




PRÀCTICA 1

ANÀLISI

DATASET

Forest Fires

BIEL GONZÁLEZ
JUDIT PANISELLO
CRISTINA SOLER

- 
- 01 INTRODUCCIÓ I OBJECTIUS**
 - 02 EXPLICACIÓ DATASET**
 - 03 APARTAT C - DESCRIPCIÓ**
 - 04 TRACTAMENT DE DADES**
 - 05 APARTAT B - REGRESSIONS**
 - 06 CONCLUSIONS**

INTRODUCCIÓ

Una de les principals preocupacions ambientals és l'aparició d'incendis forestals, que afecten la preservació dels boscos, creen danys econòmics i ecològics i causen patiment humà. Hauríem d'aconseguir partir de dades d'incendis anteriors poder predir l'extensió d'un incendi a partir de magnituds mesurables a la natura.

OBJECTIUS

1

**DESCRIURE
CORRECTAMENT LES
DADES**

2

**TROBAR UNA
REGRESSIÓ PER PREDIR
L'ÀREA CREMADA**

DESCRIPCIÓ DEL DATASET

El dataset seleccionat pel treball conté les dades d'incendis forestals en el parc natural de Montesinho, situat a la regió nord-est de Trás-os-Montes a Portugal.



APARTAT C - DESCRIPCIÓ

1 INFORMACIÓ BÀSICA

2 DESCRIPCIÓ MAPA

3 HISTOGRAMES

4 PAIR PLOTS

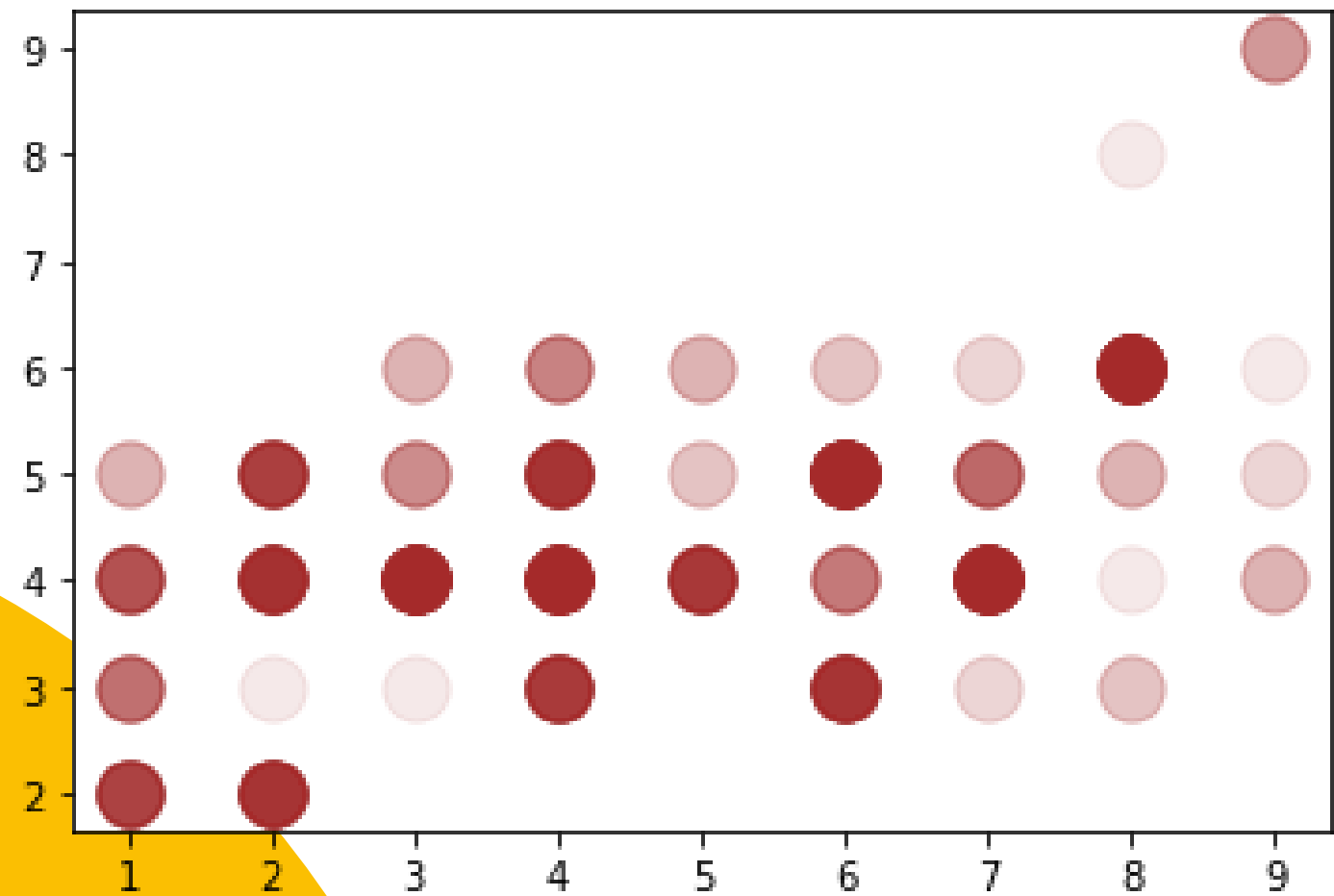
5 CORRELACIÓ DE DADES

APARTAT C

INFORMACIÓ BÀSICA

- 517 línies i 13 columnes
- 2 coordenades de posició
- 5 coeficients de l'índex FWI
- Temperatura, pluja i vent
- Variable objectiu àrea

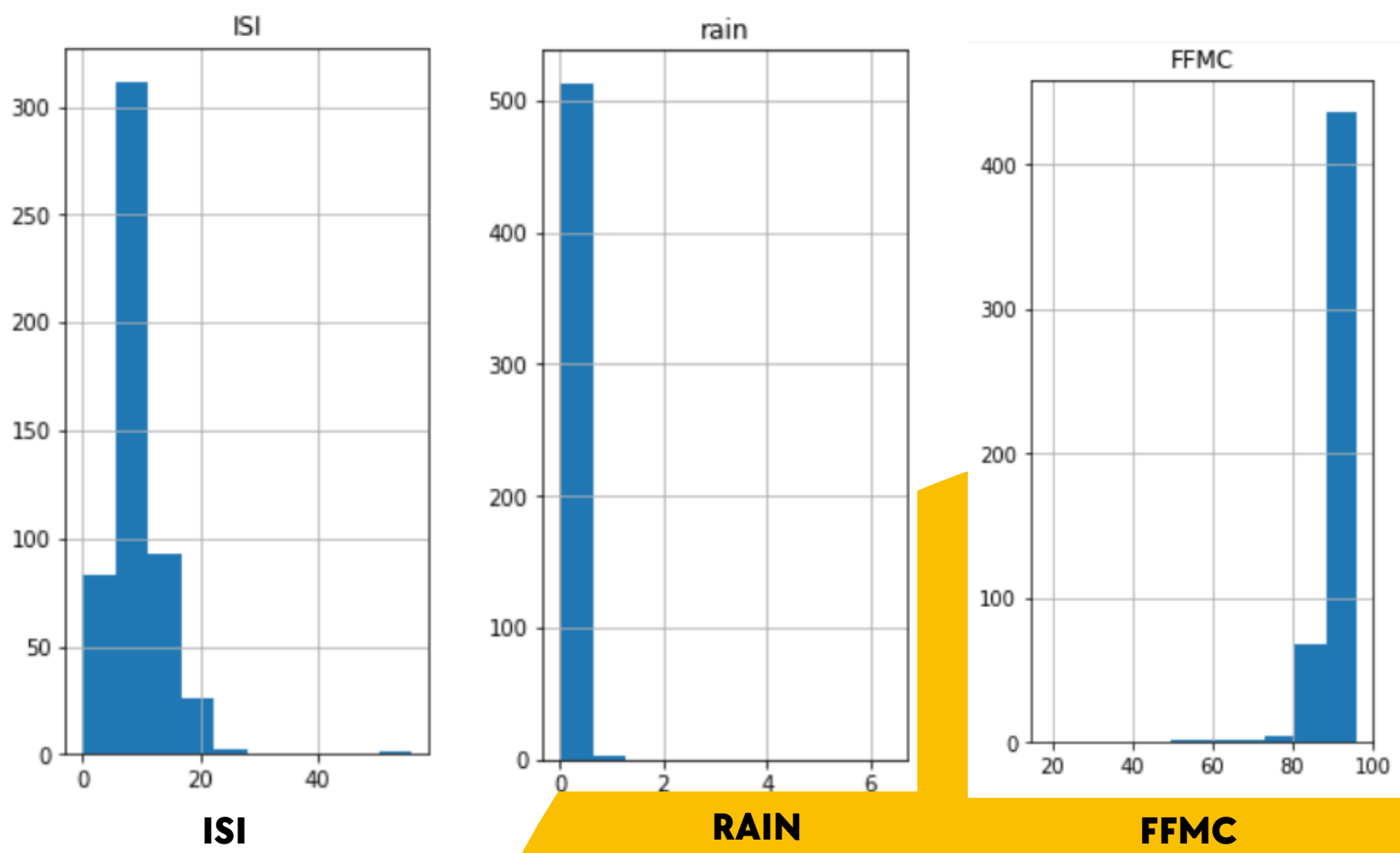
DESCRIPCIÓ MAPA



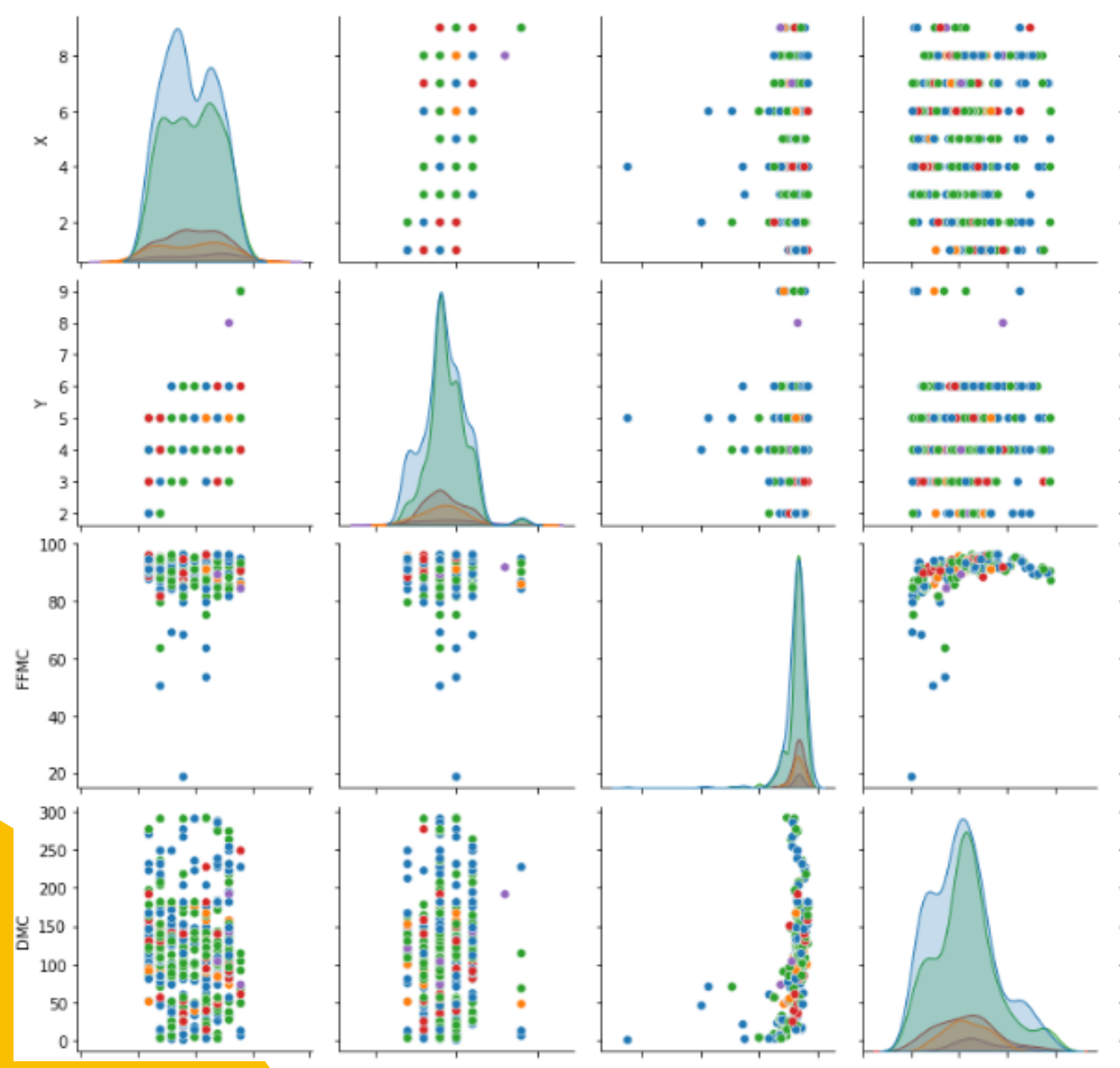
DESCRIPCIÓ PUNTS CREMATS AL PARC

APARTAT C

HISTOGRAMMES



PAIR PLOT



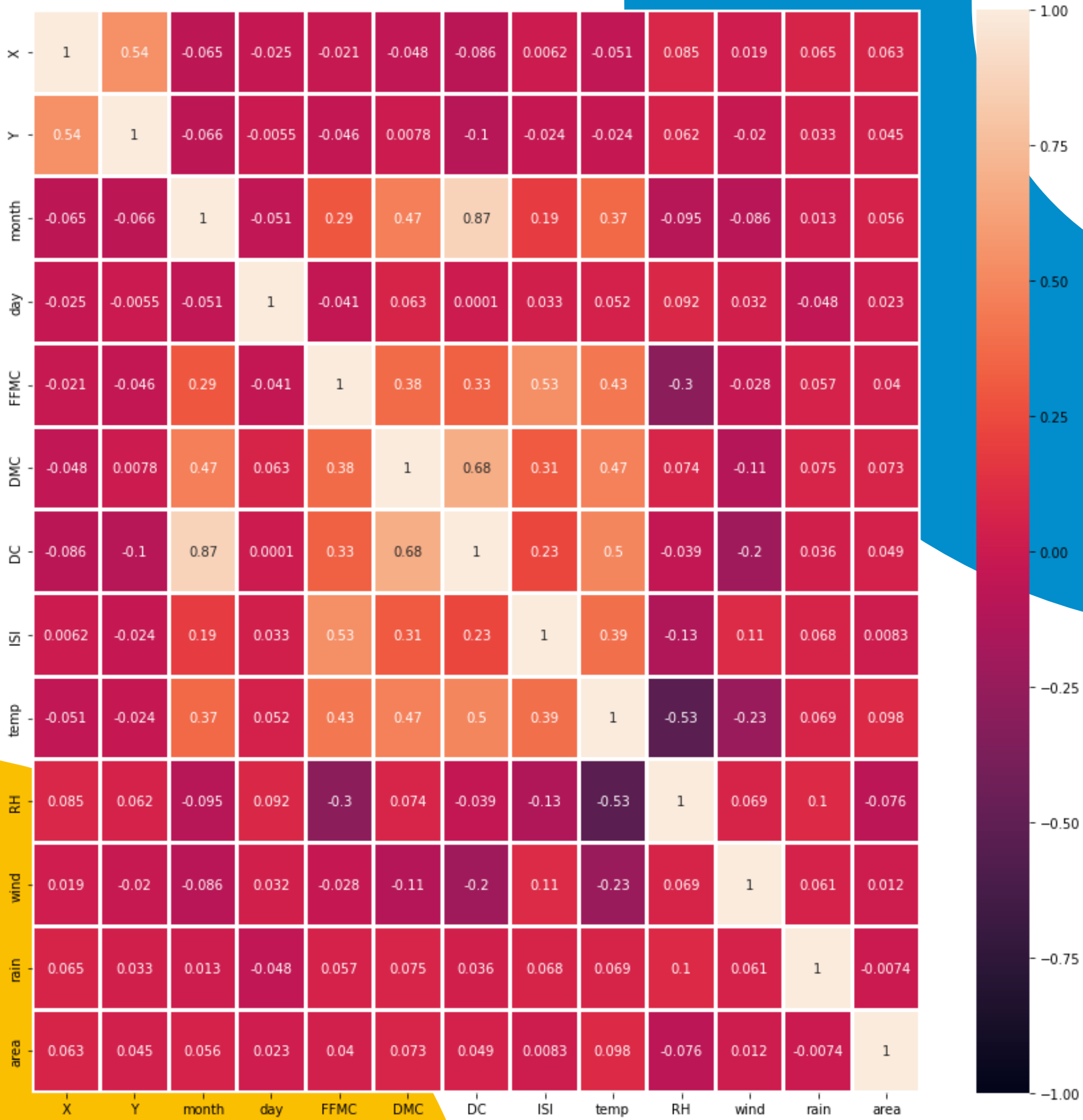
PETITA PART DEL PAIR PLOT SEGONS EL DANY FET

APARTAT C

CORRELACIÓ DE DADES

Poca correlació entre les dades i la variable objectiu

GRÀFICA DE CORRELACIÓ DE LES DADES



TRACTAMENT DE DADES

1

TRANSFORMACIÓ LOGARÍTMICA VARIABLE OBJECTIU

Millora resultats de la regressió
i simetria de les dades

2

ONE- HOT

L'usem per tractar els valors
categòrics

3

SEPAREM EL MODEL EN DUES PARTS

Tindrem dos models:

- Àrea igual a zero
- Àrea diferent a zero

APARTAT B - REGRESSIONS

1

**REGRESSIÓ LINEAL
SIMPLE I MÚLTIPLE**

2

REGRESSIÓ POLINOMIAL

3

REGRESSIÓ LOGÍSTICA

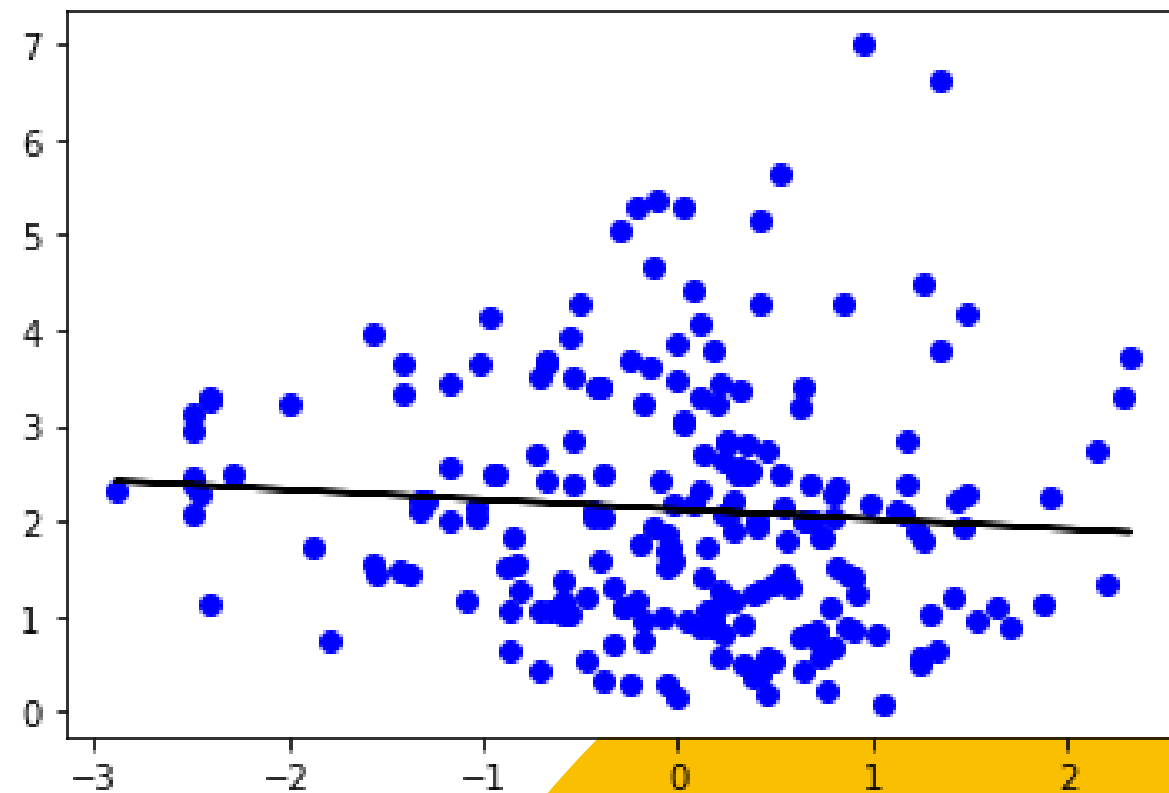
4

REGRESSIÓ LASSO

REGRESSIÓ LINEAL

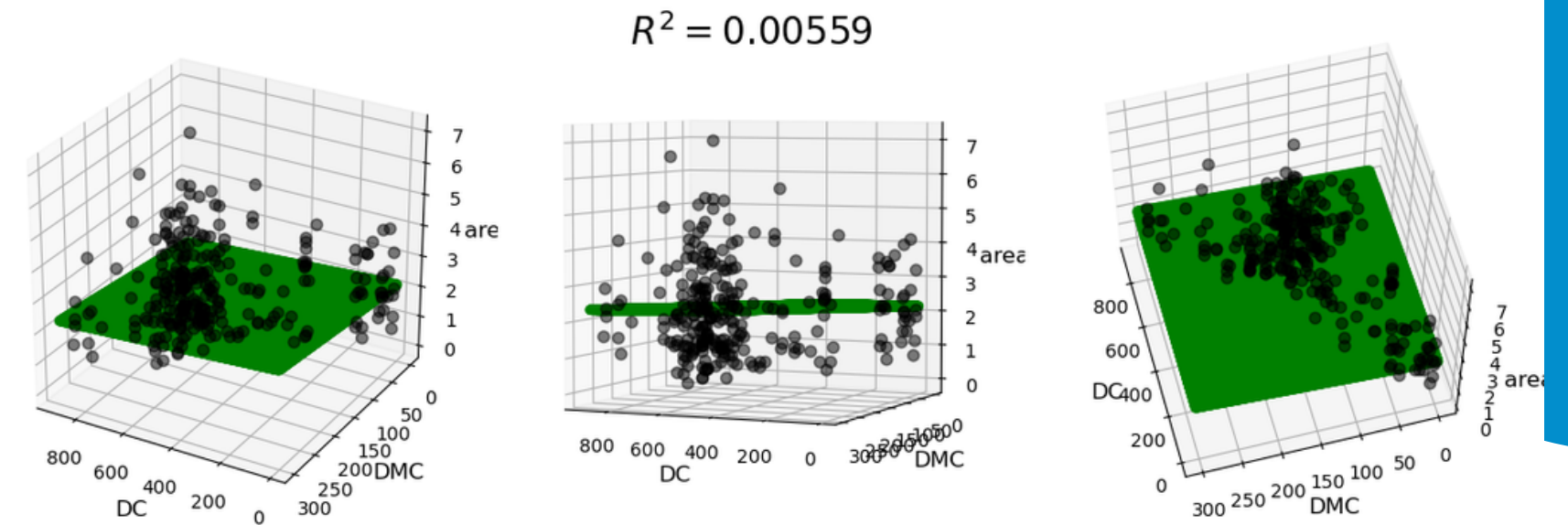
SIMPLE

- Obtenim Coeficients de determinació molt baixos (aprox. zero)



REGRESSIÓ LINEAL TEMPERATURA - ÀREA

MÚLTIPLE

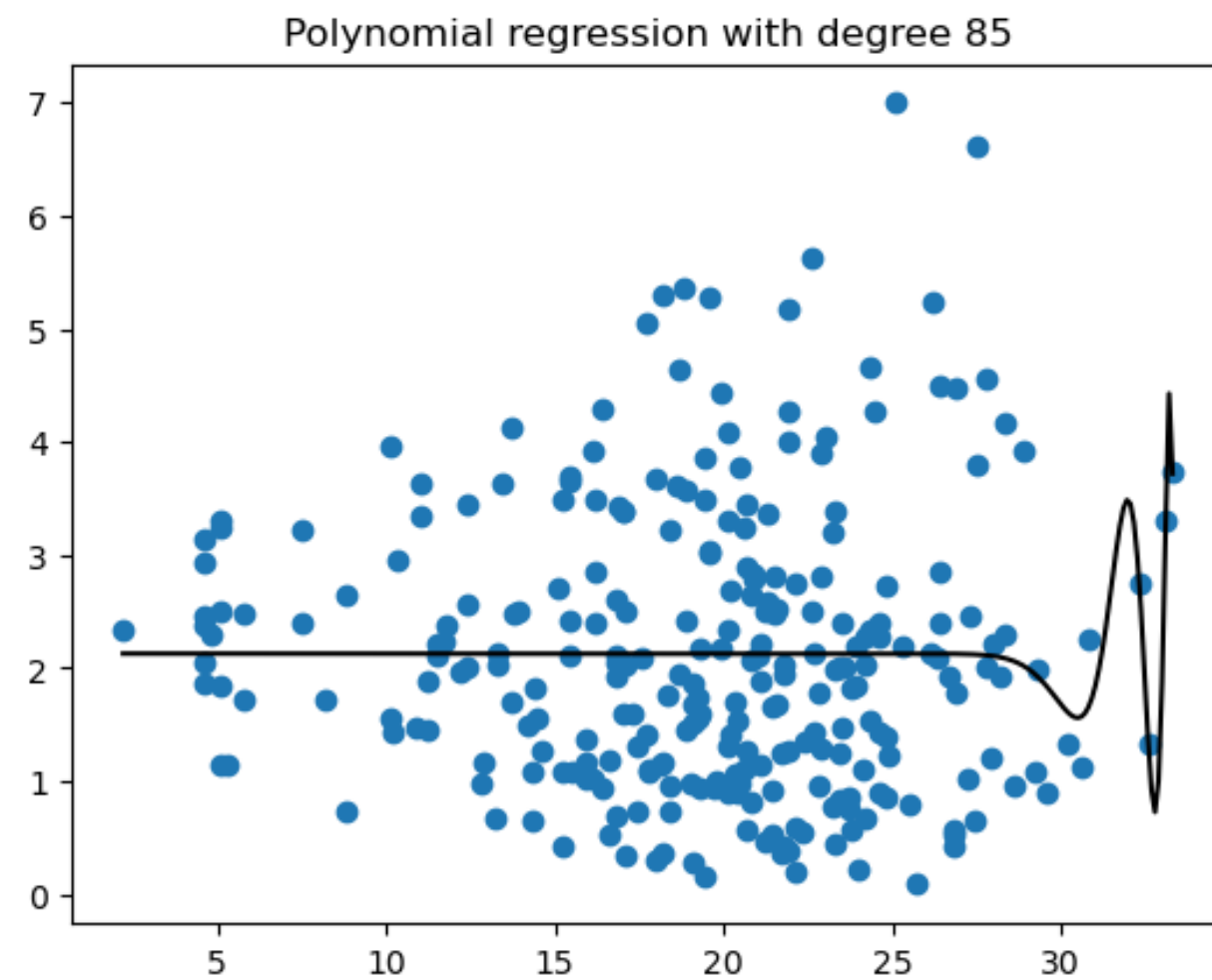


REGRESSIÓ MÚLTIPLE DMC - DC - ÀREA

- Regressió múltiple amb tots els atributs + one hot són els que millors hem obtingut coeficient de determinació igual a 0.20

POLINOMIAL

- Tornem a tenir un coeficient de determinació baix, i el grau de la polinomial és molt alt
- El polinomi no s'acaba d'ajustar al núvol de punts



REGRESSIÓ POLINÒMICA TEMPERATURA - ÀREA

LOGÍSTICA

- Separem per dany de l'àrea

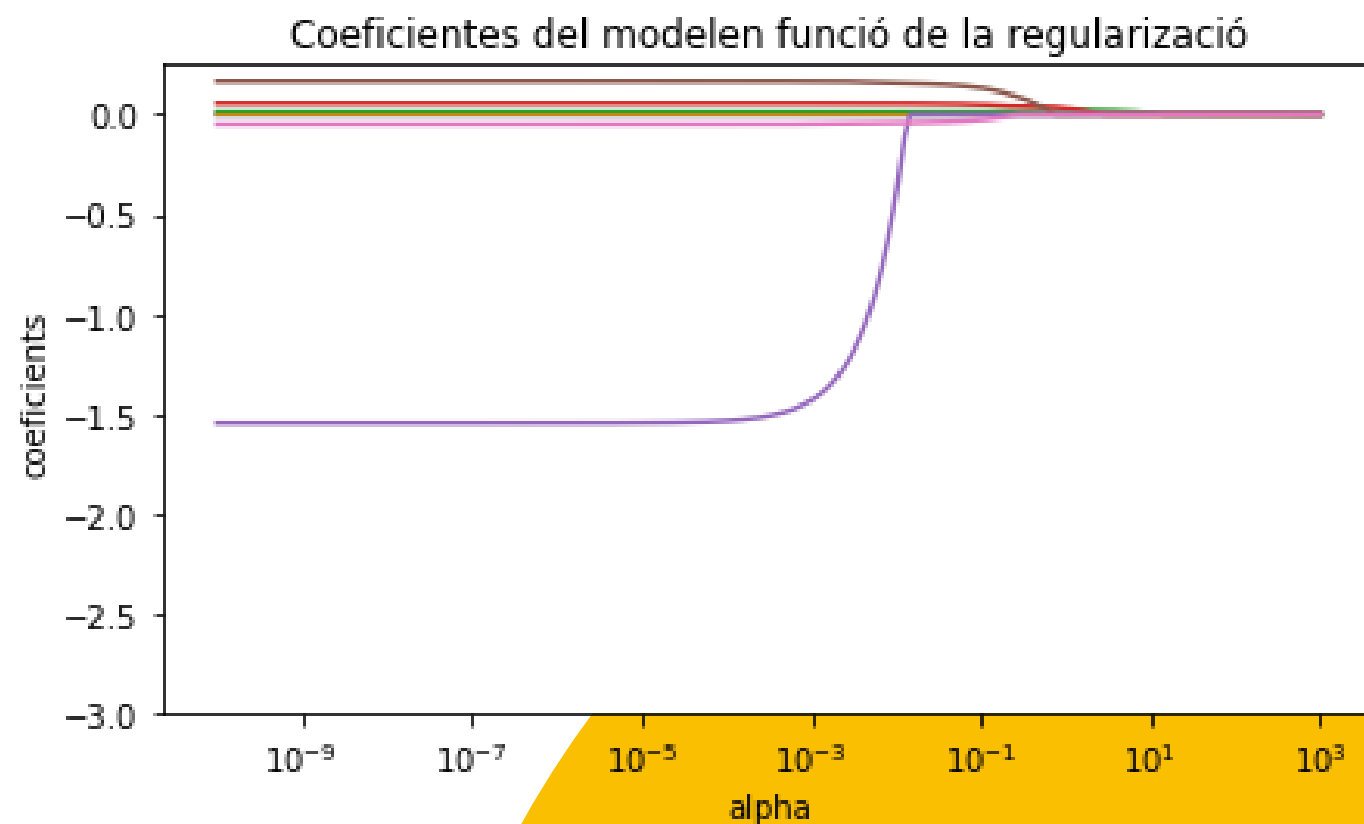
| <i>Damage Category</i> | <i>Area</i> |
|------------------------|-------------|
| Moderate (0) | ≤ 25 |
| High (1) | > 25 |

- Veiem resultats mitjanament bons al MSE, però a la matriu de confusió no.

REGRESSIÓ LASSO

LASSO

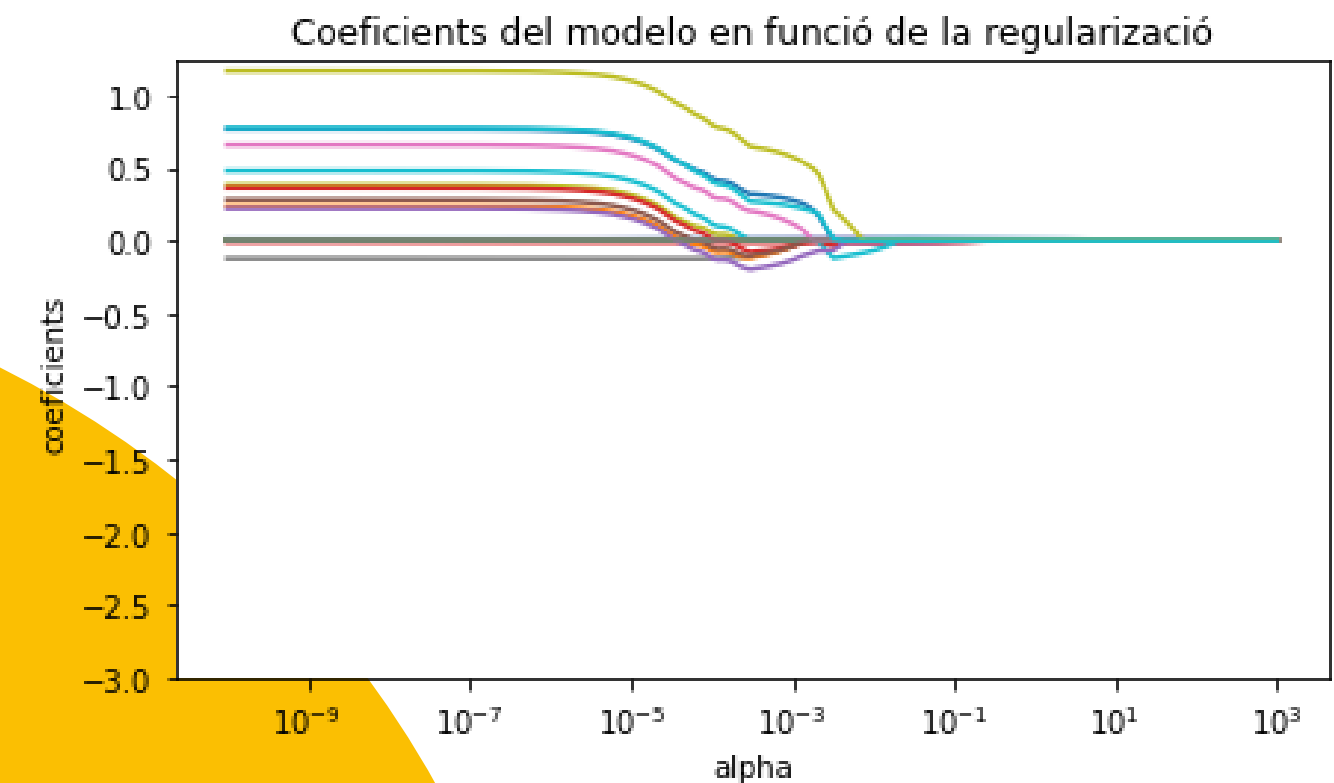
- Eliminem predictors amb poca importància i triem la millor alfa
- Obtenim un coeficient de determinació negatiu, dades relacionades inversament



GRÀFIC PER VISUALITZAR EL COMPORTAMENT DELS COEFICIENTS

LASSO + ONE-HOT

- Apliquem el One hot per millorar els resultats obtinguts a la regressió Lasso
- Obtenim que cap coeficient serveix



GRÀFIC PER VISUALITZAR EL COMPORTAMENT DELS COEFICIENTS



CONCLUSIÓ