



# RED WINE QUALITY

POL ESPINASA -- I566792

# ÍNDEX

- Introducció a la Base de Dades
- Anàlisi i estudi de les dades
- Models
- Cerca d'hiperparàmetres
- Conclusions

# INTRODUCCIÓ A LA BASE DE DADES



# ATRIBUTS

	count	mean	std	min	25%	50%	75%	max
fixed acidity	1599.0	8.319637	1.741096	4.60000	7.1000	7.90000	9.200000	15.90000
volatile acidity	1599.0	0.527821	0.179060	0.12000	0.3900	0.52000	0.640000	1.58000
citric acid	1599.0	0.270976	0.194801	0.00000	0.0900	0.26000	0.420000	1.00000
residual sugar	1599.0	2.538806	1.409928	0.90000	1.9000	2.20000	2.600000	15.50000
chlorides	1599.0	0.087467	0.047065	0.01200	0.0700	0.07900	0.090000	0.61100
free sulfur dioxide	1599.0	15.874922	10.460157	1.00000	7.0000	14.00000	21.000000	72.00000
total sulfur dioxide	1599.0	46.467792	32.895324	6.00000	22.0000	38.00000	62.000000	289.00000
density	1599.0	0.996747	0.001887	0.99007	0.9956	0.99675	0.997835	1.00369
pH	1599.0	3.311113	0.154386	2.74000	3.2100	3.31000	3.400000	4.01000
sulphates	1599.0	0.658149	0.169507	0.33000	0.5500	0.62000	0.730000	2.00000
alcohol	1599.0	10.422983	1.065668	8.40000	9.5000	10.20000	11.100000	14.90000
quality	1599.0	5.636023	0.807569	3.00000	5.0000	6.00000	6.000000	8.00000

# TIPUS DE DADES I VALORS NULLS

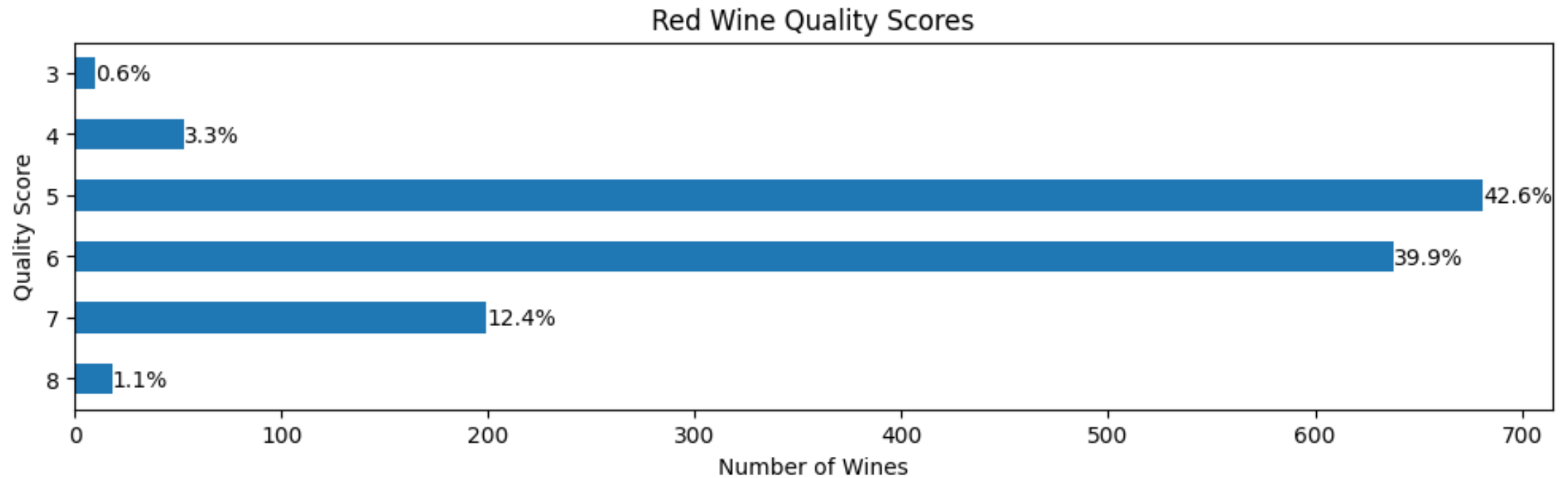
```
fixed acidity      0
volatile acidity   0
citric acid        0
residual sugar     0
chlorides          0
free sulfur dioxide 0
total sulfur dioxide 0
density           0
pH                0
sulphates          0
alcohol            0
quality            0
dtype: int64
```

#	Column	Non-Null Count	Dtype
0	fixed acidity	1599 non-null	float64
1	volatile acidity	1599 non-null	float64
2	citric acid	1599 non-null	float64
3	residual sugar	1599 non-null	float64
4	chlorides	1599 non-null	float64
5	free sulfur dioxide	1599 non-null	float64
6	total sulfur dioxide	1599 non-null	float64
7	density	1599 non-null	float64
8	pH	1599 non-null	float64
9	sulphates	1599 non-null	float64
10	alcohol	1599 non-null	float64
11	quality	1599 non-null	int64

No hi ha valors nulls!!

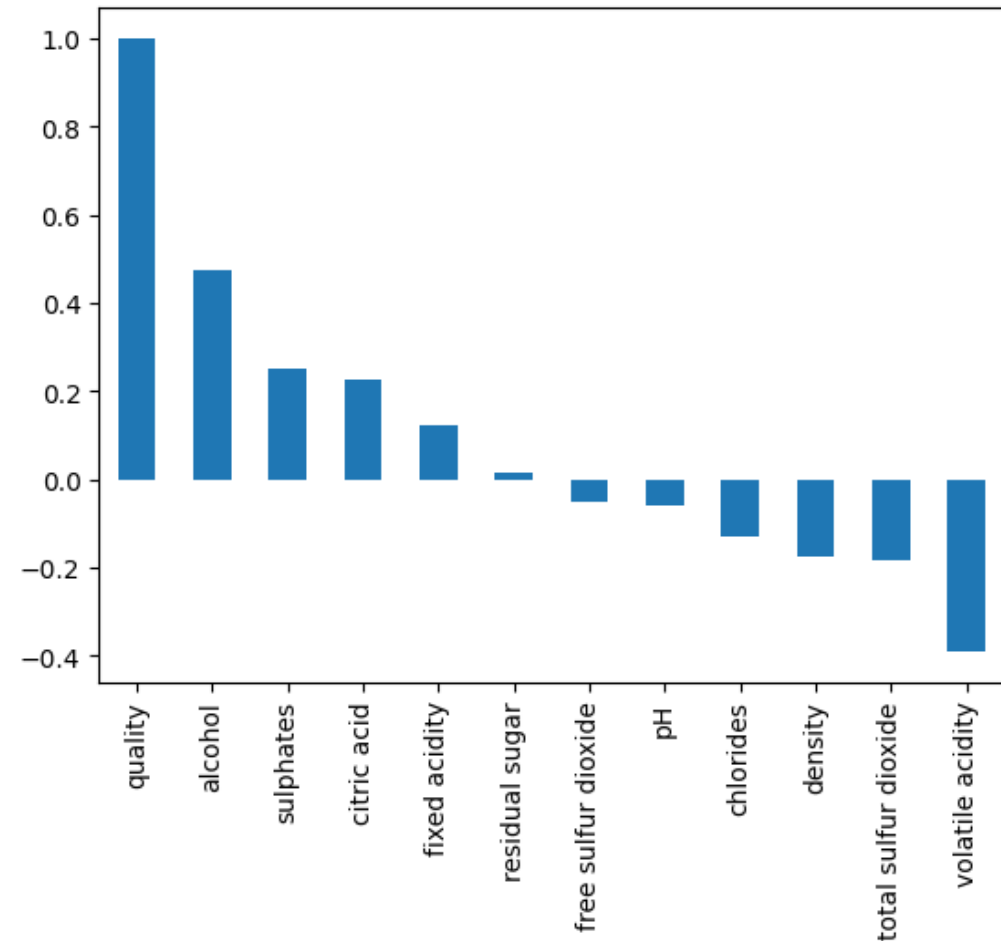
Tipus de dades senzills de tractar

# ATRIBUT OBJECTIU

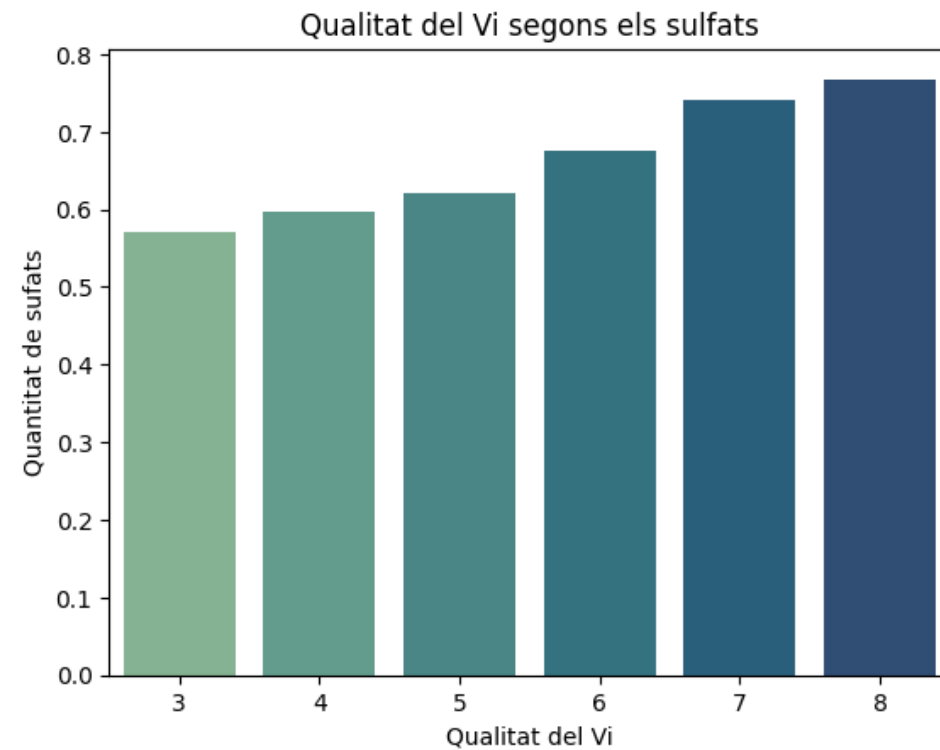
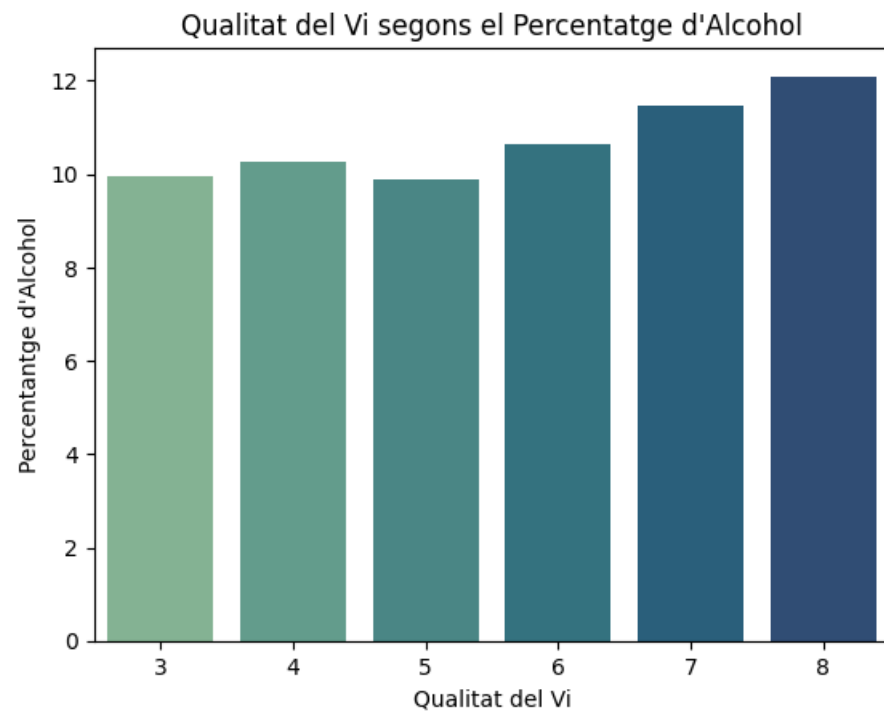


# CORRELACIÓ ENTRE ATRIBUTS

quality	1.000000
alcohol	0.476166
sulphates	0.251397
citric acid	0.226373
fixed acidity	0.124052
residual sugar	0.013732
free sulfur dioxide	-0.050656
pH	-0.057731
chlorides	-0.128907
density	-0.174919
total sulfur dioxide	-0.185100
volatile acidity	-0.390558



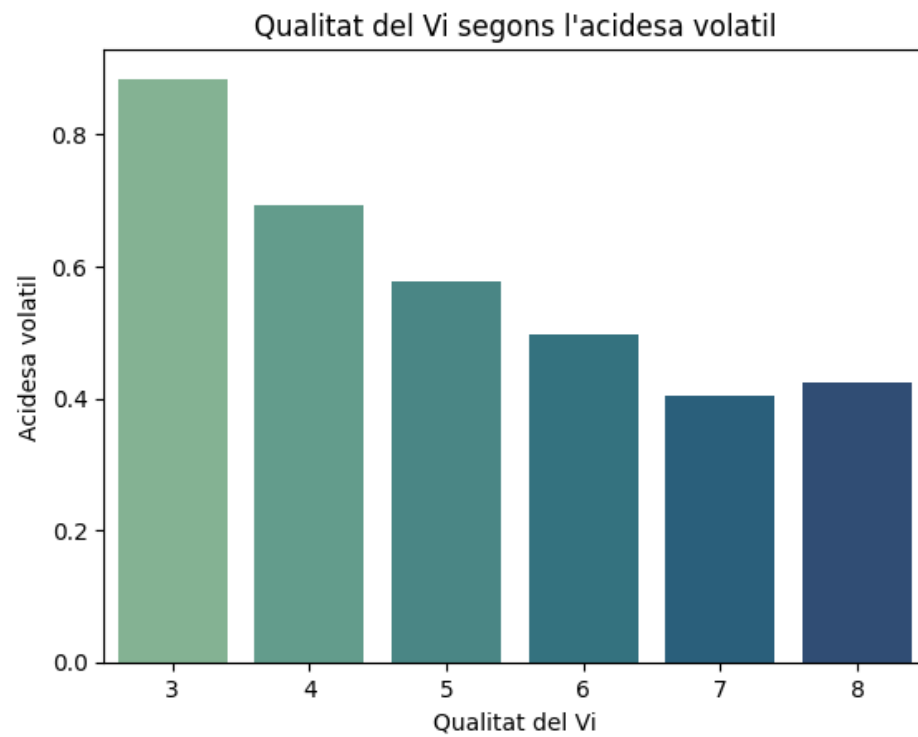
# CORRELACIÓ POSITIVA



Com més alcohol i més sulfats més alta la qualitat del ví!

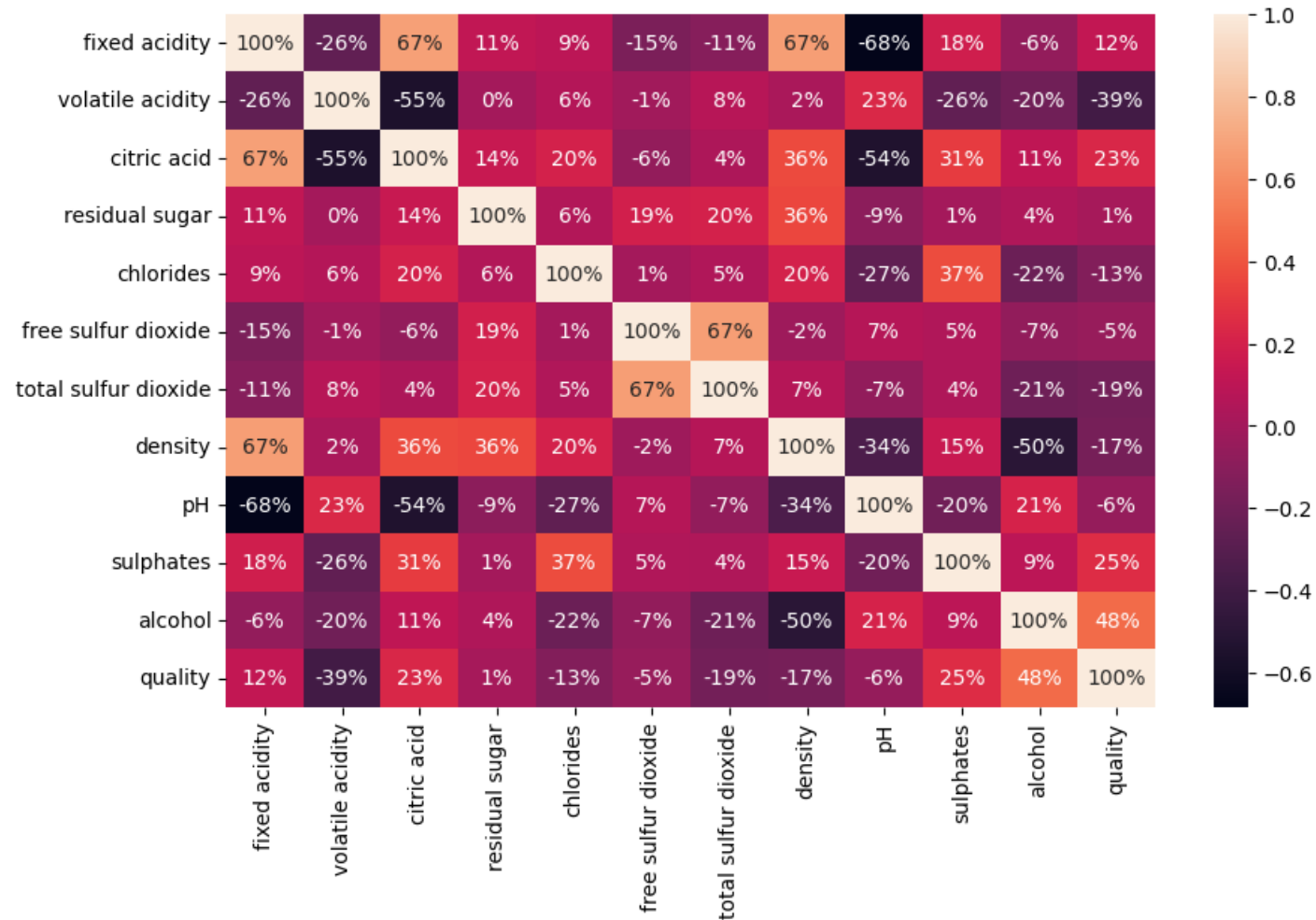


# CORRELACIÓ NEGATIVA

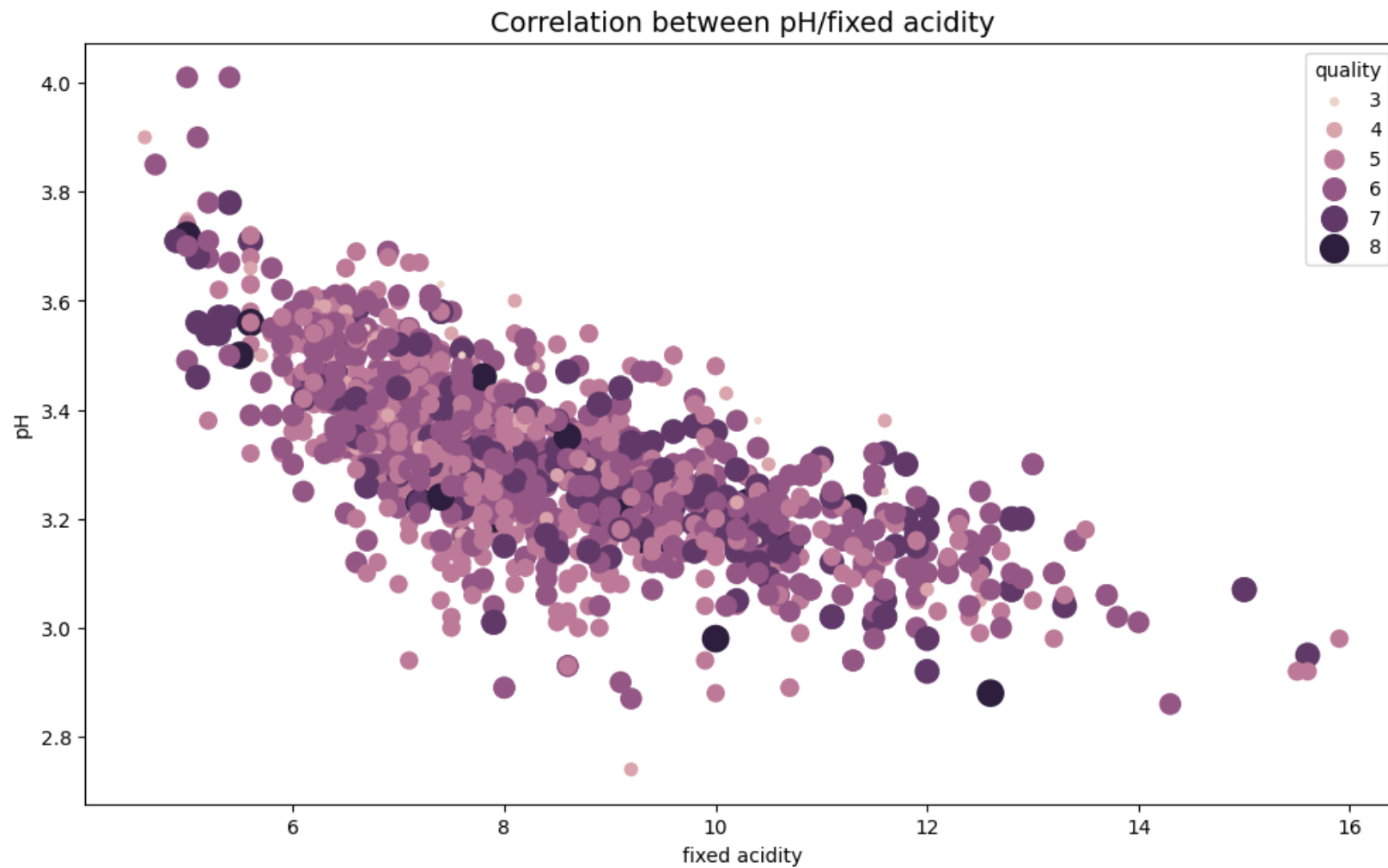


Com més acidesa volatil menys qualitat

# RELACIÓ ENTRE ATRIBUTS



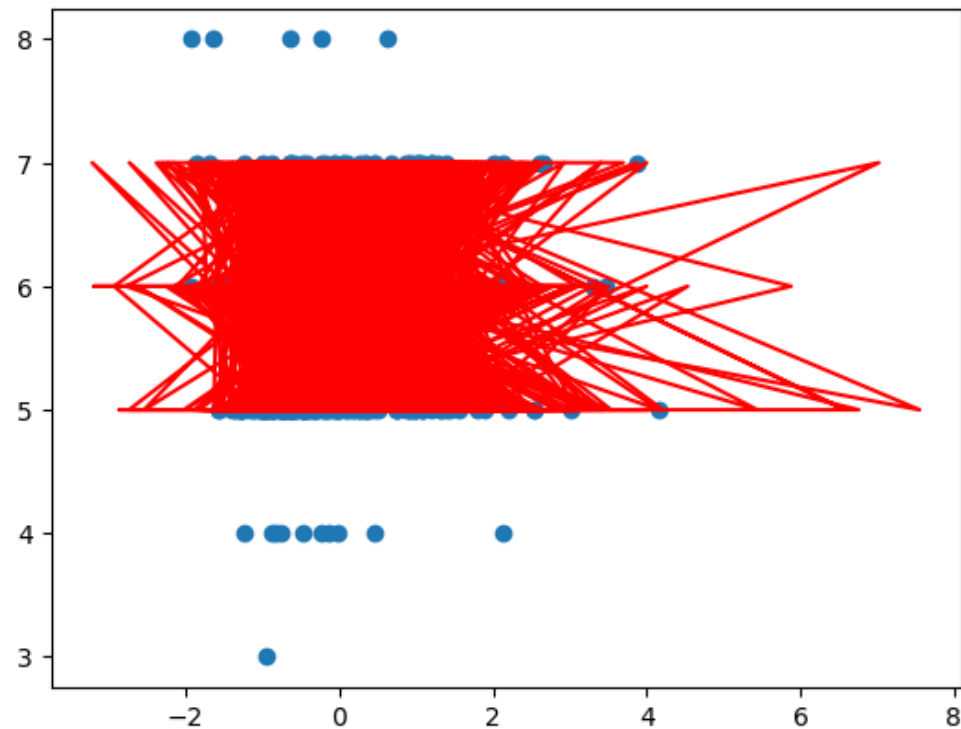
# ALTRES RELACIONES



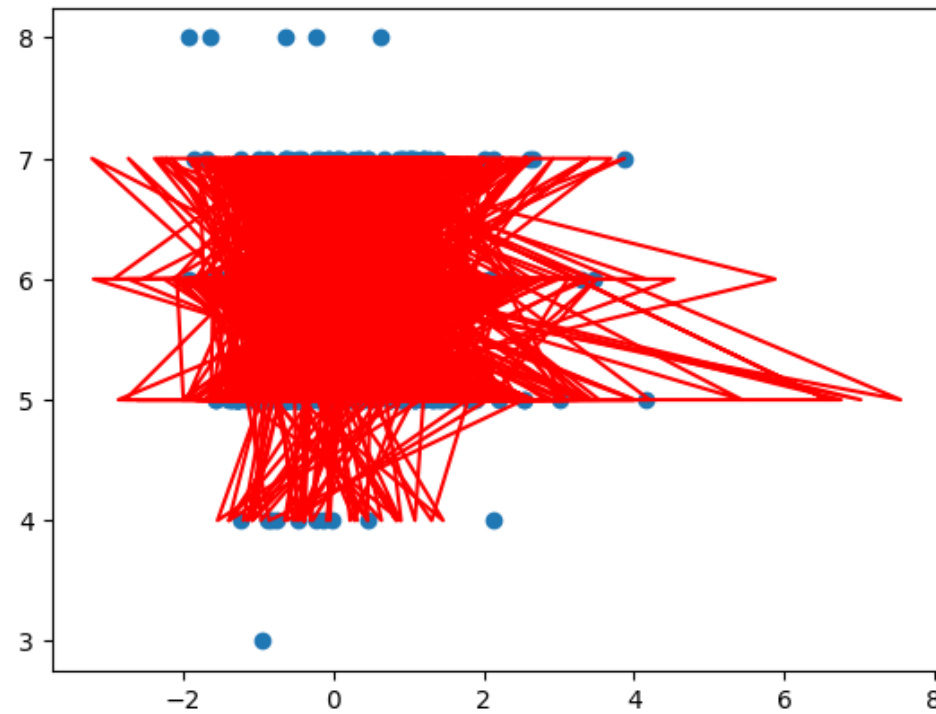
# MODELS

- Support Vector Classification
  - K-Nearest Neighbour
  - Decision Tree
  - Random Forest
- 80% Entrenament
  - 20% Test

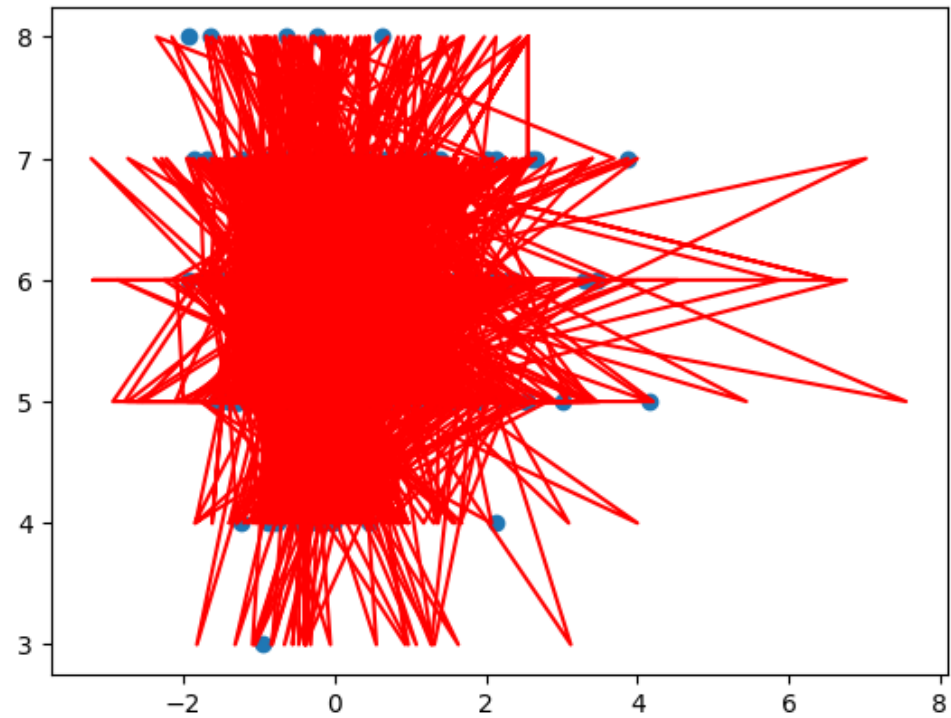
# SUPPORT VECTOR CLASSIFICATION



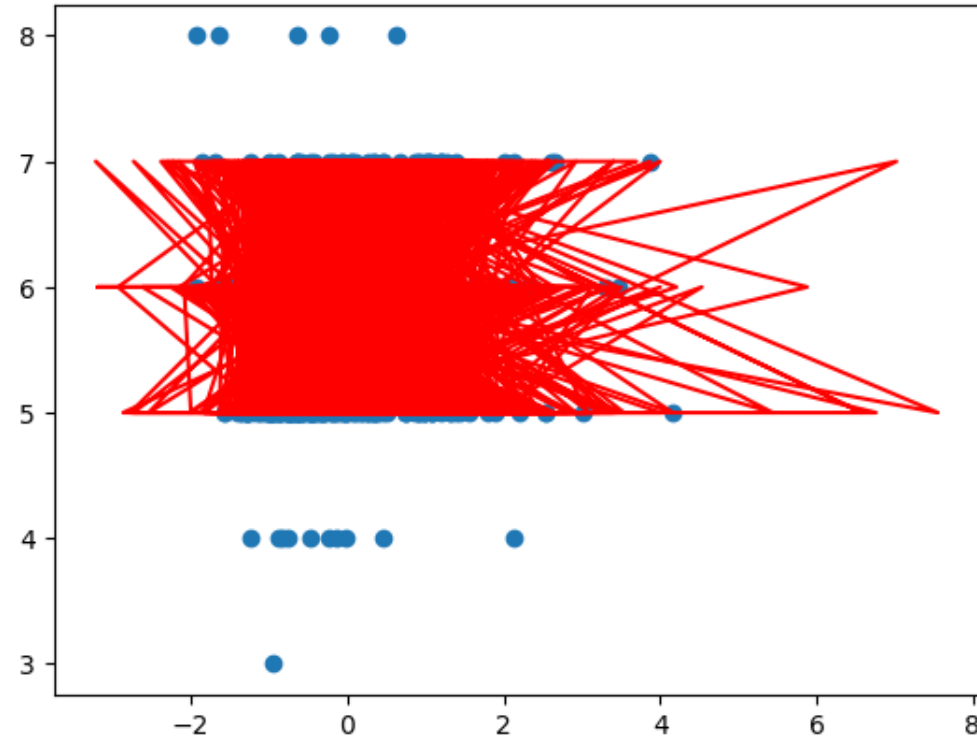
# K-NEAREST NEIGHBOUR



# DECISION TREE



# RANDOM FOREST

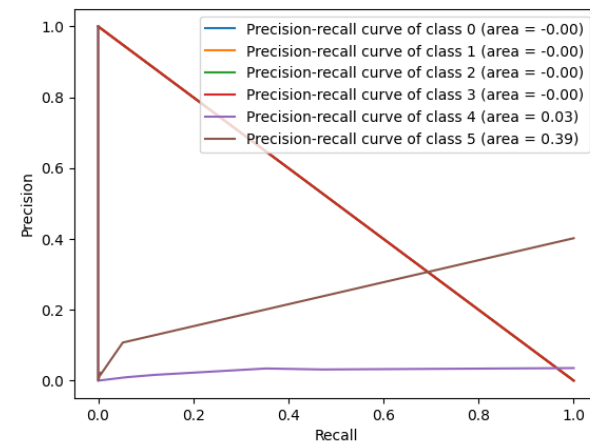
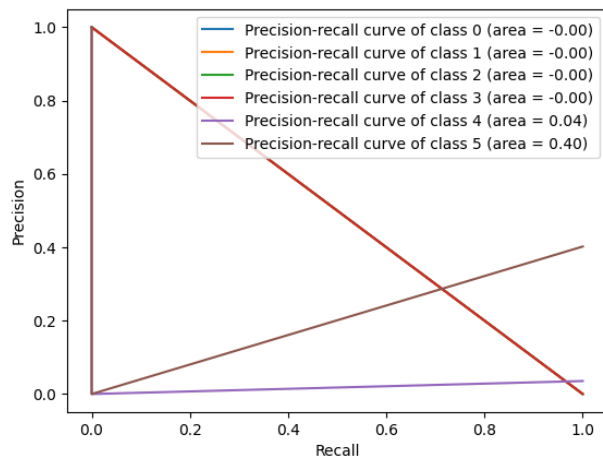
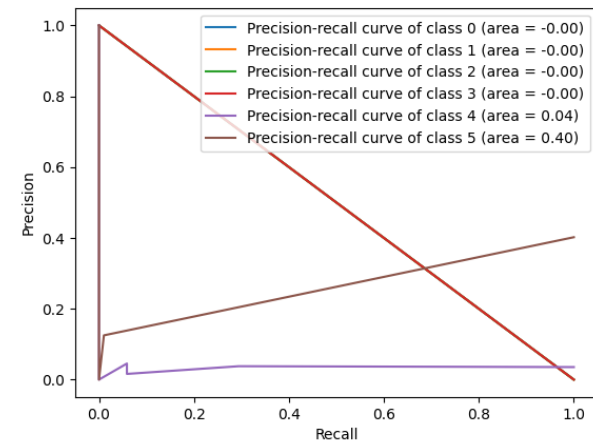
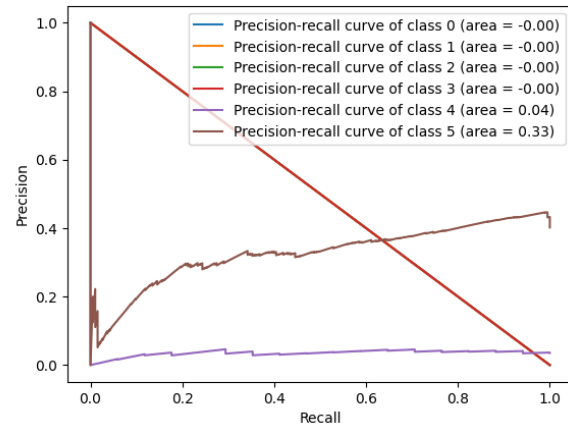




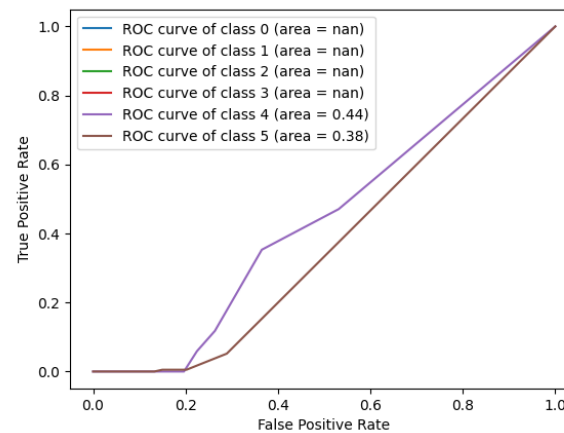
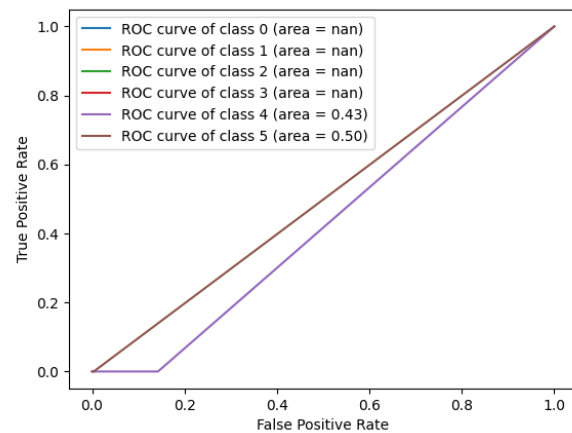
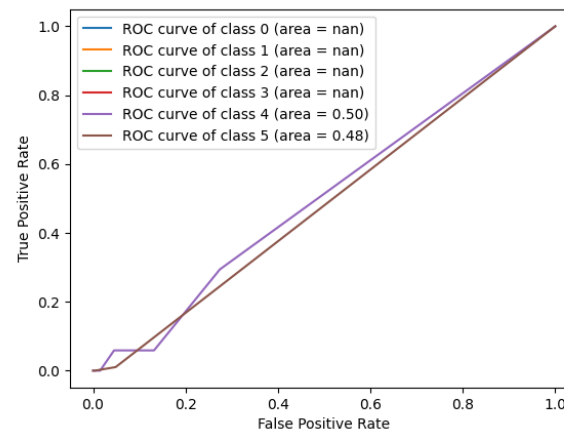
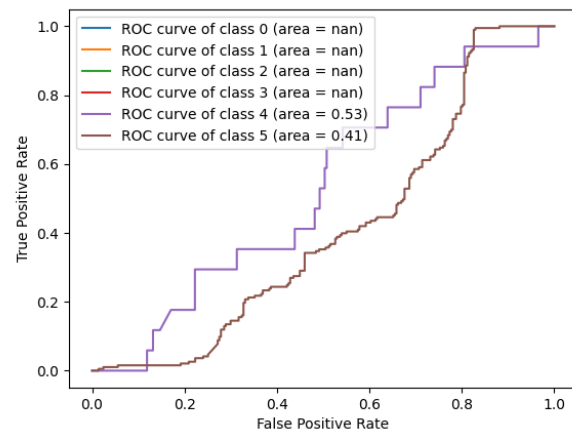
# RESULTATS

Model	Accuracy
SVC	0.719
KNN	0.672
Decision Tree	0.659
Random Forest	0.766

# CURVA PRECISION – RECALL (PR)



# ROC



# CERCA D'HIPERPARÀMETRES

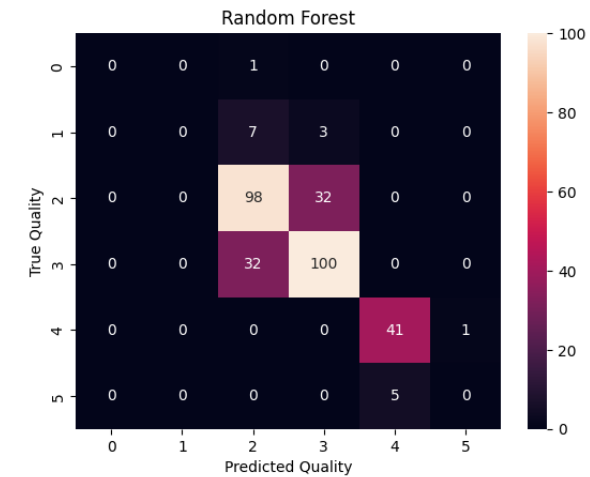
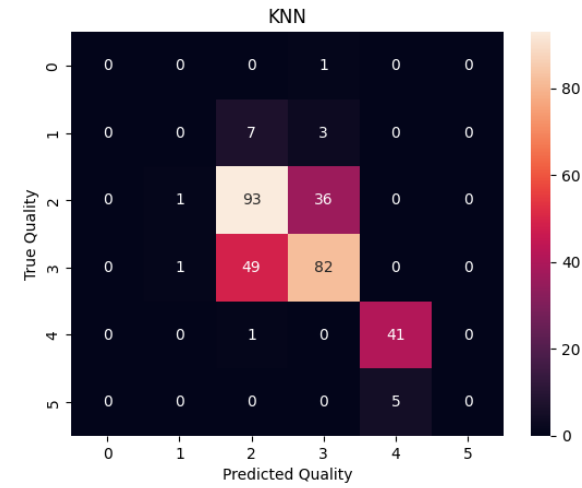
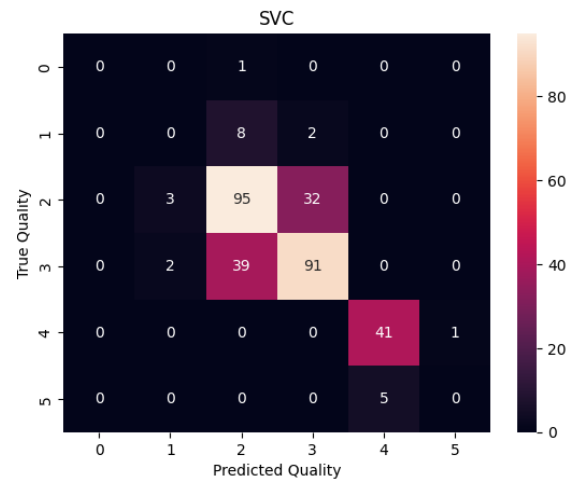
## DEFAULT

Model	Accuracy
SVC	0.719
KNN	0.672
Decision Tree	0.659
Random Forest	0.766

## HIPERPARÀMETRES

Model	Accuracy
SVC	0.728
KNN	0.704
Decision Tree	0.713
Random Forest	0.785

# MATRIUS DE CONFUSIÓ



# CONCLUSIONS

