

Tarea1__Aravena__Uribe

April 30, 2025

1 Tarea 1 DAML Prueba 2

1.0.1 Joaquín Araena Uribe

1.0.2 2021454110

```
[87]: #Importamos las librerias

import pandas as pd
import missingno as msno
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import statsmodels.api as sm
import statsmodels.formula.api as smf
import sklearn
import scipy
from scipy.stats import nbinom
from statsmodels.iolib.summary2 import summary_col
import time as time

sns.set_palette("muted")
```

```
[88]: #Funciones a utilizar

def cardinales(x):
    if x[0] == "N":
        return("N")
    elif x[0] == "S":
        return("S")
    elif x[0] == "E":
        return("E")
    else:
        return("W")
```

1.0.3 Parte 1

Tratamiento de datos

```
[89]: df = pd.read_csv("machine_failure_data.csv") #Leemos el dataframe
df["Failure_today"] = df["Failure_today"].apply(lambda x: 0 if x == "No" else 1)
↳1) #Modificamos la variable a binaria
#df["Leakage"] = df["Leakage"].apply(lambda x: 0 if x == 0 else 1)

df["Date"] = pd.to_datetime(df["Date"], format="%m/%d/%Y") #Aplicamos arreglo a
↳las fechas
df = df[~df["Date"].dt.year.isin([2007,2008,2009])]
df["Date"] = df["Date"].dt.date
df["Date"] = df["Date"].astype(str)

df["Fecha"] = pd.to_datetime(df["Date"])
df["Año"] = df["Fecha"].dt.year.apply(str)
df["Mes"] = df["Fecha"].dt.month
#df["Trimestre"] = df["Mes"].apply(lambda x: f"T{((x - 1) // 3) + 1}")
df["Mes"] = df["Mes"].apply(str)

#df["Electricity"] = df["Electricity"].fillna(0)
#df["Evaporation"] = df["Evaporation"].fillna(0)
#df["Electricity_NaN"] = df["Electricity"].isna().astype(int)
#df["Evaporation_NaN"] = df["Evaporation"].isna().astype(int)

df_mask = df[(df["Location"] == 17) | (df["Location"] == 26) | (df["Location"]
↳== 42) | (df["Location"] == 46) | (df["Location"] == 19)] #Eliminamos años con
↳pocos datos
df_filtrado = df.drop(df_mask.index)
df = df_filtrado.reset_index(drop=True)

df
```

```
[89]:
```

	Date	Location	Min_Temp	Max_Temp	Leakage	Evaporation	\
0	2010-01-01	3	19.4	31.9	5.0	NaN	
1	2010-01-02	3	18.6	29.1	12.4	NaN	
2	2010-01-03	3	12.2	29.7	0.0	NaN	
3	2010-01-04	3	14.8	32.8	0.0	NaN	
4	2010-01-05	3	15.0	35.8	0.0	NaN	
...	
114562	2017-06-20	14	19.3	33.4	0.0	6.0	
114563	2017-06-21	14	21.2	32.6	0.0	7.6	
114564	2017-06-22	14	20.7	32.8	0.0	5.6	
114565	2017-06-23	14	19.5	31.8	0.0	6.2	
114566	2017-06-24	14	20.2	31.7	0.0	5.6	

	Electricity	Parameter1_Dir	Parameter1_Speed	Parameter2_9am	...	\
0	NaN	NNE	39.0	NW	...	
1	NaN	W	56.0	S	...	
2	NaN	W	30.0	SSW	...	

3	NaN	SW	30.0	ENE	...
4	NaN	W	46.0	E	...
...
114562	11.0	ENE	35.0	SE	...
114563	8.6	E	37.0	SE	...
114564	11.0	E	33.0	E	...
114565	10.6	ESE	26.0	SE	...
114566	10.7	ENE	30.0	ENE	...

	Parameter5_9am	Parameter5_3pm	Parameter6_9am	Parameter6_3pm	\
0	1012.2	1008.5	NaN	1.0	
1	1007.8	1006.2	5.0	NaN	
2	1014.4	1012.5	NaN	NaN	
3	1017.5	1013.6	NaN	NaN	
4	1014.9	1011.5	NaN	NaN	
...	
114562	1013.9	1010.5	0.0	1.0	
114563	1014.6	1011.2	7.0	0.0	
114564	1015.3	1011.8	0.0	0.0	
114565	1014.9	1010.7	1.0	1.0	
114566	1013.9	1009.7	6.0	5.0	

	Parameter7_9am	Parameter7_3pm	Failure_today	Fecha	Año	Mes
0	23.4	30.9	1	2010-01-01	2010	1
1	20.6	28.0	1	2010-01-02	2010	1
2	18.0	28.0	0	2010-01-03	2010	1
3	21.3	30.6	0	2010-01-04	2010	1
4	23.3	34.9	0	2010-01-05	2010	1
...
114562	24.5	32.3	0	2017-06-20	2017	6
114563	24.8	32.0	0	2017-06-21	2017	6
114564	24.8	32.1	0	2017-06-22	2017	6
114565	24.8	29.2	0	2017-06-23	2017	6
114566	25.4	31.0	0	2017-06-24	2017	6

[114567 rows x 25 columns]

```
[90]: #Eliminamos columnas que no vamos a ocupar

df = df.drop(columns=["Electricity", #Muchos NaN

                    "Evaporation", #Muchos NaN
                    "Leakage",
                    "Parameter6_9am",
                    "Parameter6_3pm",
                    "Parameter1_Dir"
                    #"Parameter2_9am",
```

```

        #"Parameter2_3pm"
    ]
)

#Eliminamos datos faltantes
df.dropna(inplace=True)
#df["Parameter1_Dir"] = df["Parameter1_Dir"].apply(cardinales)
df["Parameter2_9am"] = df["Parameter2_9am"].apply(cardinales) #Aplicamos la
    ↪funcion para direcciones cardinales
df["Parameter2_3pm"] = df["Parameter2_3pm"].apply(cardinales)
df = df.reset_index(drop=True)

```

```
[91]: df #Dataframe limpio
```

```

[91]:
      Date  Location  Min_Temp  Max_Temp  Parameter1_Speed  \
0   2010-01-01         3      19.4      31.9             39.0
1   2010-01-02         3      18.6      29.1             56.0
2   2010-01-03         3      12.2      29.7             30.0
3   2010-01-04         3      14.8      32.8             30.0
4   2010-01-05         3      15.0      35.8             46.0
...      ...      ...      ...      ...      ...
91384  2017-06-20        14      19.3      33.4             35.0
91385  2017-06-21        14      21.2      32.6             37.0
91386  2017-06-22        14      20.7      32.8             33.0
91387  2017-06-23        14      19.5      31.8             26.0
91388  2017-06-24        14      20.2      31.7             30.0

      Parameter2_9am  Parameter2_3pm  Parameter3_9am  Parameter3_3pm  \
0                  N                W              9.0              9.0
1                  S                W              6.0             28.0
2                  S                S              9.0             19.0
3                  E                N             11.0              9.0
4                  E                N              4.0             17.0
...      ...      ...      ...      ...
91384              S                N              9.0             20.0
91385              S                S             13.0             11.0
91386              E                W             17.0             11.0
91387              S                N              9.0             17.0
91388              E                N             15.0              7.0

      Parameter4_9am  Parameter4_3pm  Parameter5_9am  Parameter5_3pm  \
0                70.0             40.0          1012.2          1008.5
1                88.0             48.0          1007.8          1006.2
2                57.0             32.0          1014.4          1012.5
3                55.0             24.0          1017.5          1013.6
4                46.0             13.0          1014.9          1011.5
...      ...      ...      ...      ...

```

91384	63.0	32.0	1013.9	1010.5
91385	56.0	28.0	1014.6	1011.2
91386	46.0	23.0	1015.3	1011.8
91387	62.0	58.0	1014.9	1010.7
91388	73.0	32.0	1013.9	1009.7

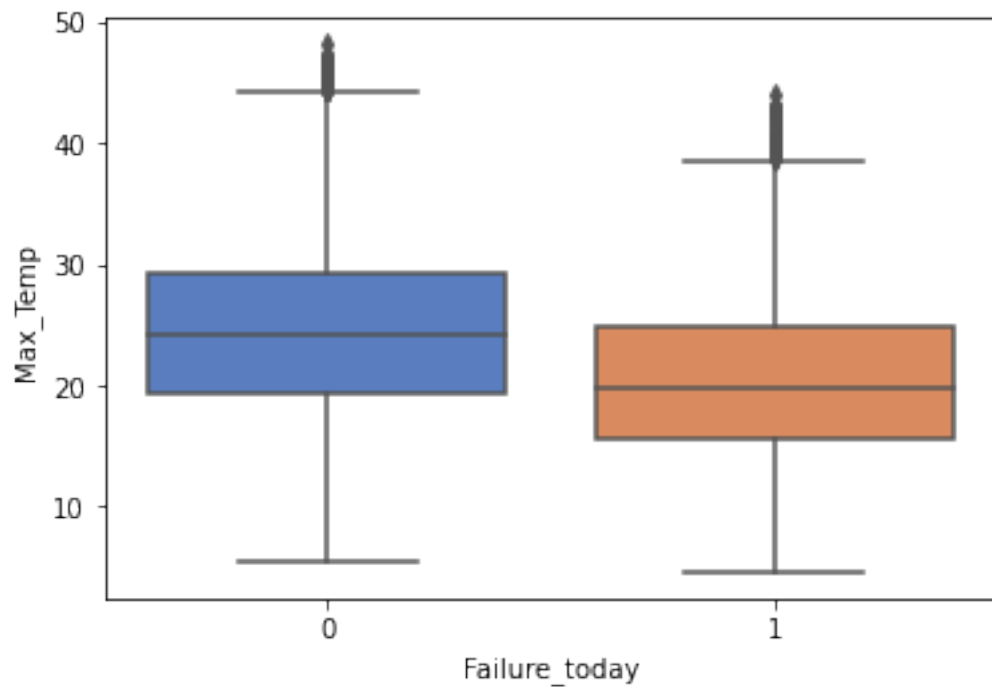
	Parameter7_9am	Parameter7_3pm	Failure_today	Fecha	Año	Mes
0	23.4	30.9	1	2010-01-01	2010	1
1	20.6	28.0	1	2010-01-02	2010	1
2	18.0	28.0	0	2010-01-03	2010	1
3	21.3	30.6	0	2010-01-04	2010	1
4	23.3	34.9	0	2010-01-05	2010	1
...
91384	24.5	32.3	0	2017-06-20	2017	6
91385	24.8	32.0	0	2017-06-21	2017	6
91386	24.8	32.1	0	2017-06-22	2017	6
91387	24.8	29.2	0	2017-06-23	2017	6
91388	25.4	31.0	0	2017-06-24	2017	6

[91389 rows x 19 columns]

Gráficos

```
[107]: sns.boxplot(df,x="Failure_today",y="Max_Temp")
```

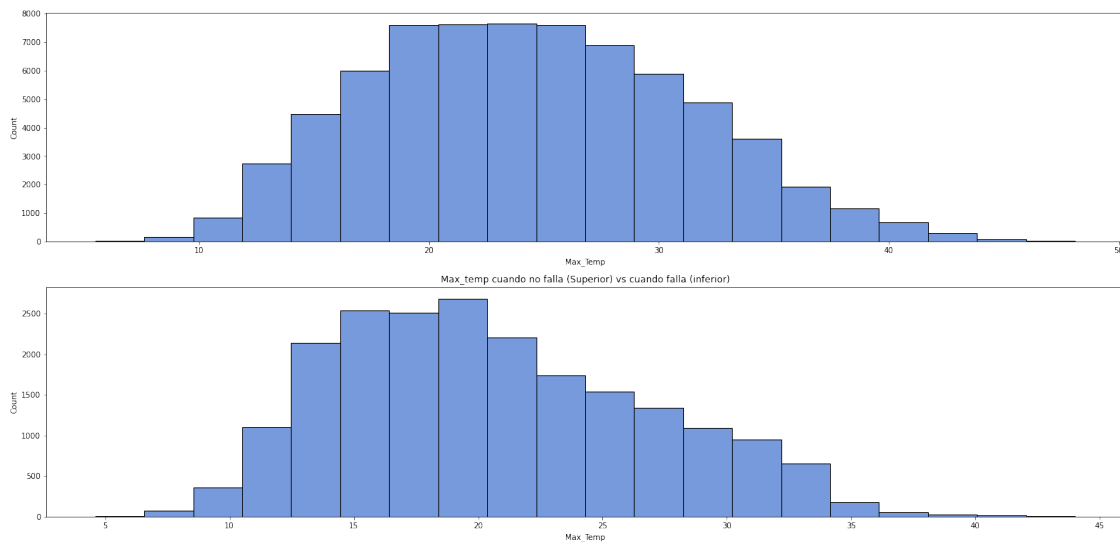
```
[107]: <Axes: xlabel='Failure_today', ylabel='Max_Temp'>
```



podemos ver que, cuando la maquina falla, su max_temp está centrada en una menor que cuando no falla

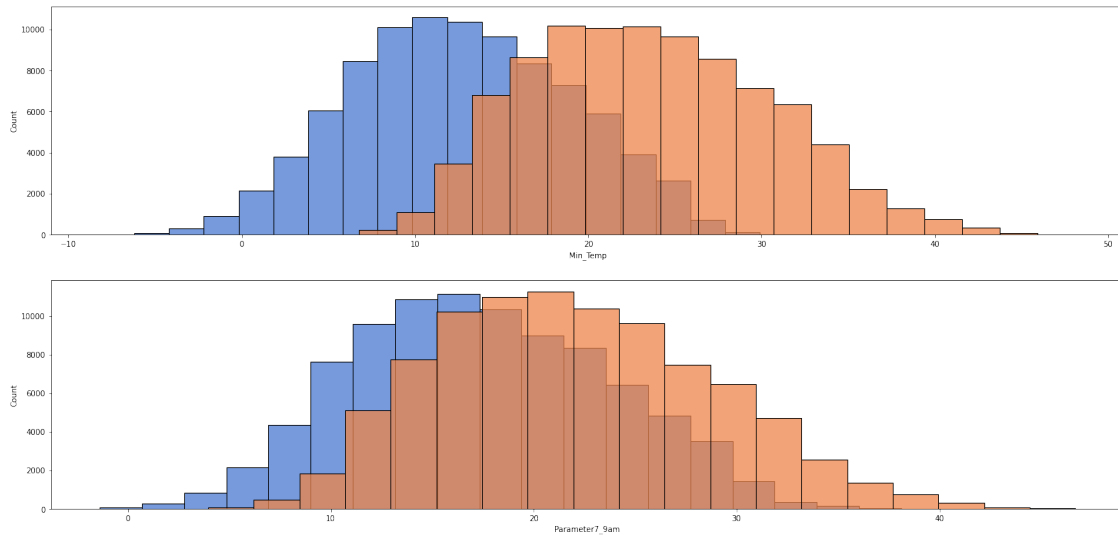
```
[130]: fig, axes = plt.subplots(2, 1, figsize=(25, 12))
plt.title("Max_temp cuando no falla (Superior) vs cuando falla (inferior)")
sns.histplot(data=df[df["Failure_today"] == 0],x="Max_Temp",bins=20,ax=axes[0])
sns.histplot(data=df[df["Failure_today"] == 1],x="Max_Temp",bins=20,ax=axes[1])
```

```
[130]: <Axes: title={'center': 'Max_temp cuando no falla (Superior) vs cuando falla (inferior)'}, xlabel='Max_Temp', ylabel='Count'>
```



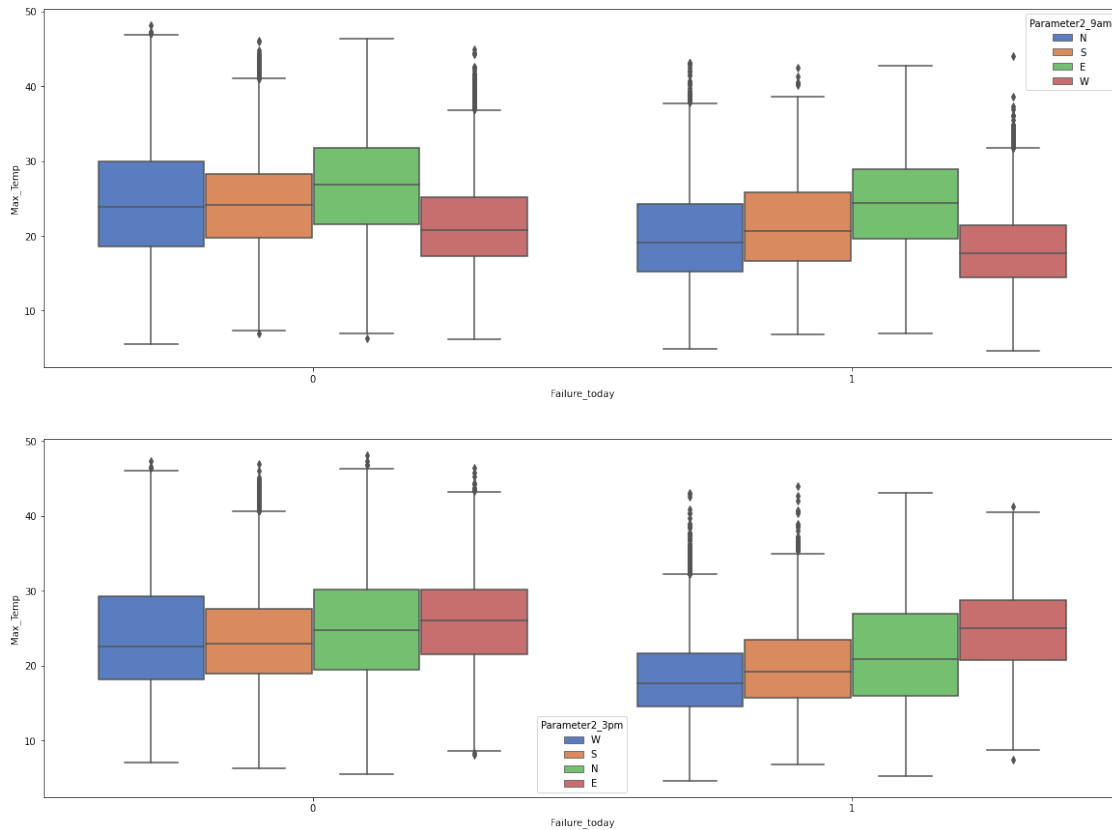
Este histograma nos entrega información diferenciada sobre la temperatura máxima en los casos que el sensor detectó un fallo y en los que no. Si bien se distribuyen de forma similar, podemos ver una leve diferencia en la forma, lo cual corrobora lo visto en el boxplot

```
[93]: fig, axes = plt.subplots(2, 1, figsize=(25, 12))
sns.histplot(data=df,x="Min_Temp",bins=20,ax=axes[0])
sns.histplot(data=df,x="Max_Temp",bins=20,ax=axes[0])
sns.histplot(data=df,x="Parameter7_9am",bins=20,ax=axes[1])
sns.histplot(data=df,x="Parameter7_3pm",bins=20,ax=axes[1])
plt.show()
```



Siguiendo con las temperaturas, nos interesa ver la relación de la temperatura maxima con la minima, y podemos notar que tanto en el parametro 7 como en la variable que mide esta se aprecia un comportamiento similar entre las 2 medidas

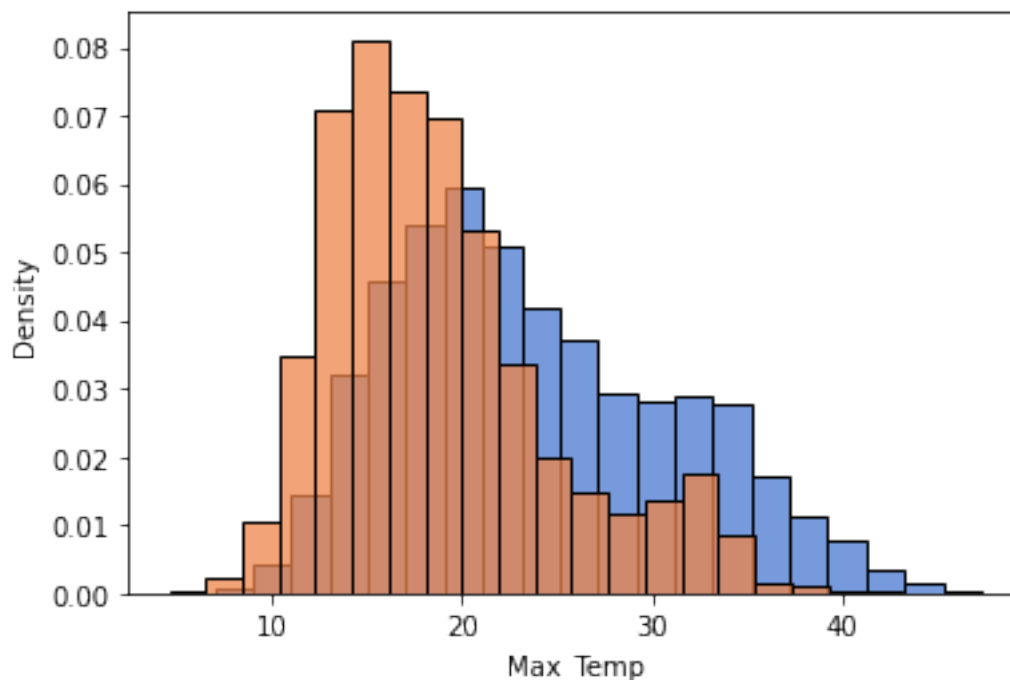
```
[94]: fig, axes = plt.subplots(2, 1, figsize=(20, 15))
sns.boxplot(data=df, y="Max_Temp",
            x="Failure_today", hue="Parameter2_9am", ax=axes[0])
sns.boxplot(data=df, y="Max_Temp",
            x="Failure_today", hue="Parameter2_3pm", ax=axes[1])
plt.show()
```



La principal diferencia la podemos encontrar a las 3pm cuando la dirección del viento corre hacia el Oeste. Para interpretar de mejor manera el fenómeno, se hace un histograma

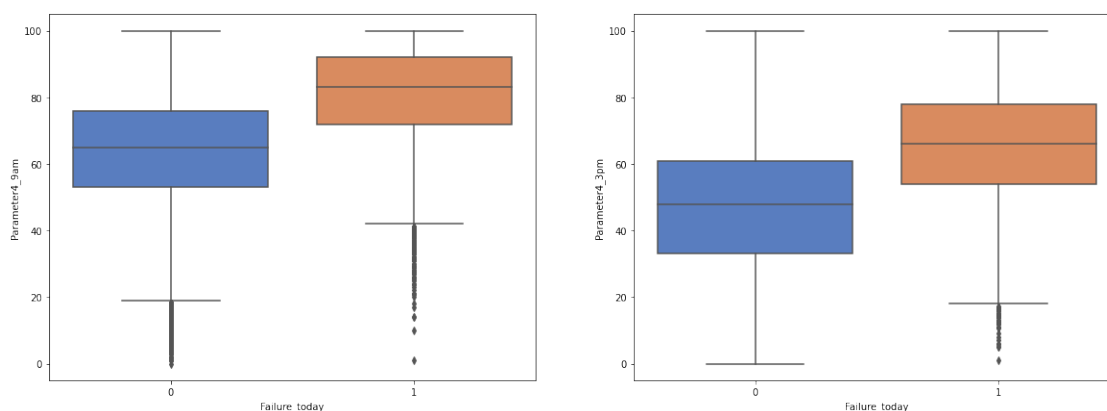
```
[95]: sns.histplot(data=df[(df["Parameter2_3pm"] == "W") & (df["Failure_today"] == 0)], x="Max_Temp", stat="density", bins=20)
sns.histplot(data=df[(df["Parameter2_3pm"] == "W") & (df["Failure_today"] == 1)], x="Max_Temp", stat="density", bins=20)
```

```
[95]: <Axes: xlabel='Max_Temp', ylabel='Density'>
```

Podemos notar que en proporcion, cuando el viento corre al oeste y la temperatura maxima es mas baja cuando falla. Este comportamiento que identificamos en el boxplot se ve de forma mas clara en el histplot. Si bien, ya sabiamos el hecho de la temperatura, la presencia del viento acrecentua este efecto al parecer.

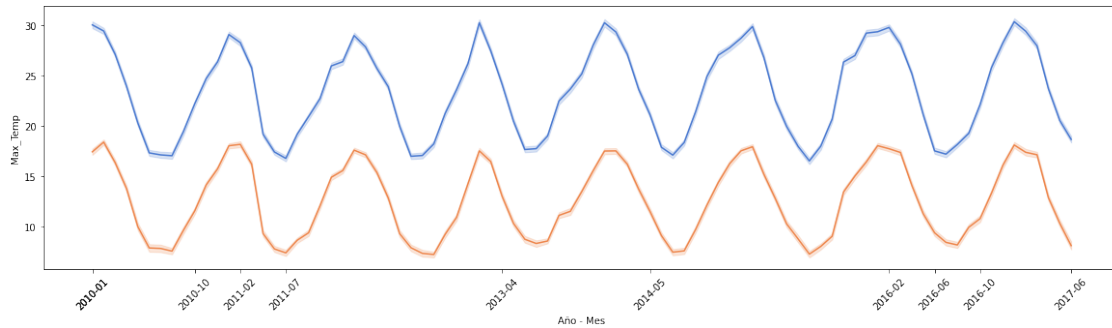
```
[96]: fig, axes = plt.subplots(1, 2, figsize=(20, 7))
sns.boxplot(df,x="Failure_today",y="Parameter4_9am",ax=axes[0])
sns.boxplot(df,x="Failure_today",y="Parameter4_3pm",ax=axes[1])
plt.show()
```



Podemos ver que claramente, en ambos horarios, hay una diferencia en el valor centrar y los

cuartiles cuando falla y cuando no el sensor. Además, durante la mañana el parametro revela un valor levemente mayor que la tarde

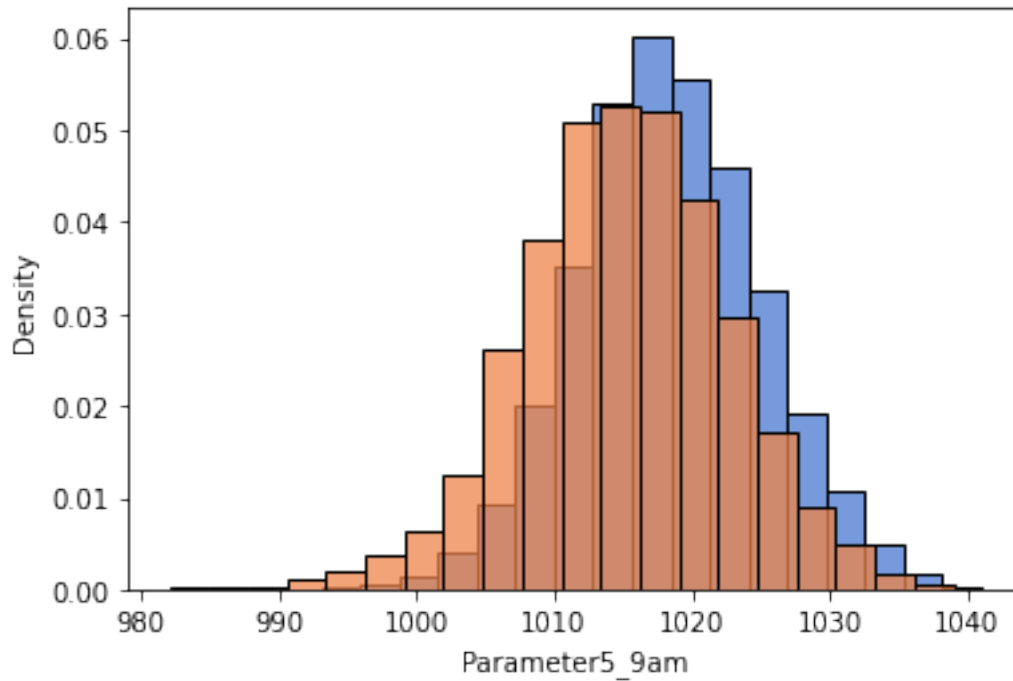
```
[132]: plt.figure(figsize=(20,5))
sns.lineplot(data=df, x="Año - Mes", y="Max_Temp"); sns.lineplot(data=df,
↪x="Año - Mes", y="Min_Temp")
plt.xticks(df["Año - Mes"][:len(df)//10], rotation=45)
plt.show()
```



Podemos notar una clara estacionalidad en la temperatura maxima y minima

```
[99]: sns.
↪histplot(df[df["Failure_today"]==0],x="Parameter5_9am",stat="density",bins=20)
sns.
↪histplot(df[df["Failure_today"]==1],x="Parameter5_9am",stat="density",bins=20)
```

```
[99]: <Axes: xlabel='Parameter5_9am', ylabel='Density'>
```

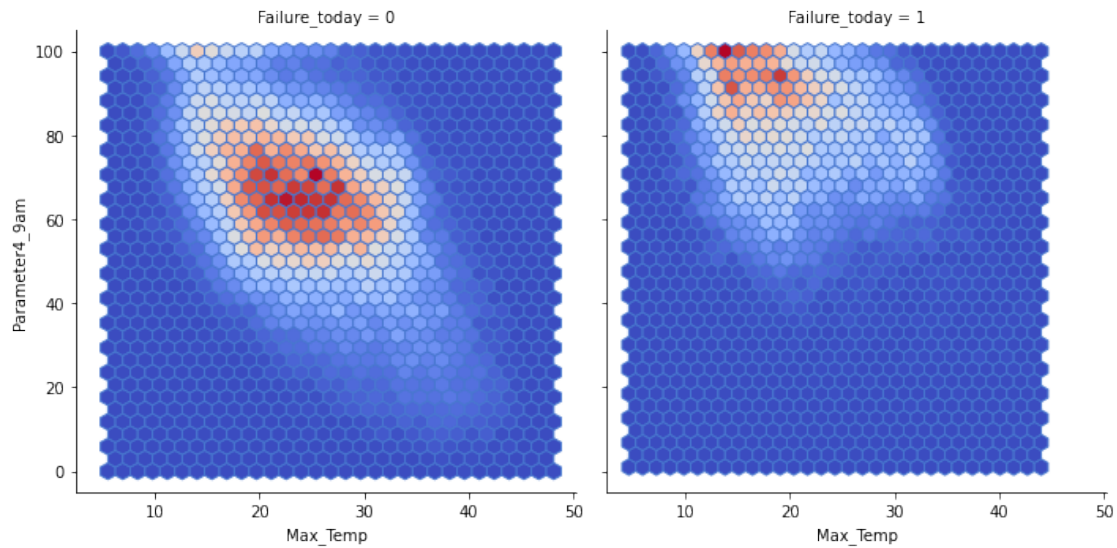


Vemos que hay un leve desplazamiento en el parametro 5 cuando falla el sensor

```
[100]: g = sns.FacetGrid(df, col="Failure_today", height=5)
g.map_dataframe(plt.hexbin, x="Max_Temp", y="Parameter4_9am", gridsize=30,
cmap="coolwarm")
```

C:\Users\joaqu\AppData\Roaming\Python\Python39\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
self._figure.tight_layout(*args, **kwargs)

```
[100]: <seaborn.axisgrid.FacetGrid at 0x21825e85a30>
```



Este grafico que nos muestra la dependencia de 2 variables y su concentración de datos nos evidencia que, si bien no parece haber una relación entre max temp y el parametro 4, vemos que cuando falla estas 2 variables la concentración de la interacción se desplazan bastante en comparación a cuando no, teniendo una temperatura maxima menor pero un valor de parametro 4 mayor.

[101]: df

```
[101]:
```

	Date	Location	Min_Temp	Max_Temp	Parameter1_Speed	\
0	2010-01-01	3	19.4	31.9	39.0	
1	2010-01-02	3	18.6	29.1	56.0	
2	2010-01-03	3	12.2	29.7	30.0	
3	2010-01-04	3	14.8	32.8	30.0	
4	2010-01-05	3	15.0	35.8	46.0	
...	
91384	2017-06-20	14	19.3	33.4	35.0	
91385	2017-06-21	14	21.2	32.6	37.0	
91386	2017-06-22	14	20.7	32.8	33.0	
91387	2017-06-23	14	19.5	31.8	26.0	
91388	2017-06-24	14	20.2	31.7	30.0	

	Parameter2_9am	Parameter2_3pm	Parameter3_9am	Parameter3_3pm	\
0	N	W	9.0	9.0	
1	S	W	6.0	28.0	
2	S	S	9.0	19.0	
3	E	N	11.0	9.0	
4	E	N	4.0	17.0	
...	
91384	S	N	9.0	20.0	
91385	S	S	13.0	11.0	

91386	E	W	17.0	11.0
91387	S	N	9.0	17.0
91388	E	N	15.0	7.0

	Parameter4_9am	Parameter4_3pm	Parameter5_9am	Parameter5_3pm	\
0	70.0	40.0	1012.2	1008.5	
1	88.0	48.0	1007.8	1006.2	
2	57.0	32.0	1014.4	1012.5	
3	55.0	24.0	1017.5	1013.6	
4	46.0	13.0	1014.9	1011.5	
...	
91384	63.0	32.0	1013.9	1010.5	
91385	56.0	28.0	1014.6	1011.2	
91386	46.0	23.0	1015.3	1011.8	
91387	62.0	58.0	1014.9	1010.7	
91388	73.0	32.0	1013.9	1009.7	

	Parameter7_9am	Parameter7_3pm	Failure_today	Fecha	Año	Mes	\
0	23.4	30.9	1	2010-01-01	2010	1	
1	20.6	28.0	1	2010-01-02	2010	1	
2	18.0	28.0	0	2010-01-03	2010	1	
3	21.3	30.6	0	2010-01-04	2010	1	
4	23.3	34.9	0	2010-01-05	2010	1	
...	
91384	24.5	32.3	0	2017-06-20	2017	6	
91385	24.8	32.0	0	2017-06-21	2017	6	
91386	24.8	32.1	0	2017-06-22	2017	6	
91387	24.8	29.2	0	2017-06-23	2017	6	
91388	25.4	31.0	0	2017-06-24	2017	6	

	Año - Mes
0	2010-01
1	2010-01
2	2010-01
3	2010-01
4	2010-01
...	...
91384	2017-06
91385	2017-06
91386	2017-06
91387	2017-06
91388	2017-06

[91389 rows x 20 columns]

1.0.4 Parte 2

OLS

```
[103]: y = df["Failure_today"]
X = df.copy()

X2 = X.copy()
X = pd.get_dummies(X, columns=["Location"], prefix=["Location"],
    ↪prefix_sep='__', drop_first=True)
X = X.drop(columns=["Failure_today",
    "Date",
    "Año",
    #"Mes",
    "Año - Mes",
    "Fecha"],axis=1)
X = pd.get_dummies(X, columns=["Parameter2_9am",
    "Parameter2_3pm",
    "Mes",
    #"Trimestre"
    ], prefix_sep='__', drop_first=True)

X=sm.add_constant(X)
model = sm.OLS(y, X)
results = model.fit(cov_type='HCO')
print(results.summary())

#fail - c(location) - 1
```

OLS Regression Results

```
=====
Dep. Variable:          Failure_today    R-squared:                0.303
Model:                  OLS             Adj. R-squared:          0.302
Method:                 Least Squares    F-statistic:             632.4
Date:                   jue, 24 abr. 2025 Prob (F-statistic):       0.00
Time:                   20:42:54         Log-Likelihood:          -34482.
No. Observations:       91389           AIC:                    6.910e+04
Df Residuals:           91322           BIC:                    6.973e+04
Df Model:                66
Covariance Type:        HCO
=====
=====
```

	coef	std err	z	P> z	[0.025
0.975]					
const	8.1858	0.255	32.150	0.000	7.687
Min_Temp	0.0112	0.001	19.495	0.000	0.010
Max_Temp	-0.0340	0.001	-29.637	0.000	-0.036

```
-----
-----
```

Parameter1_Speed 0.006	0.0055	0.000	34.996	0.000	0.005
Parameter3_9am 0.003	0.0027	0.000	12.944	0.000	0.002
Parameter3_3pm -0.004	-0.0042	0.000	-19.601	0.000	-0.005
Parameter4_9am 0.008	0.0075	0.000	55.859	0.000	0.007
Parameter4_3pm 0.003	0.0024	0.000	15.119	0.000	0.002
Parameter5_9am -0.038	-0.0401	0.001	-46.347	0.000	-0.042
Parameter5_3pm 0.033	0.0317	0.001	36.350	0.000	0.030
Parameter7_9am -0.001	-0.0030	0.001	-3.217	0.001	-0.005
Parameter7_3pm 0.030	0.0279	0.001	21.921	0.000	0.025
Location__3 -0.060	-0.0811	0.011	-7.689	0.000	-0.102
Location__4 0.125	0.1070	0.009	11.403	0.000	0.089
Location__5 -0.077	-0.0990	0.011	-9.030	0.000	-0.120
Location__6 -0.215	-0.2364	0.011	-21.411	0.000	-0.258
Location__7 -0.105	-0.1249	0.010	-12.422	0.000	-0.145
Location__8 0.024	0.0024	0.011	0.218	0.828	-0.019
Location__9 -0.037	-0.0604	0.012	-4.977	0.000	-0.084
Location__10 -0.070	-0.0909	0.011	-8.467	0.000	-0.112
Location__11 -0.012	-0.0305	0.009	-3.248	0.001	-0.049
Location__12 -0.018	-0.0405	0.012	-3.515	0.000	-0.063
Location__13 -0.148	-0.1704	0.012	-14.730	0.000	-0.193
Location__14 -0.076	-0.0989	0.012	-8.480	0.000	-0.122
Location__15 -0.050	-0.0719	0.011	-6.463	0.000	-0.094
Location__16 -0.134	-0.1550	0.011	-14.338	0.000	-0.176
Location__18 -0.126	-0.1516	0.013	-11.658	0.000	-0.177

Location__20	-0.1727	0.011	-16.381	0.000	-0.193
-0.152					
Location__21	-0.1057	0.009	-11.188	0.000	-0.124
-0.087					
Location__22	-0.0350	0.010	-3.490	0.000	-0.055
-0.015					
Location__23	-0.1038	0.011	-9.515	0.000	-0.125
-0.082					
Location__27	-0.1477	0.011	-13.266	0.000	-0.170
-0.126					
Location__28	-0.1500	0.011	-13.145	0.000	-0.172
-0.128					
Location__29	-0.0891	0.010	-8.919	0.000	-0.109
-0.070					
Location__30	-0.0118	0.010	-1.128	0.259	-0.032
0.009					
Location__32	-0.0424	0.010	-4.361	0.000	-0.061
-0.023					
Location__33	-0.0409	0.010	-4.180	0.000	-0.060
-0.022					
Location__34	-0.1339	0.011	-11.753	0.000	-0.156
-0.112					
Location__35	-0.0862	0.011	-7.673	0.000	-0.108
-0.064					
Location__36	-0.2054	0.011	-18.573	0.000	-0.227
-0.184					
Location__38	-0.1007	0.011	-8.934	0.000	-0.123
-0.079					
Location__39	-0.0814	0.011	-7.561	0.000	-0.102
-0.060					
Location__40	-0.1199	0.011	-10.854	0.000	-0.142
-0.098					
Location__41	-0.0684	0.011	-6.317	0.000	-0.090
-0.047					
Location__43	-0.0679	0.010	-6.732	0.000	-0.088
-0.048					
Location__44	-0.1086	0.011	-9.478	0.000	-0.131
-0.086					
Location__45	-0.1610	0.011	-15.168	0.000	-0.182
-0.140					
Location__47	-0.0594	0.011	-5.213	0.000	-0.082
-0.037					
Location__48	-0.1766	0.011	-16.101	0.000	-0.198
-0.155					
Location__49	-0.0908	0.009	-9.750	0.000	-0.109
-0.073					
Parameter2_9am__N	-0.0014	0.004	-0.386	0.699	-0.008
0.006					

Parameter2_9am__S 0.031	0.0238	0.004	6.753	0.000	0.017
Parameter2_9am__W 0.041	0.0317	0.005	6.842	0.000	0.023
Parameter2_3pm__N 0.010	0.0024	0.004	0.653	0.514	-0.005
Parameter2_3pm__S 0.028	0.0208	0.004	5.638	0.000	0.014
Parameter2_3pm__W 0.038	0.0293	0.004	6.772	0.000	0.021
Mes__10 0.076	0.0639	0.006	10.548	0.000	0.052
Mes__11 0.063	0.0522	0.006	9.182	0.000	0.041
Mes__12 0.041	0.0301	0.006	5.234	0.000	0.019
Mes__2 -0.007	-0.0184	0.006	-3.307	0.001	-0.029
Mes__3 0.007	-0.0035	0.006	-0.641	0.522	-0.014
Mes__4 0.049	0.0372	0.006	6.182	0.000	0.025
Mes__5 0.031	0.0176	0.007	2.671	0.008	0.005
Mes__6 0.006	-0.0085	0.007	-1.155	0.248	-0.023
Mes__7 0.038	0.0230	0.008	2.976	0.003	0.008
Mes__8 0.074	0.0595	0.007	8.227	0.000	0.045
Mes__9 0.086	0.0734	0.007	11.040	0.000	0.060

```

=====
Omnibus:                6534.458    Durbin-Watson:                1.798
Prob(Omnibus):           0.000    Jarque-Bera (JB):            7967.918
Skew:                    0.718    Prob(JB):                     0.00
Kurtosis:                2.824    Cond. No.                    2.99e+05
=====

```

Notes:

[1] Standard Errors are heteroscedasticity robust (HCO)

[2] The condition number is large, 2.99e+05. This might indicate that there are strong multicollinearity or other numerical problems.

Se puede notar que la mayoría de variables seleccionadas son significativas.

Sobre los sensores, se evidencia que el sensor 4 tiene en promedio un 10% mas de probabilidades de fallar menos que el base, mientras que por el contrario el 6 tiene un 20% menos. Los sensores no significativos como el 30 nos estan indicando que tienen un comportamiento similar al base.

Con estos resultados a simple vista podemos concluir que la ubicacion del sensor afecta en su probabilidad de fallo

Por el lado de los parametros, vemos que el parametro 5 se comporta de formas muy opuestas entre mañana y tarde, aumentando la probabilidad de fallo si aumenta su valor durante la tarde, pero disminuyendo la probabilidad casi en la misma proporcion si el aumento se da en la mañana. Por otro lado, el parametro 4 tiene un comportamiento de aumentar la probabilidad de fallo en ambos periodos, pero mucho mas marcado en las mañanas. Para la temperatura, podemos notar que si aumenta en 1 unidad la temperatura maxima, disminuye en 3% la probabilidad de falla, uno de los impactos mas grandes, lo cual tiene sentido por lo visto en los graficos, que en promedio los sensores cuando fallan tienen una temperatura maxima menor al caso cuando no fallan. El parametro 7 tambien tiene comportamiento opuesto entre mañana y tarde, pero este es atribuible a la temperatura, ya que verificamos con los graficos que representan eso, y con max temp y min temp vimos que tienen efectos opuestos. La diferencia es que, durante la tarde, se comporta muy similar a max_temp, pero durante la mañana si bien tiene el mismo efecto que min_temp de aumentar la probabilidad, la magnitud es mucho menor.

Finalmente viendo los efectos por mes, tenemos casos como Febrero, que en comparación con Enero es menos propenso a fallar, por el contrario septiembre aumenta la probabilidad de falla. Esto nos da indicios de estacionalidad en la probabilidad que los sensores detecten una falla.

El caso de la temperatura es interesante, ya que se podria interpretar como si la temperatura tenga relación a la operación de esta maquina, osea que si la temperatura maxima es alta significa que está operando. Es por esto que cuando la temperatura maxima aumenta la probabilidad de fallo baja.

1.0.5 Parte 3

```
[104]: X = X.drop(columns=["Electricity",
                        "Evaporation",
                        "Electricity_NaN",
                        "Evaporation_NaN"
                        ],axis=1)

model = sm.Probit(y, X)
probit_model = model.fit(cov_type='HCO')
print(probit_model.summary())

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

Optimization terminated successfully.

Current function value: 0.362559

Iterations 7

Probit Regression Results

```
=====
Dep. Variable:          Failure_today    No. Observations:          91389
Model:                  Probit          Df Residuals:          91322
Method:                  MLE           Df Model:              66
Date:                   jue, 24 abr. 2025    Pseudo R-squ.:          0.3317
```

Time: 20:54:09 Log-Likelihood: -33134.
 converged: True LL-Null: -49582.
 Covariance Type: HCO LLR p-value: 0.000

=====

	coef	std err	z	P> z	[0.025
0.975]					

const	28.7943	1.122	25.653	0.000	26.594
30.994					
Min_Temp	0.0770	0.003	23.695	0.000	0.071
0.083					
Max_Temp	-0.1467	0.006	-24.854	0.000	-0.158
-0.135					
Parameter1_Speed	0.0207	0.001	29.358	0.000	0.019
0.022					
Parameter3_9am	0.0086	0.001	8.765	0.000	0.007
0.011					
Parameter3_3pm	-0.0149	0.001	-14.938	0.000	-0.017
-0.013					
Parameter4_9am	0.0392	0.001	55.763	0.000	0.038
0.041					
Parameter4_3pm	0.0031	0.001	4.452	0.000	0.002
0.004					
Parameter5_9am	-0.1416	0.004	-35.427	0.000	-0.149
-0.134					
Parameter5_3pm	0.1106	0.004	27.933	0.000	0.103
0.118					
Parameter7_9am	-0.0205	0.005	-3.957	0.000	-0.031
-0.010					
Parameter7_3pm	0.0693	0.007	10.628	0.000	0.056
0.082					
Location__3	-0.3277	0.054	-6.053	0.000	-0.434
-0.222					
Location__4	0.2902	0.067	4.353	0.000	0.160
0.421					
Location__5	-0.2577	0.053	-4.897	0.000	-0.361
-0.155					
Location__6	-1.1510	0.054	-21.320	0.000	-1.257
-1.045					
Location__7	-0.5737	0.053	-10.870	0.000	-0.677
-0.470					
Location__8	0.3467	0.052	6.706	0.000	0.245
0.448					
Location__9	0.1011	0.055	1.848	0.065	-0.006
0.208					
Location__10	-0.2989	0.054	-5.570	0.000	-0.404

-0.194					
Location__11	-0.1921	0.058	-3.303	0.001	-0.306
-0.078					
Location__12	0.0552	0.052	1.068	0.286	-0.046
0.157					
Location__13	-0.7918	0.052	-15.128	0.000	-0.894
-0.689					
Location__14	-0.0126	0.057	-0.222	0.824	-0.124
0.099					
Location__15	-0.0174	0.053	-0.328	0.743	-0.122
0.087					
Location__16	-0.5619	0.052	-10.738	0.000	-0.664
-0.459					
Location__18	-0.5917	0.059	-10.086	0.000	-0.707
-0.477					
Location__20	-0.6839	0.051	-13.369	0.000	-0.784
-0.584					
Location__21	-0.6854	0.057	-12.006	0.000	-0.797
-0.574					
Location__22	0.0865	0.056	1.536	0.124	-0.024
0.197					
Location__23	-0.4711	0.050	-9.377	0.000	-0.570
-0.373					
Location__27	-0.4939	0.051	-9.778	0.000	-0.593
-0.395					
Location__28	-0.4485	0.050	-9.035	0.000	-0.546
-0.351					
Location__29	-0.5793	0.055	-10.507	0.000	-0.687
-0.471					
Location__30	0.1328	0.056	2.391	0.017	0.024
0.242					
Location__32	-0.0842	0.051	-1.639	0.101	-0.185
0.017					
Location__33	-0.0369	0.053	-0.702	0.483	-0.140
0.066					
Location__34	-0.5996	0.049	-12.124	0.000	-0.697
-0.503					
Location__35	-0.2056	0.054	-3.812	0.000	-0.311
-0.100					
Location__36	-0.7599	0.053	-14.433	0.000	-0.863
-0.657					
Location__38	-0.2390	0.051	-4.696	0.000	-0.339
-0.139					
Location__39	-0.1939	0.052	-3.722	0.000	-0.296
-0.092					
Location__40	-0.1336	0.057	-2.355	0.019	-0.245
-0.022					
Location__41	-0.2038	0.053	-3.845	0.000	-0.308

-0.100					
Location__43	-0.2616	0.055	-4.734	0.000	-0.370
-0.153					
Location__44	-0.3981	0.051	-7.829	0.000	-0.498
-0.298					
Location__45	-0.6883	0.051	-13.526	0.000	-0.788
-0.589					
Location__47	-0.1548	0.053	-2.941	0.003	-0.258
-0.052					
Location__48	-0.6166	0.052	-11.923	0.000	-0.718
-0.515					
Location__49	-0.7198	0.066	-10.986	0.000	-0.848
-0.591					
Parameter2_9am__N	-0.0039	0.020	-0.194	0.846	-0.044
0.036					
Parameter2_9am__S	0.1531	0.019	8.092	0.000	0.116
0.190					
Parameter2_9am__W	0.1581	0.022	7.223	0.000	0.115
0.201					
Parameter2_3pm__N	-0.0102	0.020	-0.514	0.607	-0.049
0.029					
Parameter2_3pm__S	0.0429	0.018	2.336	0.019	0.007
0.079					
Parameter2_3pm__W	0.0801	0.021	3.725	0.000	0.038
0.122					
Mes__10	0.1102	0.032	3.426	0.001	0.047
0.173					
Mes__11	0.1697	0.030	5.701	0.000	0.111
0.228					
Mes__12	0.1223	0.030	4.038	0.000	0.063
0.182					
Mes__2	-0.0592	0.029	-2.040	0.041	-0.116
-0.002					
Mes__3	-0.0427	0.027	-1.579	0.114	-0.096
0.010					
Mes__4	0.0361	0.029	1.234	0.217	-0.021
0.094					
Mes__5	-0.1006	0.032	-3.180	0.001	-0.163
-0.039					
Mes__6	-0.2711	0.035	-7.793	0.000	-0.339
-0.203					
Mes__7	-0.1774	0.036	-4.866	0.000	-0.249
-0.106					
Mes__8	-0.0117	0.035	-0.329	0.742	-0.081
0.058					
Mes__9	0.1237	0.033	3.713	0.000	0.058
0.189					

=====

```

=====
                Probit Marginal Effects
=====
Dep. Variable:          Failure_today
Method:                dydx
At:                    overall
=====
=====
                dy/dx      std err          z      P>|z|      [0.025
0.975]
-----
-----
Min_Temp              0.0157      0.001      23.945      0.000      0.014
0.017
Max_Temp             -0.0299      0.001     -25.169      0.000     -0.032
-0.028
Parameter1_Speed      0.0042      0.000      29.902      0.000      0.004
0.004
Parameter3_9am         0.0018      0.000       8.776      0.000      0.001
0.002
Parameter3_3pm        -0.0030      0.000     -15.001      0.000     -0.003
-0.003
Parameter4_9am         0.0080      0.000      60.363      0.000      0.008
0.008
Parameter4_3pm         0.0006      0.000       4.451      0.000      0.000
0.001
Parameter5_9am        -0.0288      0.001     -36.245      0.000     -0.030
-0.027
Parameter5_3pm         0.0225      0.001      28.339      0.000      0.021
0.024
Parameter7_9am        -0.0042      0.001      -3.961      0.000     -0.006
-0.002
Parameter7_3pm         0.0141      0.001      10.646      0.000      0.012
0.017
Location__3           -0.0667      0.011      -6.058      0.000     -0.088
-0.045
Location__4            0.0591      0.014       4.357      0.000      0.033
0.086
Location__5           -0.0525      0.011      -4.896      0.000     -0.074
-0.031
Location__6           -0.2344      0.011     -21.576      0.000     -0.256
-0.213
Location__7           -0.1169      0.011     -10.892      0.000     -0.138
-0.096
Location__8            0.0706      0.011       6.723      0.000      0.050
0.091
Location__9            0.0206      0.011       1.849      0.064     -0.001
0.042

```

Location__10 -0.039	-0.0609	0.011	-5.573	0.000	-0.082
Location__11 -0.016	-0.0391	0.012	-3.304	0.001	-0.062
Location__12 0.032	0.0112	0.011	1.068	0.286	-0.009
Location__13 -0.140	-0.1613	0.011	-15.192	0.000	-0.182
Location__14 0.020	-0.0026	0.012	-0.222	0.824	-0.025
Location__15 0.018	-0.0036	0.011	-0.328	0.743	-0.025
Location__16 -0.094	-0.1144	0.011	-10.772	0.000	-0.135
Location__18 -0.097	-0.1205	0.012	-10.103	0.000	-0.144
Location__20 -0.119	-0.1393	0.010	-13.407	0.000	-0.160
Location__21 -0.117	-0.1396	0.012	-12.035	0.000	-0.162
Location__22 0.040	0.0176	0.011	1.537	0.124	-0.005
Location__23 -0.076	-0.0960	0.010	-9.390	0.000	-0.116
Location__27 -0.080	-0.1006	0.010	-9.785	0.000	-0.121
Location__28 -0.072	-0.0914	0.010	-9.037	0.000	-0.111
Location__29 -0.096	-0.1180	0.011	-10.535	0.000	-0.140
Location__30 0.049	0.0270	0.011	2.392	0.017	0.005
Location__32 0.003	-0.0172	0.010	-1.638	0.101	-0.038
Location__33 0.013	-0.0075	0.011	-0.702	0.483	-0.029
Location__34 -0.102	-0.1221	0.010	-12.154	0.000	-0.142
Location__35 -0.020	-0.0419	0.011	-3.811	0.000	-0.063
Location__36 -0.134	-0.1548	0.011	-14.499	0.000	-0.176
Location__38 -0.028	-0.0487	0.010	-4.695	0.000	-0.069
Location__39 -0.019	-0.0395	0.011	-3.721	0.000	-0.060
Location__40 -0.005	-0.0272	0.012	-2.353	0.019	-0.050

Location__41 -0.020	-0.0415	0.011	-3.845	0.000	-0.063
Location__43 -0.031	-0.0533	0.011	-4.738	0.000	-0.075
Location__44 -0.061	-0.0811	0.010	-7.835	0.000	-0.101
Location__45 -0.120	-0.1402	0.010	-13.572	0.000	-0.160
Location__47 -0.011	-0.0315	0.011	-2.941	0.003	-0.053
Location__48 -0.105	-0.1256	0.011	-11.936	0.000	-0.146
Location__49 -0.121	-0.1466	0.013	-11.026	0.000	-0.173
Parameter2_9am__N 0.007	-0.0008	0.004	-0.194	0.846	-0.009
Parameter2_9am__S 0.039	0.0312	0.004	8.095	0.000	0.024
Parameter2_9am__W 0.041	0.0322	0.004	7.223	0.000	0.023
Parameter2_3pm__N 0.006	-0.0021	0.004	-0.514	0.607	-0.010
Parameter2_3pm__S 0.016	0.0087	0.004	2.336	0.020	0.001
Parameter2_3pm__W 0.025	0.0163	0.004	3.725	0.000	0.008
Mes__10 0.035	0.0225	0.007	3.425	0.001	0.010
Mes__11 0.046	0.0346	0.006	5.706	0.000	0.023
Mes__12 0.037	0.0249	0.006	4.040	0.000	0.013
Mes__2 -0.000	-0.0121	0.006	-2.039	0.041	-0.024
Mes__3 0.002	-0.0087	0.006	-1.578	0.114	-0.020
Mes__4 0.019	0.0074	0.006	1.234	0.217	-0.004
Mes__5 -0.008	-0.0205	0.006	-3.180	0.001	-0.033
Mes__6 -0.041	-0.0552	0.007	-7.806	0.000	-0.069
Mes__7 -0.022	-0.0361	0.007	-4.870	0.000	-0.051
Mes__8 0.012	-0.0024	0.007	-0.329	0.742	-0.017
Mes__9 0.039	0.0252	0.007	3.714	0.000	0.012

=====
=====

Para el modelo probit se consiguió un mejor ajuste y resultados muy similares al OLS

Por cada aumento de 1 en la temperatura mínima, la probabilidad de fallo aumenta en 1.57%, mientras que este cambio para la temperatura maxima disminuyedicha probabilidad en 2.99%.

Al igual que en OLS, aumentar la velocidad del viento en 1 unidad aumenta la probabilidad de fallo en un 0.43%, un impacto no tan grande pero significativo. El parametro 5 mantiene su tendencia a aumentar la probabilidad de fallo en las tardes y disminuirlo en las mañanas cuando aumenta en 1 unidad. En general, todo se comporta como en OLS.

Para la dirección del viento, si durante la mañana este se encuentra hacia el sur o el oeste, la probabilidad de fallo aumenta en un 3% en relacion a la ,dirección base que es el este, mientras que durante la tarde no se observan impactos mayores. Cuando el viento corre hacia el norte no nos da significativo, tanto en mañana como en tarde, lo cual nos indica que en ambos casos este se comporta similar a la dirección usada de referencia.

1.0.6 Parte 4

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

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

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

Optimization terminated successfully.

Current function value: 0.360972

Iterations 8

Logit Regression Results

=====
=====

Dep. Variable:	Failure_today	No. Observations:	91389
Model:	Logit	Df Residuals:	91322
Method:	MLE	Df Model:	66
Date:	jue, 24 abr. 2025	Pseudo R-squ.:	0.3347
Time:	21:51:25	Log-Likelihood:	-32989.
converged:	True	LL-Null:	-49582.
Covariance Type:	HC0	LLR p-value:	0.000

=====
=====

	coef	std err	z	P> z	[0.025
--	------	---------	---	------	--------

0.975]

const	49.6439	1.987	24.979	0.000	45.749
53.539					
Min_Temp	0.1396	0.006	24.347	0.000	0.128
0.151					
Max_Temp	-0.2663	0.011	-25.227	0.000	-0.287
-0.246					
Parameter1_Speed	0.0364	0.001	29.157	0.000	0.034
0.039					
Parameter3_9am	0.0143	0.002	8.155	0.000	0.011
0.018					
Parameter3_3pm	-0.0253	0.002	-14.226	0.000	-0.029
-0.022					
Parameter4_9am	0.0717	0.001	57.664	0.000	0.069
0.074					
Parameter4_3pm	0.0042	0.001	3.470	0.001	0.002
0.007					
Parameter5_9am	-0.2498	0.007	-35.265	0.000	-0.264
-0.236					
Parameter5_3pm	0.1962	0.007	28.031	0.000	0.182
0.210					
Parameter7_9am	-0.0371	0.009	-4.063	0.000	-0.055
-0.019					
Parameter7_3pm	0.1173	0.012	10.187	0.000	0.095
0.140					
Location__3	-0.5936	0.095	-6.229	0.000	-0.780
-0.407					
Location__4	0.5294	0.118	4.471	0.000	0.297
0.761					
Location__5	-0.3818	0.093	-4.089	0.000	-0.565
-0.199					
Location__6	-2.0802	0.094	-22.144	0.000	-2.264
-1.896					
Location__7	-1.0065	0.093	-10.871	0.000	-1.188
-0.825					
Location__8	0.7514	0.091	8.302	0.000	0.574
0.929					
Location__9	0.3524	0.095	3.705	0.000	0.166
0.539					
Location__10	-0.5106	0.095	-5.354	0.000	-0.697
-0.324					
Location__11	-0.3702	0.103	-3.579	0.000	-0.573
-0.167					
Location__12	0.2039	0.091	2.250	0.024	0.026
0.382					
Location__13	-1.3905	0.092	-15.170	0.000	-1.570

-1.211					
Location__14	0.1645	0.099	1.659	0.097	-0.030
0.359					
Location__15	0.0921	0.093	0.987	0.324	-0.091
0.275					
Location__16	-1.0109	0.093	-10.822	0.000	-1.194
-0.828					
Location__18	-1.0165	0.103	-9.874	0.000	-1.218
-0.815					
Location__20	-1.1782	0.091	-13.016	0.000	-1.356
-1.001					
Location__21	-1.2082	0.101	-12.011	0.000	-1.405
-1.011					
Location__22	0.2115	0.102	2.065	0.039	0.011
0.412					
Location__23	-0.8147	0.088	-9.233	0.000	-0.988
-0.642					
Location__27	-0.8197	0.089	-9.182	0.000	-0.995
-0.645					
Location__28	-0.7189	0.087	-8.262	0.000	-0.889
-0.548					
Location__29	-1.0446	0.097	-10.753	0.000	-1.235
-0.854					
Location__30	0.2596	0.097	2.677	0.007	0.070
0.450					
Location__32	-0.0650	0.090	-0.724	0.469	-0.241
0.111					
Location__33	0.0136	0.092	0.148	0.882	-0.167
0.194					
Location__34	-1.0339	0.087	-11.853	0.000	-1.205
-0.863					
Location__35	-0.2804	0.095	-2.946	0.003	-0.467
-0.094					
Location__36	-1.3364	0.093	-14.324	0.000	-1.519
-1.154					
Location__38	-0.3424	0.090	-3.817	0.000	-0.518
-0.167					
Location__39	-0.2860	0.094	-3.058	0.002	-0.469
-0.103					
Location__40	-0.0406	0.099	-0.408	0.683	-0.235
0.154					
Location__41	-0.3333	0.094	-3.542	0.000	-0.518
-0.149					
Location__43	-0.5093	0.098	-5.195	0.000	-0.701
-0.317					
Location__44	-0.6783	0.090	-7.560	0.000	-0.854
-0.502					
Location__45	-1.2034	0.090	-13.436	0.000	-1.379

-1.028					
Location__47	-0.2316	0.092	-2.511	0.012	-0.412
-0.051					
Location__48	-1.0316	0.092	-11.189	0.000	-1.212
-0.851					
Location__49	-1.3123	0.114	-11.484	0.000	-1.536
-1.088					
Parameter2_9am__N	-0.0167	0.036	-0.467	0.641	-0.087
0.053					
Parameter2_9am__S	0.2721	0.034	8.090	0.000	0.206
0.338					
Parameter2_9am__W	0.2741	0.039	7.072	0.000	0.198
0.350					
Parameter2_3pm__N	-0.0219	0.035	-0.625	0.532	-0.091
0.047					
Parameter2_3pm__S	0.0494	0.033	1.519	0.129	-0.014
0.113					
Parameter2_3pm__W	0.1218	0.038	3.206	0.001	0.047
0.196					
Mes__10	0.1663	0.058	2.884	0.004	0.053
0.279					
Mes__11	0.3095	0.053	5.802	0.000	0.205
0.414					
Mes__12	0.2250	0.055	4.126	0.000	0.118
0.332					
Mes__2	-0.0721	0.052	-1.393	0.164	-0.173
0.029					
Mes__3	-0.0433	0.048	-0.898	0.369	-0.138
0.051					
Mes__4	0.0836	0.052	1.615	0.106	-0.018
0.185					
Mes__5	-0.1735	0.056	-3.117	0.002	-0.283
-0.064					
Mes__6	-0.4930	0.061	-8.094	0.000	-0.612
-0.374					
Mes__7	-0.3309	0.064	-5.179	0.000	-0.456
-0.206					
Mes__8	-0.0349	0.062	-0.559	0.576	-0.157
0.087					
Mes__9	0.2117	0.059	3.580	0.000	0.096
0.328					

=====

=====

Logit Marginal Effects

=====

Dep. Variable:	Failure_today
Method:	dydx
At:	overall

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

Min_Temp	0.0159	0.001	24.595	0.000	0.015
0.017					
Max_Temp	-0.0304	0.001	-25.596	0.000	-0.033
-0.028					
Parameter1_Speed	0.0042	0.000	29.836	0.000	0.004
0.004					
Parameter3_9am	0.0016	0.000	8.164	0.000	0.001
0.002					
Parameter3_3pm	-0.0029	0.000	-14.286	0.000	-0.003
-0.002					
Parameter4_9am	0.0082	0.000	62.672	0.000	0.008
0.008					
Parameter4_3pm	0.0005	0.000	3.471	0.001	0.000
0.001					
Parameter5_9am	-0.0285	0.001	-36.277	0.000	-0.030
-0.027					
Parameter5_3pm	0.0224	0.001	28.554	0.000	0.021
0.024					
Parameter7_9am	-0.0042	0.001	-4.066	0.000	-0.006
-0.002					
Parameter7_3pm	0.0134	0.001	10.218	0.000	0.011
0.016					
Location__3	-0.0678	0.011	-6.235	0.000	-0.089
-0.046					
Location__4	0.0604	0.014	4.474	0.000	0.034
0.087					
Location__5	-0.0436	0.011	-4.090	0.000	-0.064
-0.023					
Location__6	-0.2375	0.011	-22.425	0.000	-0.258
-0.217					
Location__7	-0.1149	0.011	-10.897	0.000	-0.136
-0.094					
Location__8	0.0858	0.010	8.322	0.000	0.066
0.106					
Location__9	0.0402	0.011	3.708	0.000	0.019
0.061					
Location__10	-0.0583	0.011	-5.359	0.000	-0.080
-0.037					
Location__11	-0.0423	0.012	-3.580	0.000	-0.065
-0.019					
Location__12	0.0233	0.010	2.251	0.024	0.003
0.044					

Location__13 -0.138	-0.1587	0.010	-15.241	0.000	-0.179
Location__14 0.041	0.0188	0.011	1.660	0.097	-0.003
Location__15 0.031	0.0105	0.011	0.987	0.324	-0.010
Location__16 -0.095	-0.1154	0.011	-10.867	0.000	-0.136
Location__18 -0.093	-0.1160	0.012	-9.895	0.000	-0.139
Location__20 -0.114	-0.1345	0.010	-13.065	0.000	-0.155
Location__21 -0.115	-0.1379	0.011	-12.043	0.000	-0.160
Location__22 0.047	0.0241	0.012	2.065	0.039	0.001
Location__23 -0.073	-0.0930	0.010	-9.248	0.000	-0.113
Location__27 -0.074	-0.0936	0.010	-9.197	0.000	-0.114
Location__28 -0.063	-0.0821	0.010	-8.272	0.000	-0.102
Location__29 -0.098	-0.1192	0.011	-10.778	0.000	-0.141
Location__30 0.051	0.0296	0.011	2.677	0.007	0.008
Location__32 0.013	-0.0074	0.010	-0.724	0.469	-0.028
Location__33 0.022	0.0016	0.010	0.148	0.882	-0.019
Location__34 -0.099	-0.1180	0.010	-11.884	0.000	-0.138
Location__35 -0.011	-0.0320	0.011	-2.946	0.003	-0.053
Location__36 -0.132	-0.1526	0.011	-14.404	0.000	-0.173
Location__38 -0.019	-0.0391	0.010	-3.818	0.000	-0.059
Location__39 -0.012	-0.0327	0.011	-3.058	0.002	-0.054
Location__40 0.018	-0.0046	0.011	-0.408	0.683	-0.027
Location__41 -0.017	-0.0380	0.011	-3.543	0.000	-0.059
Location__43 -0.036	-0.0581	0.011	-5.201	0.000	-0.080
Location__44 -0.057	-0.0774	0.010	-7.570	0.000	-0.097

Location__45	-0.1374	0.010	-13.493	0.000	-0.157
-0.117					
Location__47	-0.0264	0.011	-2.511	0.012	-0.047
-0.006					
Location__48	-0.1178	0.011	-11.214	0.000	-0.138
-0.097					
Location__49	-0.1498	0.013	-11.515	0.000	-0.175
-0.124					
Parameter2_9am__N	-0.0019	0.004	-0.467	0.641	-0.010
0.006					
Parameter2_9am__S	0.0311	0.004	8.092	0.000	0.024
0.039					
Parameter2_9am__W	0.0313	0.004	7.072	0.000	0.023
0.040					
Parameter2_3pm__N	-0.0025	0.004	-0.625	0.532	-0.010
0.005					
Parameter2_3pm__S	0.0056	0.004	1.519	0.129	-0.002
0.013					
Parameter2_3pm__W	0.0139	0.004	3.206	0.001	0.005
0.022					
Mes__10	0.0190	0.007	2.884	0.004	0.006
0.032					
Mes__11	0.0353	0.006	5.809	0.000	0.023
0.047					
Mes__12	0.0257	0.006	4.128	0.000	0.013
0.038					
Mes__2	-0.0082	0.006	-1.393	0.164	-0.020
0.003					
Mes__3	-0.0049	0.006	-0.898	0.369	-0.016
0.006					
Mes__4	0.0095	0.006	1.616	0.106	-0.002
0.021					
Mes__5	-0.0198	0.006	-3.117	0.002	-0.032
-0.007					
Mes__6	-0.0563	0.007	-8.099	0.000	-0.070
-0.043					
Mes__7	-0.0378	0.007	-5.180	0.000	-0.052
-0.023					
Mes__8	-0.0040	0.007	-0.559	0.576	-0.018
0.010					
Mes__9	0.0242	0.007	3.582	0.000	0.011
0.037					

=====

=====

Odds Ratios

	Odds Ratio	5%	95%
Min_Temp	1.136922	1.162757	1.149767
Max_Temp	0.750522	0.782230	0.766212

Parameter1_Speed	1.034560	1.039639	1.037096
Parameter3_9am	1.010952	1.017945	1.014443
Parameter3_3pm	0.971649	0.978440	0.975039
Parameter4_9am	1.071758	1.076998	1.074375
Parameter4_3pm	1.001831	1.006601	1.004214
Parameter5_9am	0.768225	0.789854	0.778965
Parameter5_3pm	1.200198	1.233586	1.216777
Parameter7_9am	0.946498	0.980985	0.963587
Parameter7_3pm	1.099404	1.150186	1.124509
Location__3	0.458257	0.665793	0.552362
Location__4	1.346212	2.141445	1.697893
Location__5	0.568509	0.819725	0.682658
Location__6	0.103906	0.150163	0.124911
Location__7	0.304839	0.438214	0.365493

Para el logit, podemos identificar el mismo comportamiento que los 2 modelos anteriores, teniendo un aumento en la temperatura minima de una unidad un impacto en el 1.59% de probabilidad de fallo y en el max_temp una disminucion de 3%. Un cambio fuerte con respecto al modelo probit se da en la dirección del viento sur durante la tarde, que ahora nos da no significativa, lo que nos indica que se comporta de manera similar a la referencia del este, lo cual no ocurría en el modelo probit.

1.0.7 Parte 5

En los 3 modelos pudimos encontrar resultados bastante similares para la interpretación de las variables, con pequeños cambios en la significancia de algunas de estas, específicamente

Sobre cual modelo recomendaria, de base se descarta OLS al no ser el adecuado para trabajar con una variable dependiente binaria al no estar acotado para [0,1], aunque obviáramos esto es el que posee un R2 mas bajo entre los 3. Fuera de esto, entre logit y probit se comportan de la misma manera, pero al apreecer logit es mas exigente con la significancia. Este comportamiento más conservador puede ser beneficioso y marcar la diferencia a la hora de hacer interpretaciones precisas. Aun asi ambos modelos son extremadamente similares, por lo que no considero que este factor sea clave para elegir uno por sobre el otro de forma tajante.

Sobre la robustez de las variables, la gran mayoría de estas cumplen con esta característica, al ser significativas y tener impactos similares en todos los modelos. Algunas de estas son max_temp y min_temp, parameter1_speed, los parametros5 y algunas locaciones como la 6 y la 4, por otro lado, algunas de las no robustas son el mes 3 al no ser significativo en ningun modelo o el parameter2_3pm_S, al perder su significancia en el modelo logit.

1.0.8 Parte 6

Tratamiento de datos

```
[109]: X2.drop(columns=["Parameter2_9am",
                    "Parameter2_3pm",
                    #"Location"
                    ], inplace=True)

X2 = X2.groupby(["Año", "Mes", "Location"]).agg({
```



```

    **{col: "mean" for col in ['Min_Temp', 'Max_Temp', 'Parameter1_Speed',
        'Parameter3_9am', 'Parameter3_3pm', 'Parameter4_9am', 'Parameter4_3pm',
        'Parameter5_9am', 'Parameter5_3pm', 'Parameter7_9am', 'Parameter7_3pm']},
    "Failure_today": "sum"
}).reset_index()

X2["Año - Mes"] = X2["Año"].astype(str) + "-" + X2["Mes"].astype(str).str.
    ↪zfill(2)

```

[110]: X2

```

[110]:
      Año Mes  Location  Min_Temp  Max_Temp  Parameter1_Speed  \
0    2010   1         1  18.060714  30.917857         37.678571
1    2010   1         3  17.282759  34.420690         43.344828
2    2010   1         4  21.470000  35.153333         45.966667
3    2010   1         5  17.373684  30.884211         42.263158
4    2010   1         6  11.164286  27.246429         47.892857
...
3306  2017   6         44  10.726316  18.268421         35.421053
3307  2017   6         45   4.345000  14.870000         24.800000
3308  2017   6         47   8.827778  18.661111         37.666667
3309  2017   6         48  11.655556  17.611111         39.833333
3310  2017   6         49   5.952174  18.747826         28.000000

      Parameter3_9am  Parameter3_3pm  Parameter4_9am  Parameter4_3pm  \
0          10.857143          17.821429          43.928571          29.750000
1           9.689655          19.275862          49.724138          21.068966
2          18.700000          19.600000          38.566667          26.300000
3          10.473684          17.684211          69.052632          47.263158
4          20.357143          23.142857          58.107143          36.178571
...
3306          11.368421          12.578947          79.526316          66.894737
3307           6.200000           9.500000          97.300000          67.350000
3308          12.833333          18.222222          84.222222          68.888889
3309          15.888889          20.444444          73.111111          69.111111
3310          11.391304          13.391304          66.565217          36.608696

      Parameter5_9am  Parameter5_3pm  Parameter7_9am  Parameter7_3pm  \
0          1014.175000          1012.471429          22.932143          29.017857
1          1012.110345          1009.162069          23.110345          32.755172
2          1008.970000          1005.563333          28.276667          34.110000
3          1013.215789          1010.289474          22.500000          28.942105
4          1013.003571          1011.642857          17.835714          25.039286
...
3306          1023.984211          1021.800000          13.394737          17.205263
3307          1028.070000          1025.615000           6.650000          13.735000
3308          1022.788889          1020.844444          13.455556          17.305556

```

3309	1026.083333	1024.116667	14.705556	16.650000
3310	1029.586957	1026.939130	10.556522	18.052174

	Failure_today	Año	Mes
0		2	2010-01
1		3	2010-01
2		5	2010-01
3		4	2010-01
4		4	2010-01
...
3306		7	2017-06
3307		3	2017-06
3308		6	2017-06
3309		5	2017-06
3310		0	2017-06

[3311 rows x 16 columns]

Poisson

[111]: X3 = X2.copy()

```

y = X2['Failure_today']
X2=X2.drop(["Año - Mes",
            "Año",
            "Mes",
            #"Location",
            "Failure_today"
            ],
            axis=1)
#X2 = pd.get_dummies(X2, columns=["Año - Mes"], prefix=["Año y mes"],
                    ↪prefix_sep='__', drop_first=True)
X2 = pd.get_dummies(X2, columns=["Location"], prefix=["Location"],
                    ↪prefix_sep='__', drop_first=True)

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

#0.8101 sin location dummie

```

Generalized Linear Model Regression Results

Dep. Variable:	Failure_today	No. Observations:	3311
Model:	GLM	Df Residuals:	3261
Model Family:	Poisson	Df Model:	49
Link Function:	Log	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-7384.9

Date: jue, 24 abr. 2025 Deviance: 3484.2
Time: 22:27:31 Pearson chi2: 3.11e+03
No. Iterations: 5 Pseudo R-squ. (CS): 0.8453
Covariance Type: nonrobust

```
=====
```

	coef	std err	z	P> z	[0.025
0.975]					

const	19.4213	3.349	5.799	0.000	12.857
25.986					
Min_Temp	0.0570	0.011	5.151	0.000	0.035
0.079					
Max_Temp	-0.0149	0.026	-0.567	0.571	-0.066
0.037					
Parameter1_Speed	0.0548	0.003	18.120	0.000	0.049
0.061					
Parameter3_9am	-0.0045	0.004	-1.018	0.309	-0.013
0.004					
Parameter3_3pm	-0.0633	0.004	-14.686	0.000	-0.072
-0.055					
Parameter4_9am	0.0330	0.003	12.000	0.000	0.028
0.038					
Parameter4_3pm	-0.0025	0.003	-0.839	0.402	-0.008
0.003					
Parameter5_9am	-0.0232	0.020	-1.150	0.250	-0.063
0.016					
Parameter5_3pm	0.0040	0.021	0.194	0.846	-0.036
0.044					
Parameter7_9am	0.1293	0.016	8.039	0.000	0.098
0.161					
Parameter7_3pm	-0.1647	0.030	-5.584	0.000	-0.223
-0.107					
Location__3	-0.1381	0.072	-1.920	0.055	-0.279
0.003					
Location__4	0.0415	0.091	0.456	0.648	-0.137
0.220					
Location__5	-0.3150	0.075	-4.224	0.000	-0.461
-0.169					
Location__6	-0.4161	0.082	-5.099	0.000	-0.576
-0.256					
Location__7	-0.2279	0.072	-3.149	0.002	-0.370
-0.086					
Location__8	-0.1672	0.072	-2.331	0.020	-0.308
-0.027					
Location__9	-0.0874	0.079	-1.101	0.271	-0.243
0.068					

Location__10 0.031	-0.1236	0.079	-1.563	0.118	-0.279
Location__11 0.105	-0.0423	0.075	-0.562	0.574	-0.190
Location__12 0.105	-0.0327	0.070	-0.464	0.642	-0.171
Location__13 -0.284	-0.4245	0.072	-5.924	0.000	-0.565
Location__14 -0.242	-0.4053	0.083	-4.859	0.000	-0.569
Location__15 -0.017	-0.1759	0.081	-2.167	0.030	-0.335
Location__16 -0.438	-0.5636	0.064	-8.775	0.000	-0.690
Location__18 -0.277	-0.4312	0.079	-5.465	0.000	-0.586
Location__20 -0.136	-0.2788	0.073	-3.832	0.000	-0.421
Location__21 -0.001	-0.1644	0.083	-1.974	0.048	-0.328
Location__22 0.149	-0.0208	0.087	-0.240	0.810	-0.191
Location__23 0.084	-0.0535	0.070	-0.764	0.445	-0.191
Location__27 -0.424	-0.5528	0.066	-8.387	0.000	-0.682
Location__28 -0.400	-0.5379	0.071	-7.625	0.000	-0.676
Location__29 -0.061	-0.1955	0.069	-2.839	0.005	-0.330
Location__30 0.190	0.0484	0.072	0.669	0.503	-0.093
Location__32 0.068	-0.0619	0.067	-0.931	0.352	-0.192
Location__33 0.228	0.0879	0.072	1.225	0.220	-0.053
Location__34 -0.083	-0.2089	0.064	-3.259	0.001	-0.335
Location__35 -0.305	-0.4496	0.074	-6.074	0.000	-0.595
Location__36 -0.104	-0.2506	0.075	-3.340	0.001	-0.398
Location__38 -0.125	-0.2491	0.064	-3.920	0.000	-0.374
Location__39 0.047	-0.0881	0.069	-1.280	0.201	-0.223
Location__40 -0.257	-0.4277	0.087	-4.925	0.000	-0.598

Location__41	-0.2262	0.072	-3.143	0.002	-0.367
-0.085					
Location__43	0.0684	0.073	0.940	0.347	-0.074
0.211					
Location__44	-0.5949	0.064	-9.244	0.000	-0.721
-0.469					
Location__45	-0.4476	0.065	-6.846	0.000	-0.576
-0.319					
Location__47	-0.3397	0.068	-5.016	0.000	-0.472
-0.207					
Location__48	-0.7525	0.067	-11.173	0.000	-0.885
-0.620					
Location__49	-0.4308	0.095	-4.556	0.000	-0.616
-0.245					

```
=====
=====
```

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

```
[118]: const                27200350779.84
Min_Temp                    5.87
Max_Temp                    -1.48
Parameter1_Speed            5.64
Parameter3_9am              -0.45
Parameter3_3pm              -6.13
Parameter4_9am               3.35
Parameter4_3pm              -0.25
Parameter5_9am              -2.29
Parameter5_3pm               0.40
Parameter7_9am              13.81
Parameter7_3pm             -15.19
Location__3                 -12.90
Location__4                   4.24
Location__5                 -27.02
Location__6                 -34.04
Location__7                 -20.38
Location__8                 -15.40
Location__9                  -8.37
Location__10                -11.63
Location__11                 -4.14
Location__12                 -3.21
Location__13                -34.59
Location__14                -33.32
```

Location__15	-16.13
Location__16	-43.09
Location__18	-35.03
Location__20	-24.33
Location__21	-15.16
Location__22	-2.06
Location__23	-5.21
Location__27	-42.47
Location__28	-41.60
Location__29	-17.76
Location__30	4.96
Location__32	-6.01
Location__33	9.19
Location__34	-18.85
Location__35	-36.21
Location__36	-22.16
Location__38	-22.05
Location__39	-8.44
Location__40	-34.80
Location__41	-20.24
Location__43	7.08
Location__44	-44.84
Location__45	-36.09
Location__47	-28.80
Location__48	-52.88
Location__49	-35.00

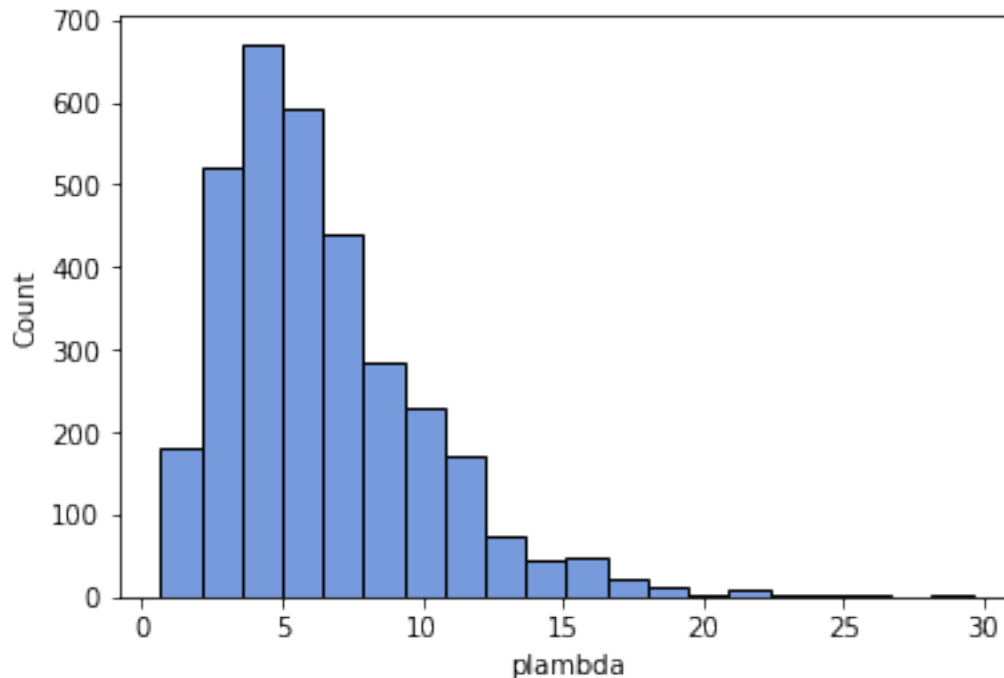
dtype: float64

Para este modelo evaluaremos la cantidad de fallos del sensor en el mes. Vemos comportamientos similares a los modelos previos, por ejemplo, ante un aumento de 1 unidad en la `min_temp`, la cantidad de fallos en el sensor durante el mes aumentan en un 5.87% en comparación al caso base. Aquí vemos un comportamiento distinto para los parametros 7, ya que un aumento en 1 unidad en la mañana impacta en un aumento del 13% a la cantidad de fallos, bastante mas que la variable de `min_temp`. Además, notamos que `Max_temp` no es significativa, al igual que el parametro 3 en la mañana, los parametros 4, y el 5 en la tarde. Lo que mas incide en la mayoría de casos son la ubicacion de los sensores, los cuales se deben comparar con el sensor base.

1.0.9 Parte 7

```
[119]: X2['plambda'] = poisson.mu
sns.histplot(data=X2, x="plambda", bins=20)
```

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



Podemos ver que hay una “cola larga” en la distribución

```
[120]: aux=((y-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:
3.918
Date:                   jue, 24 abr. 2025  Prob (F-statistic):
0.0479
Time:                   23:01:22         Log-Likelihood:
-5717.9
No. Observations:       3311            AIC:
1.144e+04
Df Residuals:           3310            BIC:
1.144e+04
Df Model:                1
Covariance Type:        nonrobust
```

	coef	std err	t	P> t	[0.025	0.975]
x1	-0.0064	0.003	-1.979	0.048	-0.013	-6.01e-05
Omnibus:		2964.467	Durbin-Watson:			1.815
Prob(Omnibus):		0.000	Jarque-Bera (JB):			167481.499
Skew:		4.044	Prob(JB):			0.00
Kurtosis:		36.891	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.

```
[123]: print(f"El alpha estimado es: {np.exp(-0.0064)}")
```

El alpha estimado es: 0.993620436379149

Al aplicar la estimación de alpha nos da un valor casi igual a 1, lo que confirma que hay sobredispersión, al no ser cercano a 0

```
[124]: nb = smf.negativebinomial("Failure_today ~ Min_Temp + Max_Temp +
    ↪Parameter1_Speed + Parameter3_9am + Parameter3_3pm + Parameter4_9am +
    ↪Parameter4_3pm + Parameter5_9am + Parameter5_3pm + Parameter7_9am +
    ↪Parameter7_3pm + C(Location)", data=X3).fit(alpha=1.24)
print(nb.summary())
```

```
c:\Users\joaqu\anaconda3\anaconda\lib\site-
packages\statsmodels\base\optimizer.py:17: FutureWarning: Keyword arguments have
been passed to the optimizer that have no effect. The list of allowed keyword
arguments for method bfgs is: gtol, norm, epsilon. The list of unsupported
keyword arguments passed include: alpha. After release 0.14, this will raise.
```

```
warnings.warn(
c:\Users\joaqu\anaconda3\anaconda\lib\site-
packages\statsmodels\discrete\discrete_model.py:2651: RuntimeWarning: divide by
zero encountered in log
```

```
llf = coeff + size*np.log(prob) + endog*np.log(1-prob)
c:\Users\joaqu\anaconda3\anaconda\lib\site-
packages\statsmodels\discrete\discrete_model.py:2651: RuntimeWarning: invalid
value encountered in multiply
```

```
llf = coeff + size*np.log(prob) + endog*np.log(1-prob)
c:\Users\joaqu\anaconda3\anaconda\lib\site-
packages\scipy\optimize\_optimize.py:1292: OptimizeWarning: Maximum number of
iterations has been exceeded.
```

```
res = _minimize_bfgs(f, x0, args, fprime, callback=callback, **opts)
```

Current function value: 2.230443

Iterations: 35
Function evaluations: 50
Gradient evaluations: 50

c:\Users\joaqu\anaconda3\anaconda\lib\site-packages\statsmodels\base\model.py:604: ConvergenceWarning: Maximum Likelihood optimization failed to converge. Check mle_retvals
warnings.warn("Maximum Likelihood optimization failed to "

NegativeBinomial Regression Results

```
=====
Dep. Variable:          Failure_today    No. Observations:          3311
Model:                NegativeBinomial    Df Residuals:              3261
Method:                MLE               Df Model:                 49
Date:                 jue, 24 abr. 2025    Pseudo R-squ.:            0.1976
Time:                 23:02:59            Log-Likelihood:           -7385.0
converged:             False              LL-Null:                  -9203.1
Covariance Type:       nonrobust          LLR p-value:              0.000
=====
```

	coef	std err	z	P> z	[0.025
0.975]					

Intercept	19.4220	3.352	5.794	0.000	12.852
25.992					
C(Location) [T.3]	-0.1373	0.072	-1.907	0.057	-0.278
0.004					
C(Location) [T.4]	0.0365	0.091	0.400	0.689	-0.142
0.215					
C(Location) [T.5]	-0.3146	0.075	-4.217	0.000	-0.461
-0.168					
C(Location) [T.6]	-0.4157	0.082	-5.092	0.000	-0.576
-0.256					
C(Location) [T.7]	-0.2258	0.072	-3.119	0.002	-0.368
-0.084					
C(Location) [T.8]	-0.1686	0.072	-2.347	0.019	-0.309
-0.028					
C(Location) [T.9]	-0.0958	0.080	-1.204	0.229	-0.252
0.060					
C(Location) [T.10]	-0.1231	0.079	-1.556	0.120	-0.278
0.032					
C(Location) [T.11]	-0.0483	0.075	-0.641	0.522	-0.196
0.099					
C(Location) [T.12]	-0.0352	0.070	-0.500	0.617	-0.173
0.103					
C(Location) [T.13]	-0.4213	0.072	-5.876	0.000	-0.562
-0.281					
C(Location) [T.14]	-0.4101	0.084	-4.889	0.000	-0.574

-0.246					
C(Location) [T.15]	-0.1813	0.081	-2.232	0.026	-0.340
-0.022					
C(Location) [T.16]	-0.5630	0.064	-8.760	0.000	-0.689
-0.437					
C(Location) [T.18]	-0.4261	0.079	-5.401	0.000	-0.581
-0.271					
C(Location) [T.20]	-0.2806	0.073	-3.854	0.000	-0.423
-0.138					
C(Location) [T.21]	-0.1715	0.083	-2.055	0.040	-0.335
-0.008					
C(Location) [T.22]	-0.0244	0.087	-0.282	0.778	-0.194
0.145					
C(Location) [T.23]	-0.0522	0.070	-0.745	0.456	-0.189
0.085					
C(Location) [T.27]	-0.5561	0.066	-8.430	0.000	-0.685
-0.427					
C(Location) [T.28]	-0.5429	0.071	-7.684	0.000	-0.681
-0.404					
C(Location) [T.29]	-0.1945	0.069	-2.824	0.005	-0.329
-0.060					
C(Location) [T.30]	0.0494	0.072	0.683	0.495	-0.092
0.191					
C(Location) [T.32]	-0.0604	0.067	-0.907	0.365	-0.191
0.070					
C(Location) [T.33]	0.0876	0.072	1.220	0.223	-0.053
0.228					
C(Location) [T.34]	-0.2036	0.064	-3.175	0.001	-0.329
-0.078					
C(Location) [T.35]	-0.4496	0.074	-6.072	0.000	-0.595
-0.304					
C(Location) [T.36]	-0.2496	0.075	-3.326	0.001	-0.397
-0.103					
C(Location) [T.38]	-0.2533	0.064	-3.983	0.000	-0.378
-0.129					
C(Location) [T.39]	-0.0939	0.069	-1.363	0.173	-0.229
0.041					
C(Location) [T.40]	-0.4362	0.087	-5.011	0.000	-0.607
-0.266					
C(Location) [T.41]	-0.2252	0.072	-3.127	0.002	-0.366
-0.084					
C(Location) [T.43]	0.0678	0.073	0.932	0.351	-0.075
0.210					
C(Location) [T.44]	-0.5925	0.064	-9.203	0.000	-0.719
-0.466					
C(Location) [T.45]	-0.4472	0.065	-6.835	0.000	-0.575
-0.319					
C(Location) [T.47]	-0.3370	0.068	-4.972	0.000	-0.470

-0.204					
C(Location) [T.48]	-0.7548	0.067	-11.193	0.000	-0.887
-0.623					
C(Location) [T.49]	-0.4377	0.095	-4.623	0.000	-0.623
-0.252					
Min_Temp	0.0580	0.011	5.240	0.000	0.036
0.080					
Max_Temp	-0.0158	0.026	-0.601	0.548	-0.067
0.036					
Parameter1_Speed	0.0546	0.003	17.953	0.000	0.049
0.061					
Parameter3_9am	-0.0043	0.004	-0.968	0.333	-0.013
0.004					
Parameter3_3pm	-0.0630	0.004	-14.528	0.000	-0.071
-0.054					
Parameter4_9am	0.0330	0.003	11.996	0.000	0.028
0.038					
Parameter4_3pm	-0.0026	0.003	-0.875	0.382	-0.009
0.003					
Parameter5_9am	-0.0207	0.020	-1.029	0.304	-0.060
0.019					
Parameter5_3pm	0.0015	0.021	0.075	0.940	-0.039
0.042					
Parameter7_9am	0.1291	0.016	8.010	0.000	0.098
0.161					
Parameter7_3pm	-0.1644	0.030	-5.569	0.000	-0.222
-0.107					
alpha	8.499e-05	0.004	0.023	0.982	-0.007
0.007					

=====

=====

```
[125]: np.exp(8.499e-05)
```

```
[125]: 1.0000849936117524
```

Y, al correr una binomial sin especificar alpha, nos entrega su estimación de 8.499e-05, que al aplicarle exponencial nos da aproximadamente 1, un valor similar al que estimamos arriba

1.0.10 Parte 8

```
[126]: negbin=sm.GLM(y,X2,family=sm.families.NegativeBinomial(alpha=np.exp(-0.0064))).
        ↪fit()
        print(negbin.summary())
```

Generalized Linear Model Regression Results

```
=====
```

Dep. Variable:	Failure_today	No. Observations:	3311
Model:	GLM	Df Residuals:	3260

Model Family: NegativeBinomial Df Model: 50
Link Function: Log Scale: 1.0000
Method: IRLS Log-Likelihood: -9225.9
Date: jue, 24 abr. 2025 Deviance: 763.67
Time: 23:04:02 Pearson chi2: 562.
No. Iterations: 9 Pseudo R-squ. (CS): 0.2530
Covariance Type: nonrobust

=====

	coef	std err	z	P> z	[0.025
0.975]					

const	30.7950	9.719	3.169	0.002	11.746
49.844					
Min_Temp	0.0615	0.029	2.109	0.035	0.004
0.119					
Max_Temp	-0.0362	0.072	-0.501	0.616	-0.178
0.105					
Parameter1_Speed	0.1052	0.013	8.208	0.000	0.080
0.130					
Parameter3_9am	-0.0099	0.012	-0.814	0.416	-0.034
0.014					
Parameter3_3pm	-0.1163	0.015	-7.772	0.000	-0.146
-0.087					
Parameter4_9am	0.0502	0.008	6.538	0.000	0.035
0.065					
Parameter4_3pm	0.0057	0.009	0.666	0.505	-0.011
0.023					
Parameter5_9am	-0.0445	0.056	-0.801	0.423	-0.153
0.064					
Parameter5_3pm	0.0126	0.057	0.223	0.824	-0.099
0.124					
Parameter7_9am	0.2184	0.043	5.137	0.000	0.135
0.302					
Parameter7_3pm	-0.2269	0.081	-2.816	0.005	-0.385
-0.069					
Location__3	-0.1146	0.196	-0.584	0.559	-0.499
0.270					
Location__4	0.0503	0.219	0.230	0.818	-0.379
0.480					
Location__5	-0.5866	0.206	-2.851	0.004	-0.990
-0.183					
Location__6	-0.7092	0.240	-2.951	0.003	-1.180
-0.238					
Location__7	-0.3248	0.199	-1.636	0.102	-0.714
0.064					
Location__8	-0.3303	0.198	-1.670	0.095	-0.718

0.057					
Location__9	-0.2285	0.228	-1.003	0.316	-0.675
0.218					
Location__10	-0.2262	0.215	-1.051	0.293	-0.648
0.196					
Location__11	0.0510	0.189	0.269	0.788	-0.320
0.422					
Location__12	-0.1705	0.203	-0.840	0.401	-0.568
0.227					
Location__13	-0.7476	0.216	-3.467	0.001	-1.170
-0.325					
Location__14	-0.8825	0.237	-3.720	0.000	-1.347
-0.418					
Location__15	-0.3656	0.227	-1.614	0.107	-0.810
0.078					
Location__16	-0.9798	0.203	-4.824	0.000	-1.378
-0.582					
Location__18	-0.7520	0.227	-3.310	0.001	-1.197
-0.307					
Location__20	-0.4806	0.211	-2.281	0.023	-0.894
-0.068					
Location__21	-0.0657	0.204	-0.322	0.747	-0.466
0.334					
Location__22	-0.0252	0.229	-0.110	0.912	-0.474
0.423					
Location__23	-0.0857	0.208	-0.413	0.680	-0.493
0.321					
Location__27	-0.9601	0.203	-4.730	0.000	-1.358
-0.562					
Location__28	-0.9429	0.211	-4.462	0.000	-1.357
-0.529					
Location__29	-0.3110	0.190	-1.640	0.101	-0.683
0.061					
Location__30	-0.0512	0.200	-0.256	0.798	-0.443
0.341					
Location__32	-0.2575	0.180	-1.432	0.152	-0.610
0.095					
Location__33	-0.0418	0.198	-0.211	0.833	-0.430
0.346					
Location__34	-0.4088	0.194	-2.110	0.035	-0.789
-0.029					
Location__35	-0.7355	0.205	-3.580	0.000	-1.138
-0.333					
Location__36	-0.5069	0.214	-2.364	0.018	-0.927
-0.087					
Location__38	-0.4168	0.187	-2.234	0.025	-0.782
-0.051					
Location__39	-0.1747	0.198	-0.881	0.378	-0.563

0.214					
Location__40	-0.8546	0.236	-3.626	0.000	-1.317
-0.393					
Location__41	-0.3407	0.196	-1.736	0.083	-0.725
0.044					
Location__43	0.1603	0.198	0.810	0.418	-0.227
0.548					
Location__44	-1.0691	0.204	-5.251	0.000	-1.468
-0.670					
Location__45	-0.7551	0.197	-3.842	0.000	-1.140
-0.370					
Location__47	-0.6306	0.199	-3.165	0.002	-1.021
-0.240					
Location__48	-1.3326	0.219	-6.085	0.000	-1.762
-0.903					
Location__49	-0.4038	0.215	-1.882	0.060	-0.825
0.017					
plambda	-0.0820	0.019	-4.336	0.000	-0.119
-0.045					

```
=====
=====
```

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

```
[127]: const                2366508221587123.00
Min_Temp                    6.35
Max_Temp                   -3.56
Parameter1_Speed           11.09
Parameter3_9am             -0.99
Parameter3_3pm            -10.98
Parameter4_9am              5.15
Parameter4_3pm              0.57
Parameter5_9am             -4.35
Parameter5_3pm              1.27
Parameter7_9am             24.41
Parameter7_3pm            -20.30
Location__3                -10.83
Location__4                 5.16
Location__5               -44.38
Location__6               -50.80
Location__7               -27.74
Location__8               -28.13
Location__9               -20.42
```

Location__10	-20.24
Location__11	5.23
Location__12	-15.67
Location__13	-52.65
Location__14	-58.63
Location__15	-30.62
Location__16	-62.46
Location__18	-52.86
Location__20	-38.16
Location__21	-6.36
Location__22	-2.49
Location__23	-8.21
Location__27	-61.71
Location__28	-61.05
Location__29	-26.73
Location__30	-4.99
Location__32	-22.70
Location__33	-4.10
Location__34	-33.56
Location__35	-52.07
Location__36	-39.76
Location__38	-34.08
Location__39	-16.03
Location__40	-57.46
Location__41	-28.87
Location__43	17.38
Location__44	-65.67
Location__45	-53.00
Location__47	-46.77
Location__48	-73.62
Location__49	-33.22
plambda	-7.87
dtype:	float64

Este modelo tiene un ajuste bastante inferior en comparacion al Poisson, pero con efectos muy similares, por ejemplo, si la `min_temp` aumenta en una unidad, la cantidad de fallos aumentará en un 6%, y en el caso de la velocidad del viento, una unidad impacta en un 11% a la cantidad de fallos en dicho mes. Al igual que en poisson, muchas variables dieron no significativas, y los cambios mas grandes se ven en la ubicacion de los sensores.

1.0.11 Parte 9

Al realizar el test de sobredispersión y obtener los resultados de que si existia, podemos descartar automaticamente el modelo poisson, que si bien nos entregaba un buen ajuste este no era adecuado para el problema. Es asi que, al aplicar el poisson, empeoró bastante, lo cual tiene sentido cuando comparamos con los modelos anteriores (Siendo estudios distintos, hace sentido que tengna un comportamiento similar). Además, si bien un modelo no era adecuado, los coeficientes de ambos se comportan similar con leves diferencias. Las variables robustas fueron `min_temp`,

parameter1_speed, parameter3_3p y parameter4_9am entre otros, pero disminuye la cantidad de robustos en comparación al caso estudiado en 2, 3 y 4