
-0.127					
Parameter1_Speed	0.0198	0.001	31.175	0.000	0.019
0.021					
Parameter3_9am	0.0098	0.001	11.025	0.000	0.008
0.012					
Parameter3_3pm	-0.0132	0.001	-14.830	0.000	-0.015
-0.011					
Parameter4_9am	0.0386	0.001	62.573	0.000	0.037
0.040					
Parameter4_3pm	0.0025	0.001	3.975	0.000	0.001
0.004					
Parameter5_9am	-0.1274	0.004	-34.927	0.000	-0.135
-0.120					
Parameter5_3pm	0.0974	0.004	27.039	0.000	0.090
0.104					
Parameter7_9am	-0.0087	0.005	-1.852	0.064	-0.018
0.001					
Parameter7_3pm	0.0556	0.006	9.619	0.000	0.044
0.067					

=====

=====

Probit Marginal Effects

=====

Dep. Variable: Failure_today

Method: dydx

At: overall

=====

=====

	dy/dx	std err	z	P> z	[0.025
0.975]					

C(Estacion) [T.Verano]	-0.0393	0.003	-12.528	0.000	-0.045
-0.033					
C(Estacion) [T.Invierno]	-0.0039	0.003	-1.214	0.225	-0.010
0.002					
C(Año) [T.2010]	0.0159	0.004	3.772	0.000	0.008
0.024					
C(Año) [T.2011]	0.0121	0.004	2.819	0.005	0.004
0.020					
C(Año) [T.2012]	0.0138	0.004	3.175	0.001	0.005
0.022					

C(Año) [T. 2013] 0.022	0.0140	0.004	3.268	0.001	0.006
C(Año) [T. 2014] 0.022	0.0134	0.004	3.132	0.002	0.005
C(Año) [T. 2015] 0.025	0.0166	0.004	3.840	0.000	0.008
C(Año) [T. 2016] 0.023	0.0149	0.004	3.545	0.000	0.007
C(Location) [T.3] -0.030	-0.0509	0.010	-4.869	0.000	-0.071
C(Location) [T.4] 0.079	0.0537	0.013	4.177	0.000	0.029
C(Location) [T.5] -0.034	-0.0543	0.010	-5.243	0.000	-0.075
C(Location) [T.6] -0.195	-0.2152	0.010	-21.201	0.000	-0.235
C(Location) [T.7] -0.085	-0.1051	0.010	-10.086	0.000	-0.125
C(Location) [T.8] 0.090	0.0709	0.010	7.073	0.000	0.051
C(Location) [T.9] 0.053	0.0323	0.010	3.089	0.002	0.012
C(Location) [T.10] -0.049	-0.0695	0.011	-6.530	0.000	-0.090
C(Location) [T.11] -0.022	-0.0435	0.011	-3.976	0.000	-0.065
C(Location) [T.12] 0.026	0.0056	0.010	0.553	0.581	-0.014
C(Location) [T.13] -0.129	-0.1488	0.010	-14.614	0.000	-0.169
C(Location) [T.14] 0.011	-0.0094	0.011	-0.882	0.378	-0.030
C(Location) [T.15] 0.029	0.0086	0.010	0.839	0.402	-0.011
C(Location) [T.16] -0.098	-0.1173	0.010	-12.024	0.000	-0.136
C(Location) [T.17] 0.035	0.0015	0.017	0.089	0.929	-0.032
C(Location) [T.18] -0.101	-0.1232	0.011	-10.872	0.000	-0.145
C(Location) [T.19] -0.007	-0.0268	0.010	-2.685	0.007	-0.046
C(Location) [T.20] -0.112	-0.1310	0.010	-13.304	0.000	-0.150
C(Location) [T.21] -0.111	-0.1326	0.011	-11.993	0.000	-0.154
C(Location) [T.22] 0.036	0.0150	0.011	1.398	0.162	-0.006

C(Location) [T.23] -0.072	-0.0914	0.010	-9.397	0.000	-0.111
C(Location) [T.26] -0.168	-0.1934	0.013	-14.998	0.000	-0.219
C(Location) [T.27] -0.076	-0.0952	0.010	-9.830	0.000	-0.114
C(Location) [T.28] -0.080	-0.0984	0.010	-10.342	0.000	-0.117
C(Location) [T.29] -0.097	-0.1177	0.010	-11.260	0.000	-0.138
C(Location) [T.30] 0.050	0.0294	0.010	2.818	0.005	0.009
C(Location) [T.32] 0.008	-0.0115	0.010	-1.129	0.259	-0.031
C(Location) [T.33] 0.016	-0.0043	0.010	-0.418	0.676	-0.024
C(Location) [T.34] -0.098	-0.1164	0.009	-12.330	0.000	-0.135
C(Location) [T.35] -0.022	-0.0431	0.011	-4.021	0.000	-0.064
C(Location) [T.36] -0.129	-0.1489	0.010	-14.968	0.000	-0.168
C(Location) [T.38] -0.024	-0.0441	0.010	-4.311	0.000	-0.064
C(Location) [T.39] -0.018	-0.0370	0.010	-3.771	0.000	-0.056
C(Location) [T.40] 0.004	-0.0181	0.011	-1.639	0.101	-0.040
C(Location) [T.41] -0.021	-0.0412	0.010	-3.947	0.000	-0.062
C(Location) [T.42] 0.076	0.0440	0.016	2.725	0.006	0.012
C(Location) [T.43] -0.038	-0.0584	0.010	-5.637	0.000	-0.079
C(Location) [T.44] -0.058	-0.0772	0.010	-7.938	0.000	-0.096
C(Location) [T.45] -0.108	-0.1272	0.010	-13.026	0.000	-0.146
C(Location) [T.46] 0.046	0.0258	0.010	2.562	0.010	0.006
C(Location) [T.47] -0.003	-0.0229	0.010	-2.262	0.024	-0.043
C(Location) [T.48] -0.102	-0.1209	0.010	-12.335	0.000	-0.140
C(Location) [T.49] -0.122	-0.1459	0.012	-11.997	0.000	-0.170
C(Parameter1_Dir) [T.N] -0.006	-0.0176	0.006	-3.025	0.002	-0.029

C(Parameter1_Dir) [T.NE] 0.000	-0.0102	0.005	-1.935	0.053	-0.021
C(Parameter1_Dir) [T.NW] 0.009	-0.0030	0.006	-0.511	0.609	-0.015
C(Parameter1_Dir) [T.S] 0.008	-0.0022	0.005	-0.431	0.667	-0.012
C(Parameter1_Dir) [T.SE] 0.002	-0.0074	0.005	-1.519	0.129	-0.017
C(Parameter1_Dir) [T.SW] 0.019	0.0087	0.005	1.617	0.106	-0.002
C(Parameter1_Dir) [T.W] 0.011	0.0002	0.006	0.028	0.978	-0.011
C(Parameter2_9am) [T.N] 0.006	-0.0041	0.005	-0.783	0.433	-0.014
C(Parameter2_9am) [T.NE] 0.023	0.0128	0.005	2.458	0.014	0.003
C(Parameter2_9am) [T.NW] 0.022	0.0114	0.005	2.151	0.031	0.001
C(Parameter2_9am) [T.S] 0.040	0.0305	0.005	6.114	0.000	0.021
C(Parameter2_9am) [T.SE] 0.029	0.0198	0.005	4.103	0.000	0.010
C(Parameter2_9am) [T.SW] 0.068	0.0582	0.005	11.332	0.000	0.048
C(Parameter2_9am) [T.W] 0.039	0.0291	0.005	5.515	0.000	0.019
C(Parameter2_3pm) [T.N] 0.002	-0.0091	0.006	-1.599	0.110	-0.020
C(Parameter2_3pm) [T.NE] -0.007	-0.0172	0.005	-3.368	0.001	-0.027
C(Parameter2_3pm) [T.NW] 0.029	0.0179	0.006	3.090	0.002	0.007
C(Parameter2_3pm) [T.S] 0.011	0.0012	0.005	0.244	0.807	-0.009
C(Parameter2_3pm) [T.SE] 0.010	0.0008	0.005	0.171	0.864	-0.008
C(Parameter2_3pm) [T.SW] 0.009	-0.0014	0.005	-0.252	0.801	-0.012
C(Parameter2_3pm) [T.W] 0.023	0.0124	0.005	2.253	0.024	0.002
Min_Temp 0.015	0.0142	0.001	23.489	0.000	0.013
Max_Temp -0.026	-0.0280	0.001	-26.127	0.000	-0.030
Parameter1_Speed 0.004	0.0040	0.000	31.623	0.000	0.004
Parameter3_9am 0.002	0.0020	0.000	11.047	0.000	0.002

Parameter3_3pm -0.002	-0.0027	0.000	-14.881	0.000	-0.003
Parameter4_9am 0.008	0.0079	0.000	65.813	0.000	0.008
Parameter4_3pm 0.001	0.0005	0.000	3.977	0.000	0.000
Parameter5_9am -0.025	-0.0260	0.001	-35.536	0.000	-0.027
Parameter5_3pm 0.021	0.0199	0.001	27.313	0.000	0.018
Parameter7_9am 0.000	-0.0018	0.001	-1.852	0.064	-0.004
Parameter7_3pm 0.014	0.0114	0.001	9.632	0.000	0.009

=====

=====

0.5 Logit

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.

Se mantiene el uso de las mismas variables explicativas.

Notamos que el cambio desde la estación base a la estación Invierno continua siendo no significativo.

Además, se mantiene el resultado de la no significancia de los coeficientes asociados a las direcciones del parámetro 1, lo que puede presumir una baja explicación de la variabilidad de Failure today en torno a esta variable.

El modelo Logit vuelve a reafirmar que si la temperatura máxima detectada por el sensor aumenta, tienden a detectarse menos fallas, y si la temperatura mínima aumenta en una unidad, la probabilidad de fallo aumenta en un 1.4%. Curioso...

```
[176]: logit = smf.logit('Failure_today ~ C(Estacion) + C(Año) + C(Location) +
↳Min_Temp + Max_Temp + C(Parameter1_Dir) + Parameter1_Speed +
↳C(Parameter2_9am) + C(Parameter2_3pm) + Parameter3_9am + Parameter3_3pm +
↳Parameter4_9am + Parameter4_3pm + Parameter5_9am + Parameter5_3pm +
↳Parameter7_9am + Parameter7_3pm' , data=df2).fit()
print(logit.summary())
mfx = logit.get_margeff()
print(mfx.summary())
```

Optimization terminated successfully.

Current function value: 0.361726

Iterations 8

Logit Regression Results

```
=====
Dep. Variable:          Failure_today    No. Observations:          105383
Model:                  Logit           Df Residuals:          105298
```


Method:	MLE	Df Model:	84
Date:	jue, 24 abr. 2025	Pseudo R-squ.:	0.3312
Time:	23:54:13	Log-Likelihood:	-38120.
converged:	True	LL-Null:	-57001.
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P> z	[0.025
0.975]					

Intercept	47.8709	1.828	26.192	0.000	44.289
51.453					
C(Estacion) [T.Verano]	-0.3514	0.027	-12.949	0.000	-0.405
-0.298					
C(Estacion) [T.Invierno]	-0.0277	0.028	-0.994	0.320	-0.082
0.027					
C(Año) [T.2010]	0.1537	0.036	4.231	0.000	0.082
0.225					
C(Año) [T.2011]	0.1184	0.037	3.192	0.001	0.046
0.191					
C(Año) [T.2012]	0.1343	0.038	3.561	0.000	0.060
0.208					
C(Año) [T.2013]	0.1207	0.037	3.236	0.001	0.048
0.194					
C(Año) [T.2014]	0.1218	0.037	3.264	0.001	0.049
0.195					
C(Año) [T.2015]	0.1458	0.038	3.877	0.000	0.072
0.220					
C(Año) [T.2016]	0.1337	0.036	3.672	0.000	0.062
0.205					
C(Location) [T.3]	-0.4672	0.091	-5.117	0.000	-0.646
-0.288					
C(Location) [T.4]	0.4582	0.117	3.931	0.000	0.230
0.687					
C(Location) [T.5]	-0.4001	0.090	-4.446	0.000	-0.577
-0.224					
C(Location) [T.6]	-1.9144	0.089	-21.590	0.000	-2.088
-1.741					
C(Location) [T.7]	-0.9243	0.091	-10.167	0.000	-1.102
-0.746					
C(Location) [T.8]	0.7258	0.087	8.330	0.000	0.555
0.897					
C(Location) [T.9]	0.4351	0.090	4.836	0.000	0.259
0.611					
C(Location) [T.10]	-0.5850	0.093	-6.272	0.000	-0.768
-0.402					
C(Location) [T.11]	-0.4205	0.098	-4.306	0.000	-0.612

-0.229					
C(Location) [T.12]	0.1506	0.088	1.710	0.087	-0.022
0.323					
C(Location) [T.13]	-1.2819	0.088	-14.598	0.000	-1.454
-1.110					
C(Location) [T.14]	0.0834	0.092	0.904	0.366	-0.097
0.264					
C(Location) [T.15]	0.1946	0.089	2.190	0.028	0.020
0.369					
C(Location) [T.16]	-1.0206	0.086	-11.874	0.000	-1.189
-0.852					
C(Location) [T.17]	0.2076	0.148	1.399	0.162	-0.083
0.499					
C(Location) [T.18]	-1.0445	0.098	-10.649	0.000	-1.237
-0.852					
C(Location) [T.19]	-0.2096	0.087	-2.409	0.016	-0.380
-0.039					
C(Location) [T.20]	-1.1124	0.086	-12.910	0.000	-1.281
-0.944					
C(Location) [T.21]	-1.1602	0.097	-11.903	0.000	-1.351
-0.969					
C(Location) [T.22]	0.1650	0.096	1.724	0.085	-0.023
0.353					
C(Location) [T.23]	-0.7712	0.084	-9.173	0.000	-0.936
-0.606					
C(Location) [T.26]	-1.6673	0.113	-14.804	0.000	-1.888
-1.447					
C(Location) [T.27]	-0.7715	0.084	-9.147	0.000	-0.937
-0.606					
C(Location) [T.28]	-0.7891	0.082	-9.582	0.000	-0.951
-0.628					
C(Location) [T.29]	-1.0498	0.092	-11.453	0.000	-1.229
-0.870					
C(Location) [T.30]	0.2861	0.092	3.125	0.002	0.107
0.466					
C(Location) [T.32]	-0.0197	0.088	-0.224	0.823	-0.192
0.153					
C(Location) [T.33]	0.0370	0.090	0.413	0.680	-0.139
0.213					
C(Location) [T.34]	-0.9852	0.082	-12.041	0.000	-1.146
-0.825					
C(Location) [T.35]	-0.3035	0.093	-3.253	0.001	-0.486
-0.121					
C(Location) [T.36]	-1.2750	0.087	-14.626	0.000	-1.446
-1.104					
C(Location) [T.38]	-0.3140	0.089	-3.536	0.000	-0.488
-0.140					
C(Location) [T.39]	-0.2675	0.086	-3.103	0.002	-0.436

-0.099					
C(Location) [T.40]	0.0119	0.096	0.124	0.901	-0.177
0.201					
C(Location) [T.41]	-0.3323	0.091	-3.654	0.000	-0.511
-0.154					
C(Location) [T.42]	0.3038	0.150	2.025	0.043	0.010
0.598					
C(Location) [T.43]	-0.5392	0.091	-5.905	0.000	-0.718
-0.360					
C(Location) [T.44]	-0.6407	0.084	-7.606	0.000	-0.806
-0.476					
C(Location) [T.45]	-1.1003	0.085	-12.917	0.000	-1.267
-0.933					
C(Location) [T.46]	0.2745	0.088	3.133	0.002	0.103
0.446					
C(Location) [T.47]	-0.1526	0.088	-1.743	0.081	-0.324
0.019					
C(Location) [T.48]	-0.9911	0.086	-11.545	0.000	-1.159
-0.823					
C(Location) [T.49]	-1.3271	0.110	-12.044	0.000	-1.543
-1.111					
C(Parameter1_Dir) [T.N]	-0.1703	0.051	-3.320	0.001	-0.271
-0.070					
C(Parameter1_Dir) [T.NE]	-0.0869	0.046	-1.873	0.061	-0.178
0.004					
C(Parameter1_Dir) [T.NW]	-0.0421	0.052	-0.808	0.419	-0.144
0.060					
C(Parameter1_Dir) [T.S]	-0.0440	0.045	-0.974	0.330	-0.132
0.044					
C(Parameter1_Dir) [T.SE]	-0.0785	0.042	-1.856	0.063	-0.161
0.004					
C(Parameter1_Dir) [T.SW]	0.0610	0.047	1.292	0.196	-0.031
0.153					
C(Parameter1_Dir) [T.W]	-0.0163	0.048	-0.336	0.737	-0.111
0.079					
C(Parameter2_9am) [T.N]	-0.0540	0.046	-1.161	0.246	-0.145
0.037					
C(Parameter2_9am) [T.NE]	0.1027	0.046	2.213	0.027	0.012
0.194					
C(Parameter2_9am) [T.NW]	0.0808	0.047	1.725	0.084	-0.011
0.173					
C(Parameter2_9am) [T.S]	0.2555	0.044	5.846	0.000	0.170
0.341					
C(Parameter2_9am) [T.SE]	0.1721	0.043	4.032	0.000	0.088
0.256					
C(Parameter2_9am) [T.SW]	0.4944	0.045	11.024	0.000	0.406
0.582					
C(Parameter2_9am) [T.W]	0.2442	0.046	5.298	0.000	0.154

0.335					
C(Parameter2_3pm) [T.N]	-0.0729	0.050	-1.456	0.145	-0.171
0.025					
C(Parameter2_3pm) [T.NE]	-0.1450	0.045	-3.238	0.001	-0.233
-0.057					
C(Parameter2_3pm) [T.NW]	0.1554	0.051	3.071	0.002	0.056
0.255					
C(Parameter2_3pm) [T.S]	0.0002	0.044	0.005	0.996	-0.086
0.086					
C(Parameter2_3pm) [T.SE]	-0.0085	0.040	-0.214	0.831	-0.087
0.069					
C(Parameter2_3pm) [T.SW]	-0.0262	0.047	-0.557	0.577	-0.118
0.066					
C(Parameter2_3pm) [T.W]	0.0982	0.048	2.049	0.040	0.004
0.192					
Min_Temp	0.1280	0.005	23.696	0.000	0.117
0.139					
Max_Temp	-0.2507	0.010	-26.257	0.000	-0.269
-0.232					
Parameter1_Speed	0.0347	0.001	30.783	0.000	0.033
0.037					
Parameter3_9am	0.0164	0.002	10.334	0.000	0.013
0.020					
Parameter3_3pm	-0.0225	0.002	-14.095	0.000	-0.026
-0.019					
Parameter4_9am	0.0705	0.001	62.932	0.000	0.068
0.073					
Parameter4_3pm	0.0034	0.001	3.010	0.003	0.001
0.006					
Parameter5_9am	-0.2248	0.006	-34.660	0.000	-0.238
-0.212					
Parameter5_3pm	0.1729	0.006	27.040	0.000	0.160
0.185					
Parameter7_9am	-0.0177	0.008	-2.090	0.037	-0.034
-0.001					
Parameter7_3pm	0.0962	0.010	9.284	0.000	0.076
0.117					

=====

=====

Logit Marginal Effects

=====

Dep. Variable:	Failure_today
Method:	dydx
At:	overall

=====

=====

	dy/dx	std err	z	P> z	[0.025
0.975]					

C(Estacion) [T.Verano] -0.034	-0.0403	0.003	-12.983	0.000	-0.046
C(Estacion) [T.Invierno] 0.003	-0.0032	0.003	-0.994	0.320	-0.009
C(Año) [T.2010] 0.026	0.0176	0.004	4.232	0.000	0.009
C(Año) [T.2011] 0.022	0.0136	0.004	3.193	0.001	0.005
C(Año) [T.2012] 0.024	0.0154	0.004	3.562	0.000	0.007
C(Año) [T.2013] 0.022	0.0138	0.004	3.236	0.001	0.005
C(Año) [T.2014] 0.022	0.0140	0.004	3.264	0.001	0.006
C(Año) [T.2015] 0.025	0.0167	0.004	3.878	0.000	0.008
C(Año) [T.2016] 0.023	0.0153	0.004	3.673	0.000	0.007
C(Location) [T.3] -0.033	-0.0535	0.010	-5.120	0.000	-0.074
C(Location) [T.4] 0.079	0.0525	0.013	3.932	0.000	0.026
C(Location) [T.5] -0.026	-0.0458	0.010	-4.448	0.000	-0.066
C(Location) [T.6] -0.200	-0.2193	0.010	-21.772	0.000	-0.239
C(Location) [T.7] -0.086	-0.1059	0.010	-10.186	0.000	-0.126
C(Location) [T.8] 0.103	0.0832	0.010	8.336	0.000	0.064
C(Location) [T.9] 0.070	0.0498	0.010	4.837	0.000	0.030
C(Location) [T.10] -0.046	-0.0670	0.011	-6.278	0.000	-0.088
C(Location) [T.11] -0.026	-0.0482	0.011	-4.307	0.000	-0.070
C(Location) [T.12] 0.037	0.0173	0.010	1.710	0.087	-0.003
C(Location) [T.13] -0.127	-0.1469	0.010	-14.655	0.000	-0.167
C(Location) [T.14] 0.030	0.0096	0.011	0.904	0.366	-0.011
C(Location) [T.15] 0.042	0.0223	0.010	2.190	0.028	0.002
C(Location) [T.16] -0.098	-0.1169	0.010	-11.908	0.000	-0.136

C(Location) [T.17] 0.057	0.0238	0.017	1.399	0.162	-0.010
C(Location) [T.18] -0.098	-0.1197	0.011	-10.673	0.000	-0.142
C(Location) [T.19] -0.004	-0.0240	0.010	-2.410	0.016	-0.044
C(Location) [T.20] -0.108	-0.1274	0.010	-12.953	0.000	-0.147
C(Location) [T.21] -0.111	-0.1329	0.011	-11.933	0.000	-0.155
C(Location) [T.22] 0.040	0.0189	0.011	1.724	0.085	-0.003
C(Location) [T.23] -0.070	-0.0884	0.010	-9.187	0.000	-0.107
C(Location) [T.26] -0.166	-0.1910	0.013	-14.864	0.000	-0.216
C(Location) [T.27] -0.069	-0.0884	0.010	-9.164	0.000	-0.107
C(Location) [T.28] -0.072	-0.0904	0.009	-9.601	0.000	-0.109
C(Location) [T.29] -0.100	-0.1203	0.010	-11.478	0.000	-0.141
C(Location) [T.30] 0.053	0.0328	0.010	3.126	0.002	0.012
C(Location) [T.32] 0.018	-0.0023	0.010	-0.224	0.823	-0.022
C(Location) [T.33] 0.024	0.0042	0.010	0.413	0.680	-0.016
C(Location) [T.34] -0.095	-0.1129	0.009	-12.074	0.000	-0.131
C(Location) [T.35] -0.014	-0.0348	0.011	-3.254	0.001	-0.056
C(Location) [T.36] -0.127	-0.1461	0.010	-14.689	0.000	-0.166
C(Location) [T.38] -0.016	-0.0360	0.010	-3.537	0.000	-0.056
C(Location) [T.39] -0.011	-0.0306	0.010	-3.104	0.002	-0.050
C(Location) [T.40] 0.023	0.0014	0.011	0.124	0.901	-0.020
C(Location) [T.41] -0.018	-0.0381	0.010	-3.655	0.000	-0.058
C(Location) [T.42] 0.069	0.0348	0.017	2.025	0.043	0.001
C(Location) [T.43] -0.041	-0.0618	0.010	-5.908	0.000	-0.082
C(Location) [T.44] -0.055	-0.0734	0.010	-7.614	0.000	-0.092

C(Location) [T.45] -0.107	-0.1261	0.010	-12.959	0.000	-0.145
C(Location) [T.46] 0.051	0.0314	0.010	3.133	0.002	0.012
C(Location) [T.47] 0.002	-0.0175	0.010	-1.743	0.081	-0.037
C(Location) [T.48] -0.094	-0.1135	0.010	-11.577	0.000	-0.133
C(Location) [T.49] -0.127	-0.1520	0.013	-12.072	0.000	-0.177
C(Parameter1_Dir) [T.N] -0.008	-0.0195	0.006	-3.321	0.001	-0.031
C(Parameter1_Dir) [T.NE] 0.000	-0.0100	0.005	-1.874	0.061	-0.020
C(Parameter1_Dir) [T.NW] 0.007	-0.0048	0.006	-0.808	0.419	-0.017
C(Parameter1_Dir) [T.S] 0.005	-0.0050	0.005	-0.974	0.330	-0.015
C(Parameter1_Dir) [T.SE] 0.001	-0.0090	0.005	-1.856	0.063	-0.018
C(Parameter1_Dir) [T.SW] 0.018	0.0070	0.005	1.292	0.196	-0.004
C(Parameter1_Dir) [T.W] 0.009	-0.0019	0.006	-0.336	0.737	-0.013
C(Parameter2_9am) [T.N] 0.004	-0.0062	0.005	-1.161	0.246	-0.017
C(Parameter2_9am) [T.NE] 0.022	0.0118	0.005	2.213	0.027	0.001
C(Parameter2_9am) [T.NW] 0.020	0.0093	0.005	1.725	0.084	-0.001
C(Parameter2_9am) [T.S] 0.039	0.0293	0.005	5.849	0.000	0.019
C(Parameter2_9am) [T.SE] 0.029	0.0197	0.005	4.033	0.000	0.010
C(Parameter2_9am) [T.SW] 0.067	0.0566	0.005	11.045	0.000	0.047
C(Parameter2_9am) [T.W] 0.038	0.0280	0.005	5.300	0.000	0.018
C(Parameter2_3pm) [T.N] 0.003	-0.0084	0.006	-1.457	0.145	-0.020
C(Parameter2_3pm) [T.NE] -0.007	-0.0166	0.005	-3.238	0.001	-0.027
C(Parameter2_3pm) [T.NW] 0.029	0.0178	0.006	3.071	0.002	0.006
C(Parameter2_3pm) [T.S] 0.010	2.753e-05	0.005	0.005	0.996	-0.010
C(Parameter2_3pm) [T.SE] 0.008	-0.0010	0.005	-0.214	0.831	-0.010

C(Parameter2_3pm) [T.SW] 0.008	-0.0030	0.005	-0.557	0.577	-0.014
C(Parameter2_3pm) [T.W] 0.022	0.0113	0.005	2.049	0.040	0.000
Min_Temp 0.016	0.0147	0.001	23.898	0.000	0.013
Max_Temp -0.027	-0.0287	0.001	-26.564	0.000	-0.031
Parameter1_Speed 0.004	0.0040	0.000	31.369	0.000	0.004
Parameter3_9am 0.002	0.0019	0.000	10.357	0.000	0.002
Parameter3_3pm -0.002	-0.0026	0.000	-14.154	0.000	-0.003
Parameter4_9am 0.008	0.0081	0.000	67.472	0.000	0.008
Parameter4_3pm 0.001	0.0004	0.000	3.011	0.003	0.000
Parameter5_9am -0.024	-0.0258	0.001	-35.438	0.000	-0.027
Parameter5_3pm 0.021	0.0198	0.001	27.399	0.000	0.018
Parameter7_9am -0.000	-0.0020	0.001	-2.091	0.037	-0.004
Parameter7_3pm 0.013	0.0110	0.001	9.302	0.000	0.009
=====					
=====					

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 investigación y por qué? ¿Qué variables resultaron ser robustas a la especificación?

R: Las principales diferencias se encuentran en la significancia de ciertas variables en algunos modelos y además de la magnitud de los coeficientes. Estas diferencias se originan por la forma de resolución que hay tras cada modelo. Consideremos que OLS busca predecir una variable continua y en este caso forzamos a predecir una variable binaria, lo que puede resultar en interpretaciones erróneas.

Probit y Logit si nos permite operar con una estimación de una variable binaria, además permite estudiar la no linealidad del modelo. Además restringe el valor de probabilidad obtenido entre [0,1]. Considerando la similitud de los resultados, se recomendaría cualquiera de los dos para poder la pregunta de investigación.

En estos 3 modelos, las principales variables significativas fueron, Max Temp, Min Temp y el cambio estacional a Verano, que fueron consistentes con la significancia y el impacto en la estimación de la variable.

0.6 Agrupación por mes

```
[177]: df_mes=df2.copy()
```

Creamos una variable que junte el mes y el año para facilitar el estudio de los datos

```
[178]: df_mes['MesYear']= df_mes['Mes'] + " " + df_mes['Año'].astype(str)
```

Agrupamos por Mes/Año y Location (sensor), de esta forma tendremos el promedio de los parámetros de cada sensor, en cada mes y en cada año, además de la suma de todas las fallas detectadas por cada sensor en cada mes y en cada año.

Para continuar con el estudio de la variación estacional, se mantiene la variable estación.

```
[179]: df_mes=df_mes.groupby(['MesYear','Location']).agg({'Min_Temp': 'mean',
                                                         'Max_Temp': 'mean',
                                                         'Parameter1_Speed': 'mean',
                                                         'Parameter3_9am': 'mean',
                                                         'Parameter3_3pm': 'mean',
                                                         'Parameter4_9am': 'mean',
                                                         'Parameter4_3pm': 'mean',
                                                         'Parameter5_9am': 'mean',
                                                         'Parameter5_3pm': 'mean',
                                                         'Parameter7_9am': 'mean',
                                                         'Parameter7_3pm': 'mean',
                                                         'Estacion': 'first',
                                                         'Failure_today': 'sum',
                                                         }).reset_index()
```

```
[180]: df_mes
```

```
[180]:
```

	MesYear	Location	Min_Temp	Max_Temp	Parameter1_Speed	\
0	Abril 2009	1	13.21	22.79	35.86	
1	Abril 2009	3	9.09	20.84	37.38	
2	Abril 2009	4	13.73	28.67	38.96	
3	Abril 2009	5	12.35	22.23	35.07	
4	Abril 2009	6	6.56	18.02	44.11	
...	
3821	Septiembre 2016	45	8.69	16.63	34.19	
3822	Septiembre 2016	46	10.25	21.39	49.92	
3823	Septiembre 2016	47	7.49	15.63	45.04	
3824	Septiembre 2016	48	12.99	19.44	43.23	
3825	Septiembre 2016	49	8.28	20.70	48.50	

	Parameter3_9am	Parameter3_3pm	Parameter4_9am	Parameter4_3pm	\
0	10.04	16.50	57.82	45.86	
1	9.76	14.05	67.62	44.19	
2	15.72	18.28	30.48	16.92	
3	12.44	16.78	75.96	57.67	

4	19.43	20.89	78.64	52.93
...
3821	10.50	12.92	82.65	62.85
3822	21.62	26.15	67.15	54.38
3823	16.04	20.07	71.64	65.86
3824	18.10	18.03	62.13	61.73
3825	23.30	26.90	58.77	32.40

	Parameter5_9am	Parameter5_3pm	Parameter7_9am	Parameter7_3pm	Estacion	\
0	1019.62	1017.43	17.61	21.36	Otro	
1	1019.12	1016.09	13.90	19.96	Otro	
2	1018.35	1014.53	20.83	27.87	Otro	
3	1019.46	1017.17	17.43	20.95	Otro	
4	1018.72	1016.76	11.67	16.67	Otro	
...	
3821	1015.36	1013.15	11.90	15.31	Otro	
3822	1016.32	1013.29	16.10	20.02	Otro	
3823	1019.45	1017.58	12.39	13.80	Otro	
3824	1014.82	1012.27	16.68	18.12	Otro	
3825	1016.95	1013.74	13.61	19.51	Otro	

	Failure_today
0	6
1	6
2	0
3	9
4	8
...	...
3821	10
3822	4
3823	16
3824	4
3825	7

[3826 rows x 15 columns]

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.

```
[181]: poisson=smf.glm('Failure_today ~ C(Estacion) + C(Location) + Min_Temp +
    ↪Max_Temp + Parameter1_Speed + Parameter3_9am + Parameter3_3pm +
    ↪Parameter4_9am + Parameter4_3pm + Parameter5_9am + Parameter5_3pm +
    ↪Parameter7_9am + Parameter7_3pm' , data=df_mes,family=sm.families.Poisson()).
    ↪fit()
```

```
print(poisson.summary())
```

Generalized Linear Model Regression Results

```
=====
Dep. Variable:          Failure_today    No. Observations:          3826
Model:                  GLM              Df Residuals:            3769
Model Family:           Poisson          Df Model:                56
Link Function:          Log              Scale:                  1.0000
Method:                 IRLS             Log-Likelihood:          -8489.8
Date:                   jue, 24 abr. 2025 Deviance:                 4055.0
Time:                   23:54:51          Pearson chi2:             3.64e+03
No. Iterations:         5                 Pseudo R-squ. (CS):       0.8599
Covariance Type:        nonrobust
=====
```

```
=====
                                coef      std err          z      P>|z|      [0.025
0.975]
-----
-----
Intercept                    20.6973      3.120        6.633      0.000      14.582
26.813
C(Estacion) [T.Verano]      -0.1492      0.024       -6.283      0.000      -0.196
-0.103
C(Estacion) [T.Invierno]    -0.0984      0.024       -4.167      0.000      -0.145
-0.052
C(Location) [T.3]           -0.1741      0.069       -2.539      0.011      -0.308
-0.040
C(Location) [T.4]            0.1165      0.089        1.312      0.190      -0.058
0.291
C(Location) [T.5]           -0.2963      0.072       -4.123      0.000      -0.437
-0.155
C(Location) [T.6]           -0.5123      0.077       -6.614      0.000      -0.664
-0.360
C(Location) [T.7]           -0.2671      0.069       -3.847      0.000      -0.403
-0.131
C(Location) [T.8]           -0.0341      0.070       -0.484      0.628      -0.172
0.104
C(Location) [T.9]            0.0315      0.080        0.392      0.695      -0.126
0.189
C(Location) [T.10]          -0.1984      0.076       -2.605      0.009      -0.348
-0.049
C(Location) [T.11]          -0.0366      0.072       -0.509      0.611      -0.178
0.104
C(Location) [T.12]           0.0062      0.068        0.091      0.928      -0.128
0.140
C(Location) [T.13]          -0.5371      0.070       -7.723      0.000      -0.673
-0.401
C(Location) [T.14]          -0.2878      0.084       -3.434      0.001      -0.452
=====
```

-0.124					
C(Location) [T.15]	-0.0581	0.079	-0.740	0.459	-0.212
0.096					
C(Location) [T.16]	-0.6611	0.063	-10.572	0.000	-0.784
-0.539					
C(Location) [T.17]	-0.5012	0.127	-3.936	0.000	-0.751
-0.252					
C(Location) [T.18]	-0.5405	0.073	-7.355	0.000	-0.685
-0.396					
C(Location) [T.19]	-0.1002	0.065	-1.542	0.123	-0.227
0.027					
C(Location) [T.20]	-0.3402	0.069	-4.920	0.000	-0.476
-0.205					
C(Location) [T.21]	-0.1565	0.079	-1.984	0.047	-0.311
-0.002					
C(Location) [T.22]	0.0223	0.082	0.271	0.786	-0.139
0.184					
C(Location) [T.23]	-0.1182	0.066	-1.782	0.075	-0.248
0.012					
C(Location) [T.26]	-0.2905	0.091	-3.175	0.001	-0.470
-0.111					
C(Location) [T.27]	-0.5549	0.064	-8.720	0.000	-0.680
-0.430					
C(Location) [T.28]	-0.5425	0.067	-8.045	0.000	-0.675
-0.410					
C(Location) [T.29]	-0.2420	0.066	-3.676	0.000	-0.371
-0.113					
C(Location) [T.30]	0.0460	0.069	0.663	0.507	-0.090
0.182					
C(Location) [T.32]	-0.0371	0.064	-0.581	0.561	-0.162
0.088					
C(Location) [T.33]	0.1028	0.068	1.514	0.130	-0.030
0.236					
C(Location) [T.34]	-0.3147	0.062	-5.094	0.000	-0.436
-0.194					
C(Location) [T.35]	-0.4918	0.072	-6.837	0.000	-0.633
-0.351					
C(Location) [T.36]	-0.3023	0.072	-4.205	0.000	-0.443
-0.161					
C(Location) [T.38]	-0.2412	0.063	-3.821	0.000	-0.365
-0.117					
C(Location) [T.39]	-0.0521	0.065	-0.800	0.424	-0.180
0.076					
C(Location) [T.40]	-0.2894	0.086	-3.364	0.001	-0.458
-0.121					
C(Location) [T.41]	-0.2616	0.069	-3.774	0.000	-0.397
-0.126					
C(Location) [T.42]	0.0356	0.116	0.308	0.758	-0.191

0.263					
C(Location) [T.43]	0.0180	0.070	0.259	0.796	-0.118
0.154					
C(Location) [T.44]	-0.6514	0.062	-10.556	0.000	-0.772
-0.530					
C(Location) [T.45]	-0.5025	0.063	-7.997	0.000	-0.626
-0.379					
C(Location) [T.46]	0.1006	0.069	1.454	0.146	-0.035
0.236					
C(Location) [T.47]	-0.3757	0.065	-5.821	0.000	-0.502
-0.249					
C(Location) [T.48]	-0.7622	0.064	-11.833	0.000	-0.888
-0.636					
C(Location) [T.49]	-0.4235	0.090	-4.705	0.000	-0.600
-0.247					
Min_Temp	0.0544	0.011	5.123	0.000	0.034
0.075					
Max_Temp	-0.0135	0.025	-0.541	0.588	-0.062
0.035					
Parameter1_Speed	0.0585	0.003	20.656	0.000	0.053
0.064					
Parameter3_9am	-0.0027	0.004	-0.670	0.503	-0.011
0.005					
Parameter3_3pm	-0.0678	0.004	-17.250	0.000	-0.075
-0.060					
Parameter4_9am	0.0313	0.003	12.176	0.000	0.026
0.036					
Parameter4_3pm	0.0020	0.003	0.685	0.493	-0.004
0.008					
Parameter5_9am	-0.0583	0.019	-3.075	0.002	-0.096
-0.021					
Parameter5_3pm	0.0379	0.019	1.993	0.046	0.001
0.075					
Parameter7_9am	0.0991	0.016	6.111	0.000	0.067
0.131					
Parameter7_3pm	-0.1424	0.028	-5.090	0.000	-0.197
-0.088					

=====

=====

Calculamos los odds ratios para medir la variación porcentual que aporta cada coeficiente:

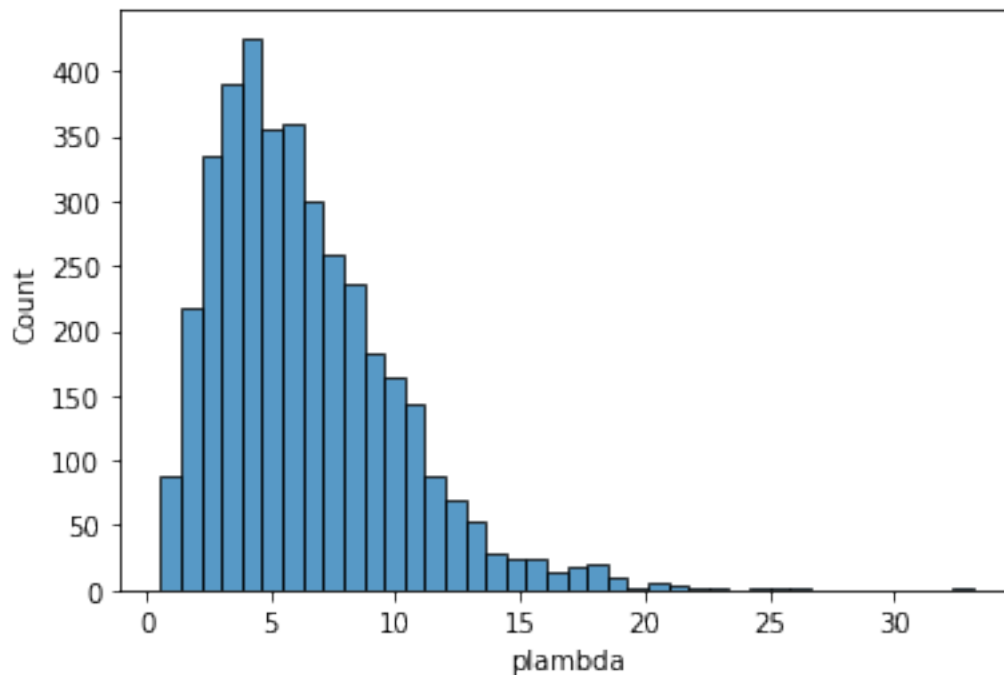
```
[182]: pd.set_option('display.float_format', '{:.2f}'.format)
coef=poisson.params
odds_ratios = np.exp(coef)
odds_ratios = 100*(odds_ratios-1)
odds_ratios
```

[182]: Intercept	97441192603.56
C(Estacion) [T.Verano]	-13.86
C(Estacion) [T.Invierno]	-9.37
C(Location) [T.3]	-15.98
C(Location) [T.4]	12.36
C(Location) [T.5]	-25.64
C(Location) [T.6]	-40.09
C(Location) [T.7]	-23.44
C(Location) [T.8]	-3.36
C(Location) [T.9]	3.20
C(Location) [T.10]	-18.00
C(Location) [T.11]	-3.60
C(Location) [T.12]	0.62
C(Location) [T.13]	-41.55
C(Location) [T.14]	-25.01
C(Location) [T.15]	-5.65
C(Location) [T.16]	-48.37
C(Location) [T.17]	-39.42
C(Location) [T.18]	-41.76
C(Location) [T.19]	-9.53
C(Location) [T.20]	-28.83
C(Location) [T.21]	-14.48
C(Location) [T.22]	2.26
C(Location) [T.23]	-11.14
C(Location) [T.26]	-25.21
C(Location) [T.27]	-42.59
C(Location) [T.28]	-41.87
C(Location) [T.29]	-21.49
C(Location) [T.30]	4.71
C(Location) [T.32]	-3.64
C(Location) [T.33]	10.82
C(Location) [T.34]	-27.00
C(Location) [T.35]	-38.85
C(Location) [T.36]	-26.09
C(Location) [T.38]	-21.43
C(Location) [T.39]	-5.07
C(Location) [T.40]	-25.13
C(Location) [T.41]	-23.01
C(Location) [T.42]	3.63
C(Location) [T.43]	1.82
C(Location) [T.44]	-47.87
C(Location) [T.45]	-39.50
C(Location) [T.46]	10.59
C(Location) [T.47]	-31.32
C(Location) [T.48]	-53.34
C(Location) [T.49]	-34.53
Min_Temp	5.59

Max_Temp	-1.34
Parameter1_Speed	6.03
Parameter3_9am	-0.27
Parameter3_3pm	-6.55
Parameter4_9am	3.18
Parameter4_3pm	0.20
Parameter5_9am	-5.67
Parameter5_3pm	3.86
Parameter7_9am	10.42
Parameter7_3pm	-13.27
dtype:	float64

```
[183]: df_mes['plambda'] = poisson.mu
sns.histplot(data=df_mes, x="plambda", bins=40)
```

```
[183]: <AxesSubplot: xlabel='plambda', ylabel='Count'>
```



R: Para construir la distribución Poisson, consideramos solo variables continuas, además de incorporar las variables categóricas de Location y Estacion, para poder analizar el cambio en torno a las categorías base.

Interpretando los coeficientes más significativos, podemos notar que al cambiar de la estación Otros a Verano o Invierno, hay una disminución del 13.7% y un 9.94% respectivamente en la tasa esperada de fallas para las categorías bases.

Además, al aumentar en una unidad la temperatura mínima, se espera un incremento de 5.26% en

la tasa esperada de fallas.

Se destaca que la variación de la temperatura máxima detectada, no se considera significativa para estimar una variación en la tasa esperada.

Notamos que la mayor cantidad de errores estimados por sensor se encuentra entre 4 y 7, y el gráfico de ocurrencias sigue gráficamente una distribución Poisson. Lo que podría indicar que no existe una sobre dispersión.

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

Los resultados del test de dispersión nos arrojan que el parámetro de la regresión auxiliar es significativo, lo que puede demostrar cierto nivel de sobredispersión, es decir, el modelo de Poisson podría no capturar toda la varianza, por lo que convendría usar Binomial Negativa.

```
[184]: aux=((df_mes['Failure_today']-poisson.mu)**2-poisson.mu)/poisson.mu
auxr=sm.OLS(aux,poisson.mu).fit()
print(auxr.summary())
```

```

                                OLS Regression Results
=====
Dep. Variable:                Failure_today    R-squared (uncentered):
0.001
Model:                        OLS             Adj. R-squared (uncentered):
0.001
Method:                      Least Squares    F-statistic:
4.148
Date:                        jue, 24 abr. 2025    Prob (F-statistic):
0.0418
Time:                        23:54:52           Log-Likelihood:
-6700.8
No. Observations:            3826             AIC:
1.340e+04
Df Residuals:                3825             BIC:
1.341e+04
Df Model:                    1
Covariance Type:             nonrobust
=====
                                coef    std err          t      P>|t|      [0.025    0.975]
-----
x1                -0.0062    0.003      -2.037    0.042    -0.012    -0.000
=====
Omnibus:                 3697.998    Durbin-Watson:                 1.814
Prob(Omnibus):           0.000    Jarque-Bera (JB):              309363.869
Skew:                    4.452    Prob(JB):                      0.00
Kurtosis:                46.143    Cond. No.                      1.00
=====
```


Notes:

[1] R^2 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.

Para binomial negativa, utilizamos la exponencial del coeficiente obtenido en la regresión auxiliar para poder estimar el valor de alpha:

```
[185]: a=np.exp(-0.0068)
```

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

R: A modo general , el sensor (location) y la estación del año son factores relevantes y explican bastante de la variación del número de fallas.

Se continúa con el resultado de a mayor mínima, más fallas; a mayor máxima, menos, pero con magnitudes distintas a Poisson, sin embargo, BN detecta que ninguna de estas dos variables es significativas.

Notamos como el parámetro 7 tiene una variación positiva si aumenta en una unidad por la mañana y un efecto completamente contrario si aumenta en una unidad por la tarde. (17% y -16% aproximadamente).

Por último, la mayoría de los sensores tiene efectos negativos (menor riesgo de falla respecto a la ubicación de referencia). Los sensores más relevantes son:

- 44: -54.44%
- 48: -57.26%

```
[186]: negbin=smf.glm('Failure_today ~ C(Estacion) + C(Location) + Min_Temp + Max_Temp_
↪ Parameter1_Speed + Parameter3_9am + Parameter3_3pm + Parameter4_9am +_
↪ Parameter4_3pm + Parameter5_9am + Parameter5_3pm + Parameter7_9am +_
↪ Parameter7_3pm',
                    data=df_mes,
                    family=sm.families.NegativeBinomial(alpha=a)).fit()
print(negbin.summary())
```

Generalized Linear Model Regression Results

```
=====
Dep. Variable:          Failure_today    No. Observations:          3826
Model:                  GLM              Df Residuals:            3769
Model Family:          NegativeBinomial  Df Model:                  56
Link Function:          Log              Scale:                   1.0000
Method:                 IRLS             Log-Likelihood:         -10596.
Date:                   jue, 24 abr. 2025 Deviance:                 937.88
Time:                   23:54:52          Pearson chi2:              686.
No. Iterations:         9                 Pseudo R-squ. (CS):       0.2673
Covariance Type:        nonrobust
```

	coef	std err	z	P> z	[0.025
0.975]					

Intercept	21.5672	8.816	2.446	0.014	4.289
38.846					
C(Estacion) [T.Verano]	-0.1676	0.067	-2.496	0.013	-0.299
-0.036					
C(Estacion) [T.Invierno]	-0.1333	0.064	-2.094	0.036	-0.258
-0.009					
C(Location) [T.3]	-0.0902	0.187	-0.484	0.629	-0.456
0.275					
C(Location) [T.4]	0.0460	0.208	0.222	0.825	-0.361
0.453					
C(Location) [T.5]	-0.4150	0.194	-2.144	0.032	-0.794
-0.036					
C(Location) [T.6]	-0.5424	0.220	-2.464	0.014	-0.974
-0.111					
C(Location) [T.7]	-0.2159	0.188	-1.150	0.250	-0.584
0.152					
C(Location) [T.8]	-0.2023	0.195	-1.039	0.299	-0.584
0.179					
C(Location) [T.9]	-0.1824	0.230	-0.794	0.427	-0.633
0.268					
C(Location) [T.10]	-0.1668	0.203	-0.820	0.412	-0.566
0.232					
C(Location) [T.11]	0.0040	0.179	0.022	0.982	-0.348
0.356					
C(Location) [T.12]	-0.1206	0.197	-0.612	0.540	-0.507
0.266					
C(Location) [T.13]	-0.6068	0.201	-3.024	0.002	-1.000
-0.214					
C(Location) [T.14]	-0.7174	0.239	-2.997	0.003	-1.186
-0.248					
C(Location) [T.15]	-0.1850	0.219	-0.845	0.398	-0.614
0.244					
C(Location) [T.16]	-0.6745	0.177	-3.808	0.000	-1.022
-0.327					
C(Location) [T.17]	-0.9506	0.329	-2.885	0.004	-1.596
-0.305					
C(Location) [T.18]	-0.5707	0.202	-2.831	0.005	-0.966
-0.176					
C(Location) [T.19]	-0.0555	0.190	-0.292	0.770	-0.428
0.317					
C(Location) [T.20]	-0.3193	0.195	-1.639	0.101	-0.701
0.063					

C(Location) [T.21] 0.322	-0.0572	0.194	-0.296	0.768	-0.437
C(Location) [T.22] 0.431	0.0079	0.216	0.037	0.971	-0.416
C(Location) [T.23] 0.243	-0.1429	0.197	-0.726	0.468	-0.529
C(Location) [T.26] 0.257	-0.2138	0.240	-0.891	0.373	-0.684
C(Location) [T.27] -0.286	-0.6422	0.182	-3.536	0.000	-0.998
C(Location) [T.28] -0.332	-0.7118	0.194	-3.670	0.000	-1.092
C(Location) [T.29] 0.103	-0.2499	0.180	-1.388	0.165	-0.603
C(Location) [T.30] 0.318	-0.0573	0.191	-0.299	0.765	-0.433
C(Location) [T.32] 0.126	-0.2151	0.174	-1.237	0.216	-0.556
C(Location) [T.33] 0.332	-0.0380	0.189	-0.202	0.840	-0.408
C(Location) [T.34] -0.062	-0.4245	0.185	-2.292	0.022	-0.787
C(Location) [T.35] -0.165	-0.5365	0.189	-2.833	0.005	-0.908
C(Location) [T.36] 0.057	-0.3316	0.198	-1.671	0.095	-0.721
C(Location) [T.38] 0.087	-0.2688	0.182	-1.479	0.139	-0.625
C(Location) [T.39] 0.325	-0.0404	0.187	-0.217	0.828	-0.406
C(Location) [T.40] -0.105	-0.5615	0.233	-2.412	0.016	-1.018
C(Location) [T.41] 0.167	-0.1958	0.185	-1.058	0.290	-0.559
C(Location) [T.42] 0.538	0.0468	0.251	0.187	0.852	-0.444
C(Location) [T.43] 0.438	0.0698	0.188	0.372	0.710	-0.298
C(Location) [T.44] -0.446	-0.8009	0.181	-4.418	0.000	-1.156
C(Location) [T.45] -0.189	-0.5389	0.178	-3.022	0.003	-0.888
C(Location) [T.46] 0.482	0.0950	0.197	0.482	0.630	-0.292
C(Location) [T.47] -0.177	-0.5453	0.188	-2.901	0.004	-0.914
C(Location) [T.48] -0.508	-0.8603	0.180	-4.789	0.000	-1.212

C(Location) [T.49]	-0.3432	0.202	-1.702	0.089	-0.738
0.052					
Min_Temp	0.0398	0.027	1.467	0.142	-0.013
0.093					
Max_Temp	0.0056	0.067	0.084	0.933	-0.126
0.137					
Parameter1_Speed	0.0691	0.008	8.563	0.000	0.053
0.085					
Parameter3_9am	-0.0007	0.011	-0.064	0.949	-0.022
0.021					
Parameter3_3pm	-0.0858	0.011	-7.932	0.000	-0.107
-0.065					
Parameter4_9am	0.0400	0.007	5.867	0.000	0.027
0.053					
Parameter4_3pm	-6.162e-05	0.008	-0.008	0.994	-0.015
0.015					
Parameter5_9am	-0.0673	0.052	-1.298	0.194	-0.169
0.034					
Parameter5_3pm	0.0454	0.052	0.870	0.384	-0.057
0.148					
Parameter7_9am	0.1615	0.041	3.902	0.000	0.080
0.243					
Parameter7_3pm	-0.1979	0.075	-2.637	0.008	-0.345
-0.051					

=====

=====

Obtenemos los odds ratios para medir las variaciones porcentuales:

```
[187]: coef=negbin.params
odds_ratios = np.exp(coef)
odds_ratios = 100*(odds_ratios-1)
odds_ratios
```

```
[187]: Intercept                232559407886.95
C(Estacion) [T.Verano]         -15.43
C(Estacion) [T.Invierno]       -12.48
C(Location) [T.3]               -8.62
C(Location) [T.4]               4.71
C(Location) [T.5]              -33.96
C(Location) [T.6]              -41.86
C(Location) [T.7]              -19.42
C(Location) [T.8]              -18.32
C(Location) [T.9]              -16.67
C(Location) [T.10]             -15.37
C(Location) [T.11]              0.40
C(Location) [T.12]             -11.36
C(Location) [T.13]             -45.49
```

C(Location) [T.14]	-51.20
C(Location) [T.15]	-16.89
C(Location) [T.16]	-49.06
C(Location) [T.17]	-61.35
C(Location) [T.18]	-43.49
C(Location) [T.19]	-5.40
C(Location) [T.20]	-27.33
C(Location) [T.21]	-5.56
C(Location) [T.22]	0.79
C(Location) [T.23]	-13.32
C(Location) [T.26]	-19.25
C(Location) [T.27]	-47.39
C(Location) [T.28]	-50.92
C(Location) [T.29]	-22.11
C(Location) [T.30]	-5.57
C(Location) [T.32]	-19.35
C(Location) [T.33]	-3.73
C(Location) [T.34]	-34.59
C(Location) [T.35]	-41.52
C(Location) [T.36]	-28.22
C(Location) [T.38]	-23.57
C(Location) [T.39]	-3.96
C(Location) [T.40]	-42.97
C(Location) [T.41]	-17.78
C(Location) [T.42]	4.79
C(Location) [T.43]	7.23
C(Location) [T.44]	-55.11
C(Location) [T.45]	-41.66
C(Location) [T.46]	9.97
C(Location) [T.47]	-42.03
C(Location) [T.48]	-57.70
C(Location) [T.49]	-29.05
Min_Temp	4.06
Max_Temp	0.57
Parameter1_Speed	7.15
Parameter3_9am	-0.07
Parameter3_3pm	-8.22
Parameter4_9am	4.08
Parameter4_3pm	-0.01
Parameter5_9am	-6.51
Parameter5_3pm	4.64
Parameter7_9am	17.53
Parameter7_3pm	-17.96
dtype: float64	

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 investigación y por qué? ¿Qué variables resultaron ser robustas a la especificación?

En primer lugar, en los cambios estacionales los efectos(odds ratios) entre Poisson BN son relativamente similares, aunque en Poisson son levemente menos negativos, lo cual puede indicar que el modelo Poisson subestima un poco la reducción de fallas en algunas estaciones.

En torno al cambio de un sensor a otro, el modelo Poisson tiende a exagerar efectos (más negativos o más positivos) en algunos sensores (Como el 3, 4, 33 y 38), probablemente como resultado de no ajustar adecuadamente por la dispersión extra. La Binomial Negativa, al incorporar el parámetro alpha, suaviza estos efectos.

En general, poisson tiende a inflar o subestimar algunos efectos, particularmente cuando hay mucha varianza en los datos, además el patrón de los signos (positivo/negativo) se mantiene en la mayoría de las variables, pero las magnitudes difieren, sobre todo en los sensores.

En mi opinión, y luego de realizar el test de dispersión, es adecuado utilizar el modelo de Binomial Negativa, ya que al presentar sobre dispersión en los datos, el modelo BN ajusta este error con la ayuda del coeficiente alpha, obteniendo estimaciones más precisas. Por último, algunas variables robustas en el estudio, que mantuvieron su significancia, coherencia y nivel de impacto sobre la variable explicativa fueron el Parámetro 1 de velocidad y el cambio a la estación Invierno.

[]: