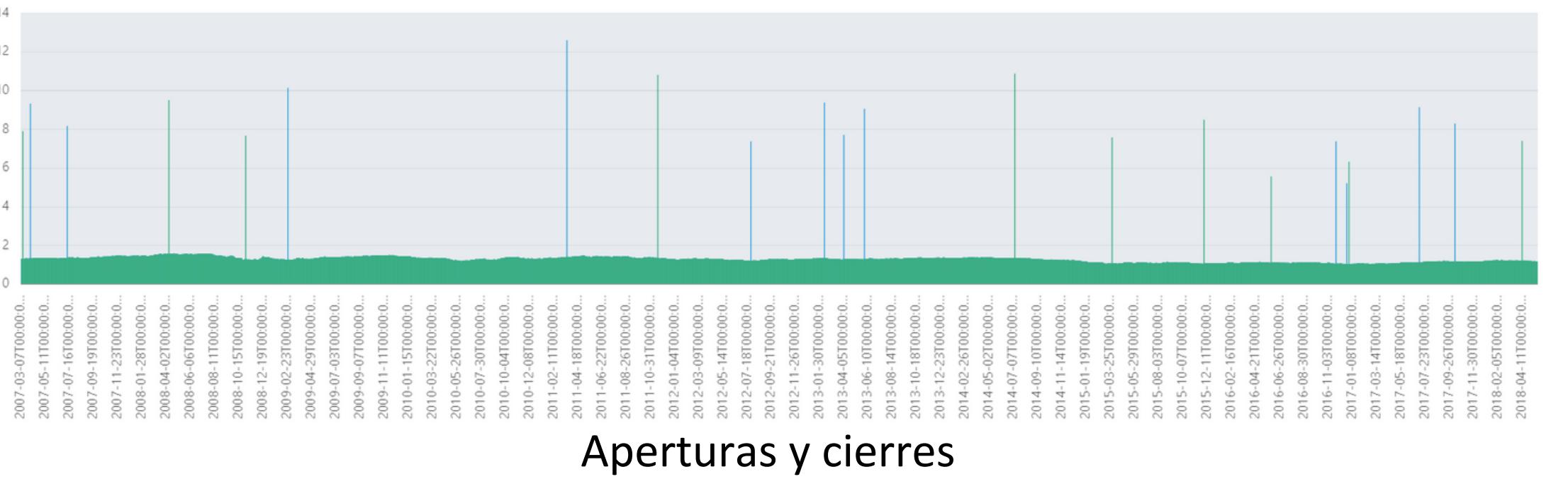


# ENSEÑA\_X\_ORACLE RETO 3

Diedros Squad: Ander Aquerreta y Mikel Aldea

# ANÁLISIS EXPLORATORIO DE LOS DATOS





# PREPROCESAMIENTO DE DATOS

## Areglamos formatos

Time	object	Time	datetime64[ns]
Open	float64	Open	float64
High	float64	High	float64
Low	float64	Low	float64
Close	float64	Close	float64
Volume	float64	Volume	float64
label	int64	label	int64
dtype:	object	dtype:	object

## Valores nulos

Numero de nulos totales del dataset: 208

Time	0
Open	18
High	56
Low	67
Close	39
Volume	28
label	0
dtype:	int64

Time	0
Open	0
High	0
Low	0
Close	0
Volume	0
label	0
dtype:	int64

En nuestro programa hemos decidido eliminarlos por simplificar, pero podríamos haber usado alguna técnica de imputación por valores o imputación por regresión estocástica

# ELIMINACIÓN DE OUTLIERS

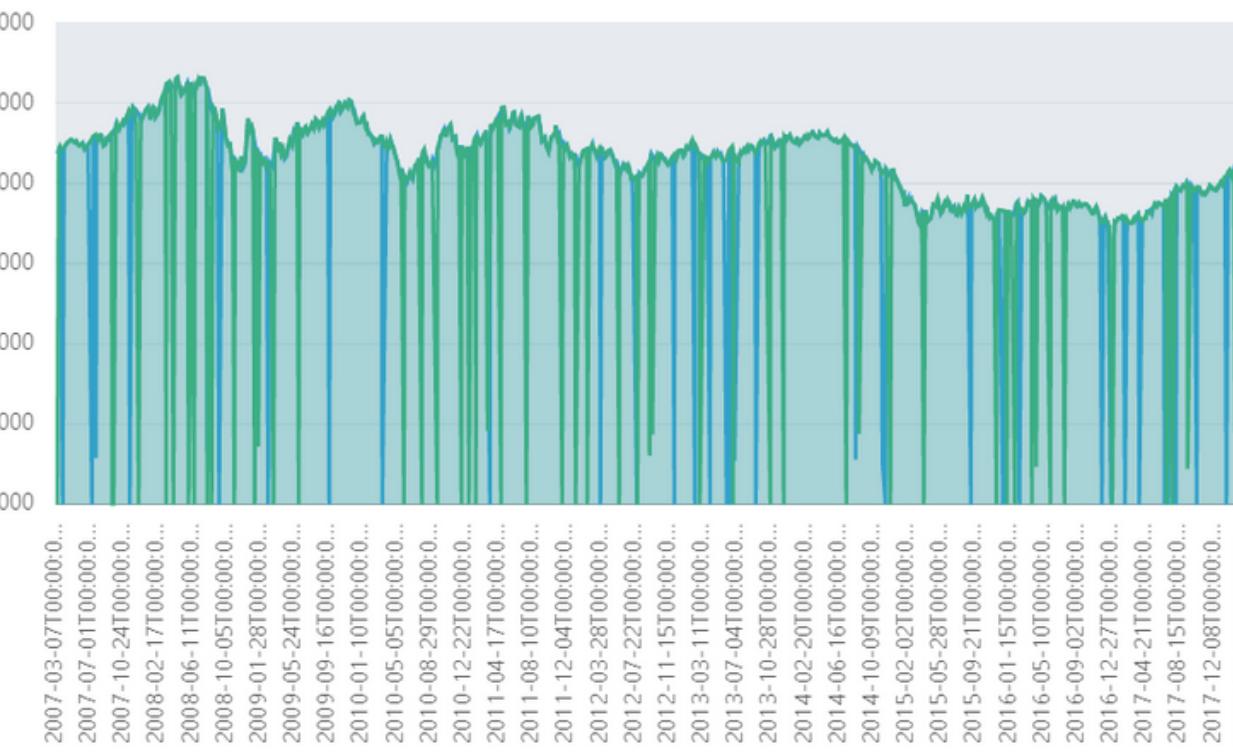
Finalmente nos hemos decidido quedar con los valores que están entre 1 y 2 al ser considerados como valores anómalos según la regla del IQR.

- Rango inter-cuartil
  - $IQR = Q3 - Q1$
- Outlier si
  - $\text{dato} > Q3 + 1.5 * IQR$
  - $\text{dato} < Q1 - 1.5 * IQR$
- Outlier extremo si
  - $\text{dato} > Q3 + 3 * IQR$
  - $\text{dato} < Q1 - 3 * IQR$

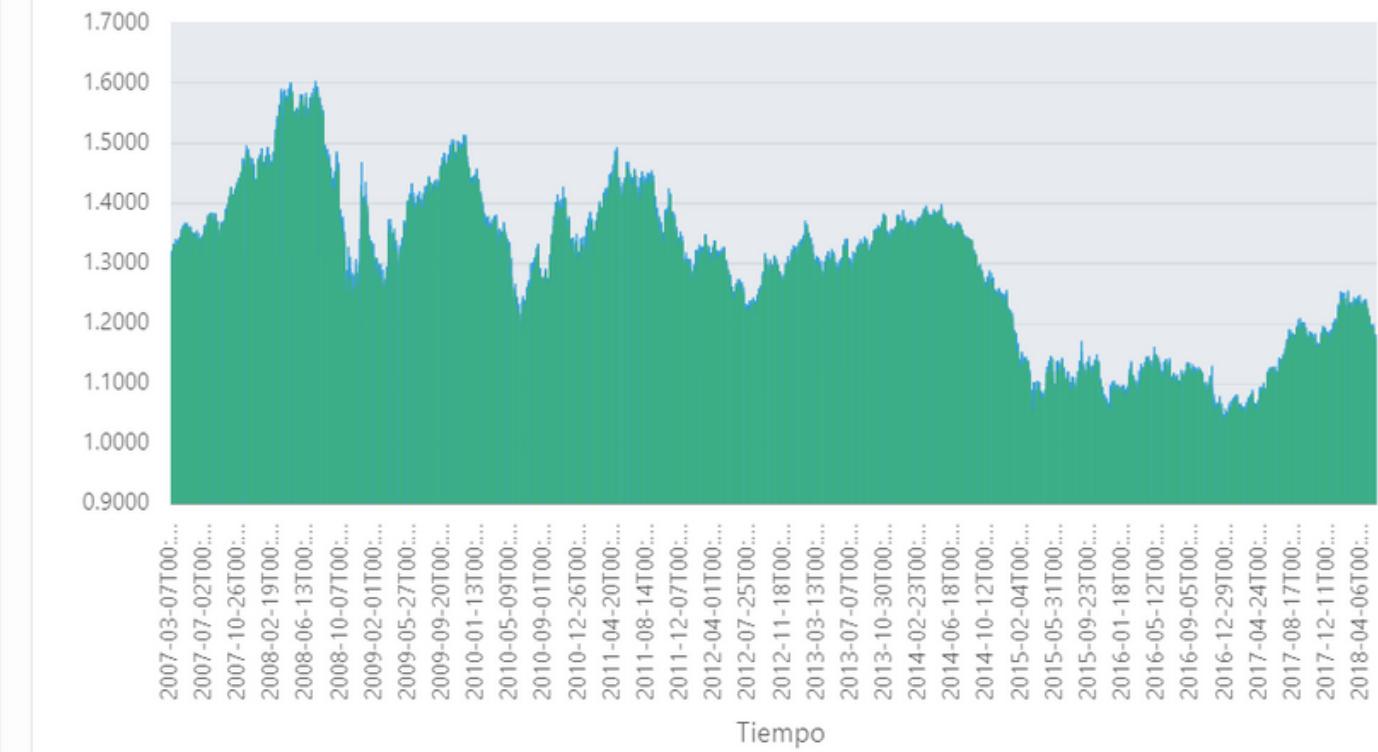


# VISUALIZACIÓN DE LOS DATOS PREPROCESADOS

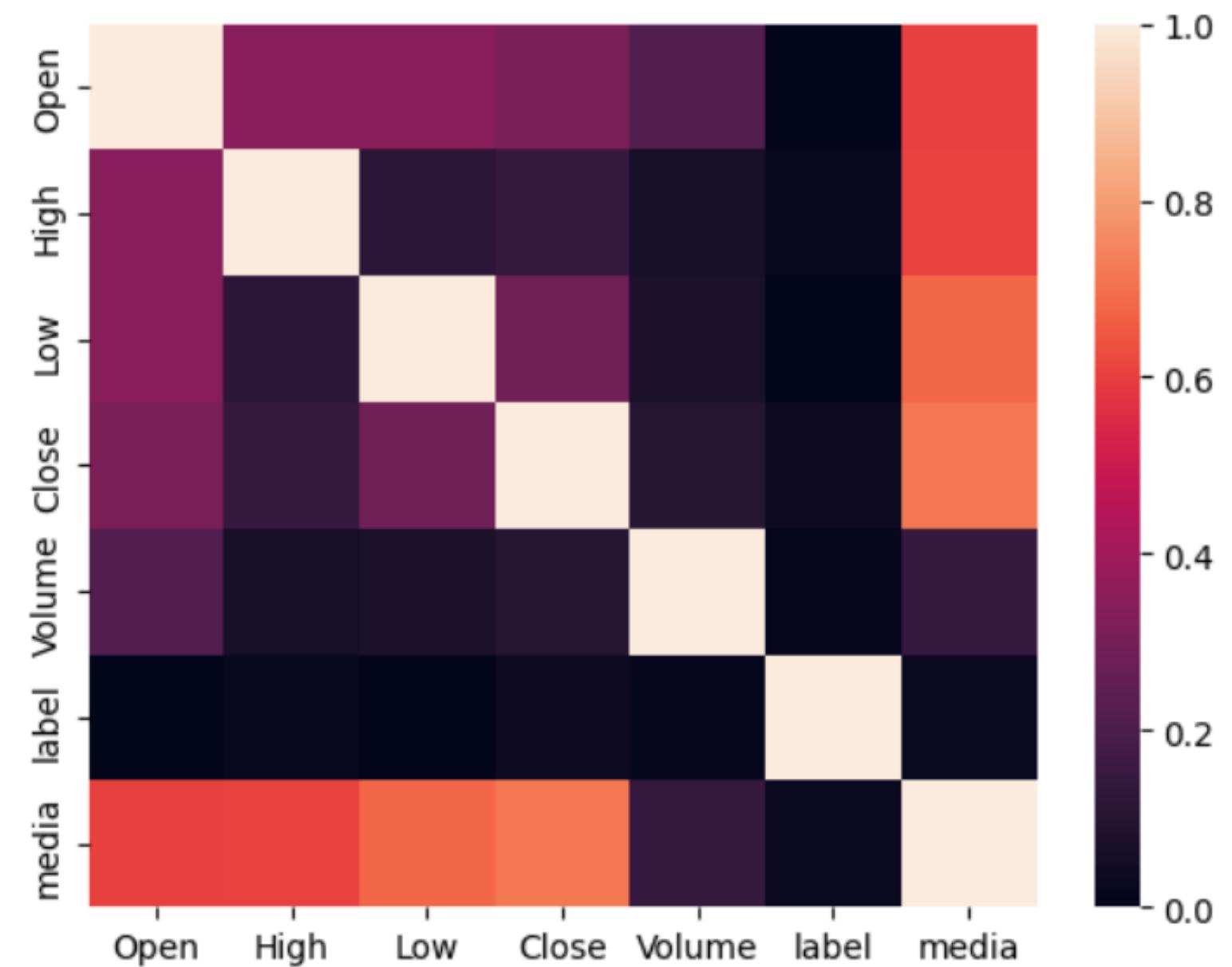
Aperturas/Cierres



Máximos y mínimos

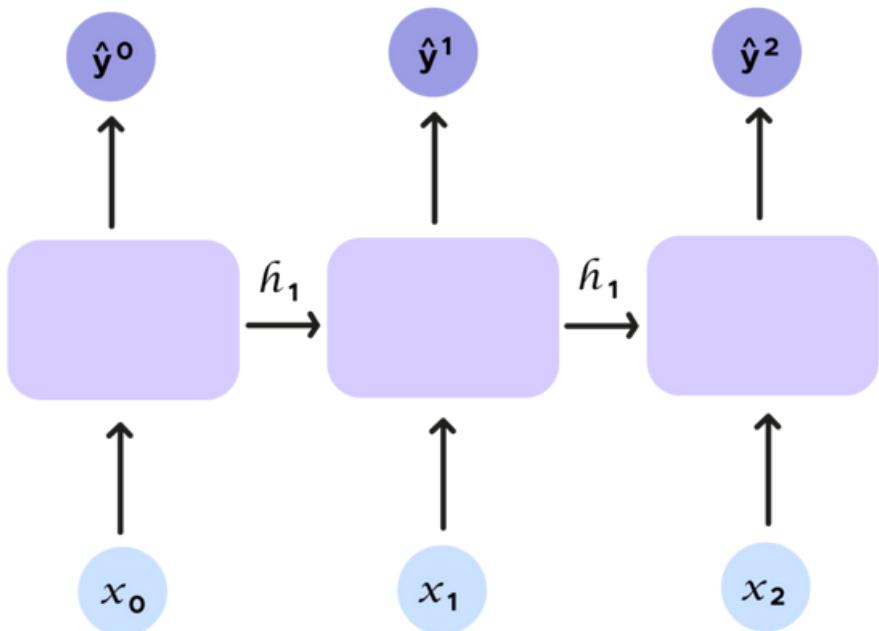


# VISUALIZACIÓN LA CORRELACIÓN ENTRE VARIABLES



# ELECCIÓN DEL MODELO

Para la elección del modelo hay que tener en cuenta que trabajamos con datos secuenciales. Teniendo esto en cuenta podemos utilizar arquitecturas ya creadas como LSTM, GRU o transformers. Sin embargo, para este reto hemos decidido utilizar como modelo una red neuronal recurrente de doble capa, debido a la simplicidad de crearlo y al manejo de tensorflow que tenemos.



# RESULTADOS

Your F1-score is...

0.41383352872215706

GOT IT

Probando con las arquitecturas mencionadas anteriormente obtenemos los mismos resultados en todas.

Resulta que nuestro modelo siempre predice 0, para todas las instancias. Como no hemos tenido demasiado tiempo lo entregaremos así, tal vez se puede deber a errores en los datos.

# THANK YOU

