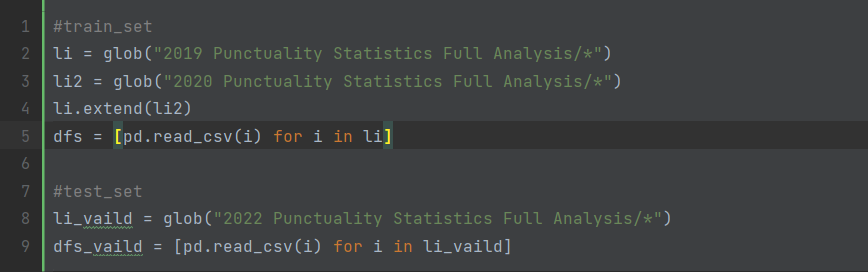
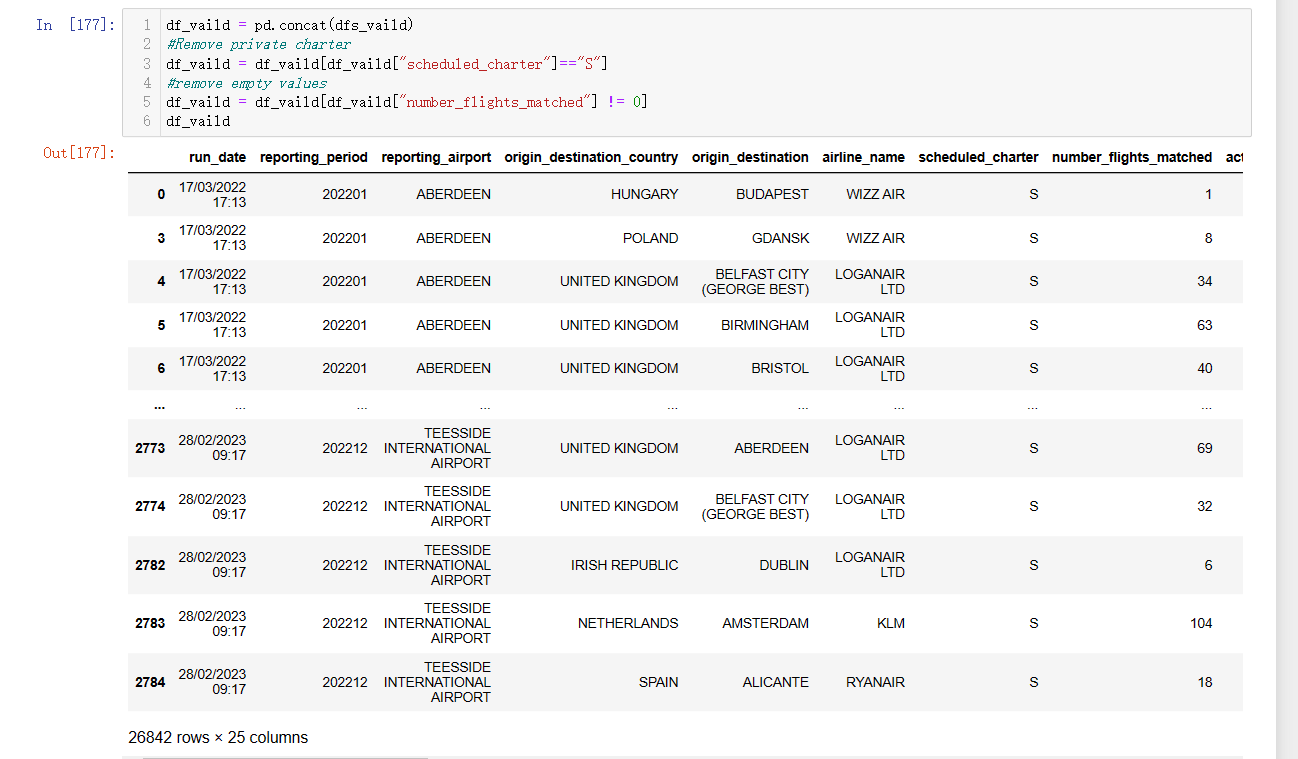
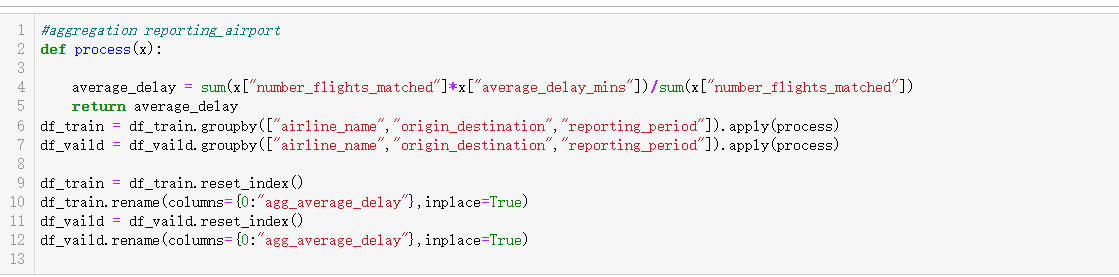
concat 19 and 20 years of data

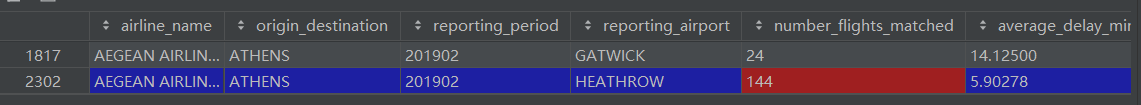


Remove private charter，remove number\_flights\_matched ！=0



aggregation reporting\_airport





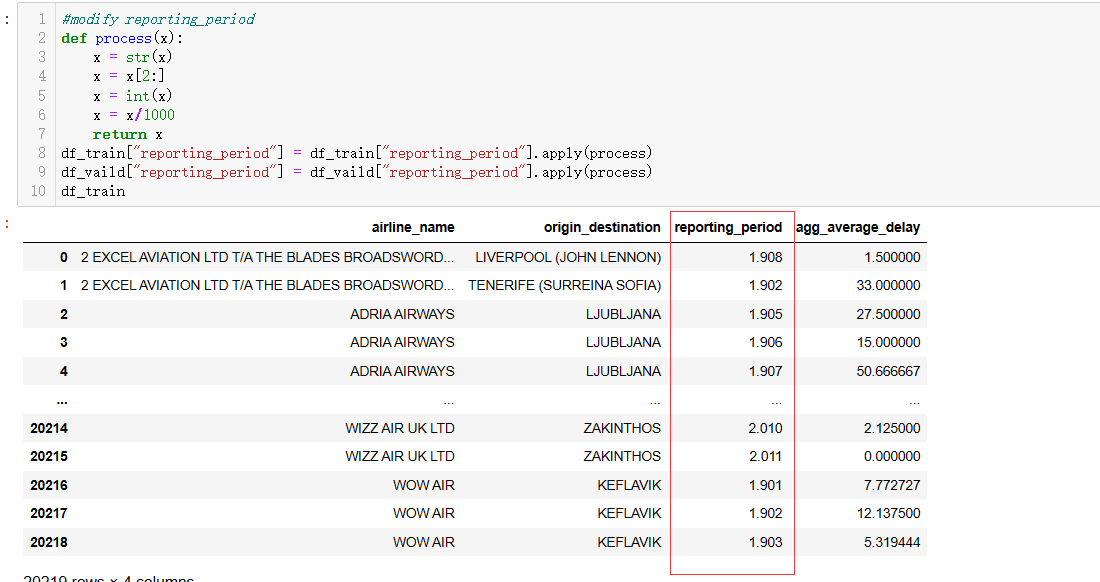
n = The number of aggregated data

Σ(number\_flights\_matched\*average\_delay\_mins)/Σ(number\_flights\_matched)

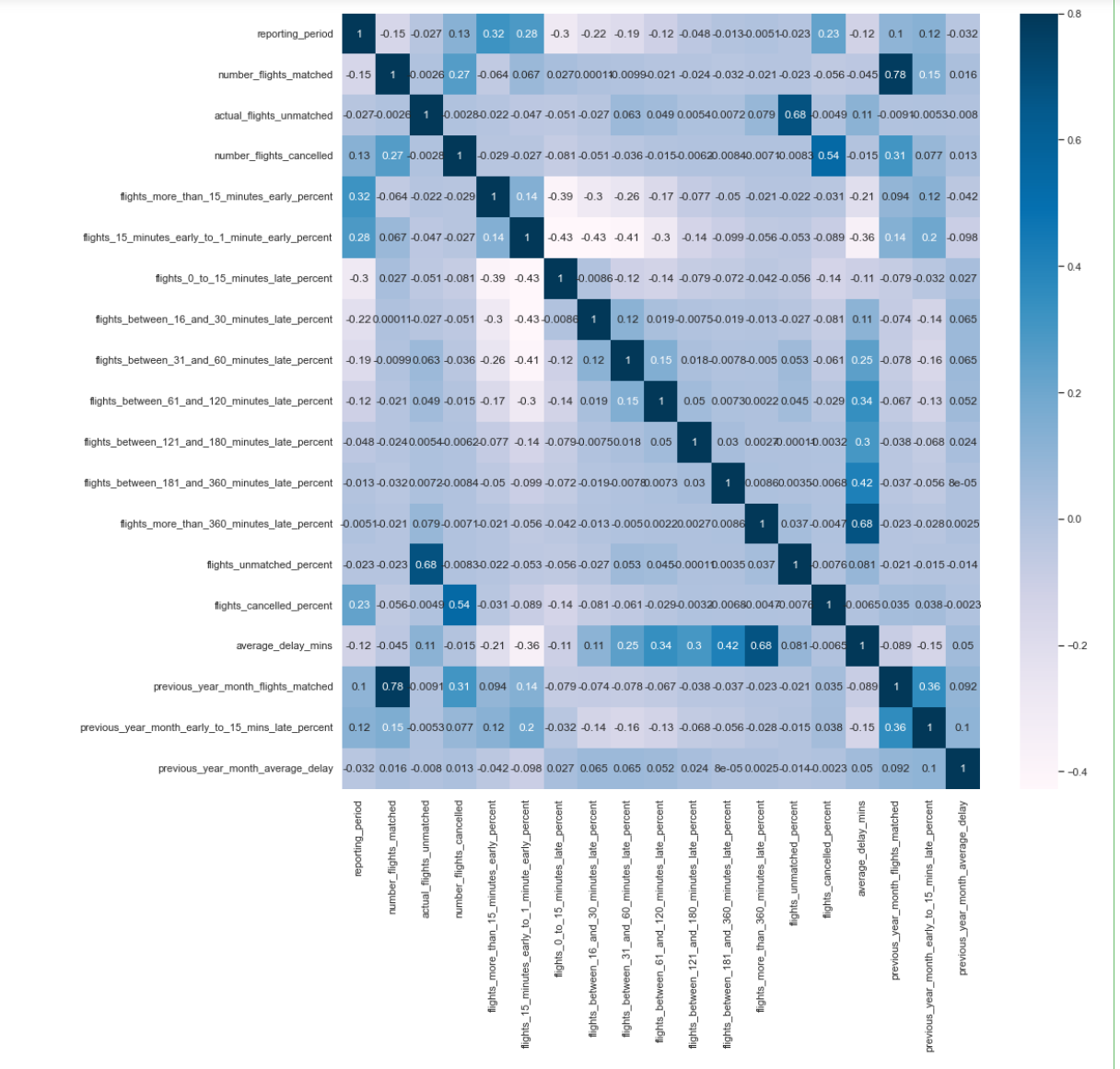
modify reporting\_period

201908 -> 1.908

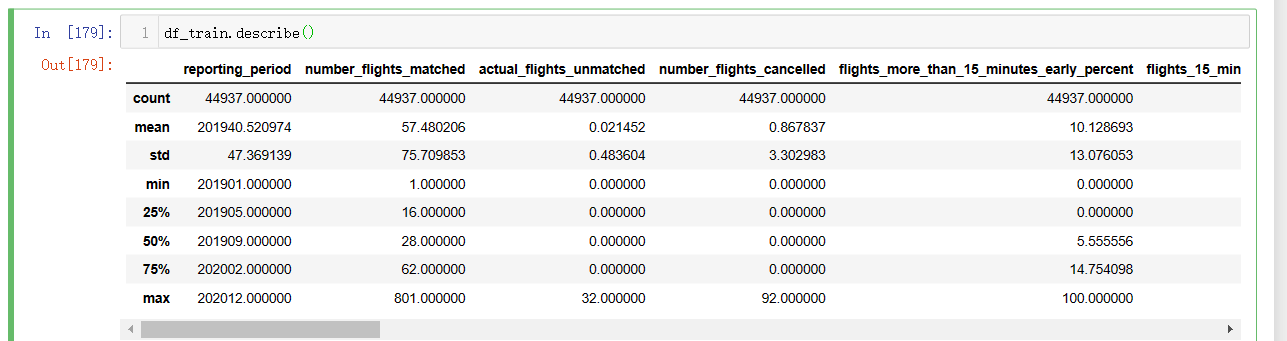
Reduce data to prevent gradient explosion



The larger the value of the heat map, the higher the correlation between the two data



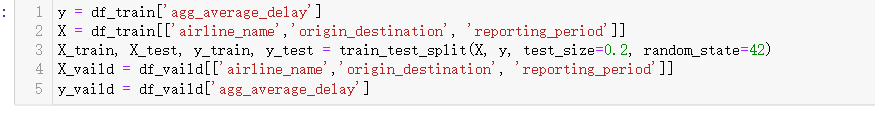
Data details, including mean, variance, min, etc.



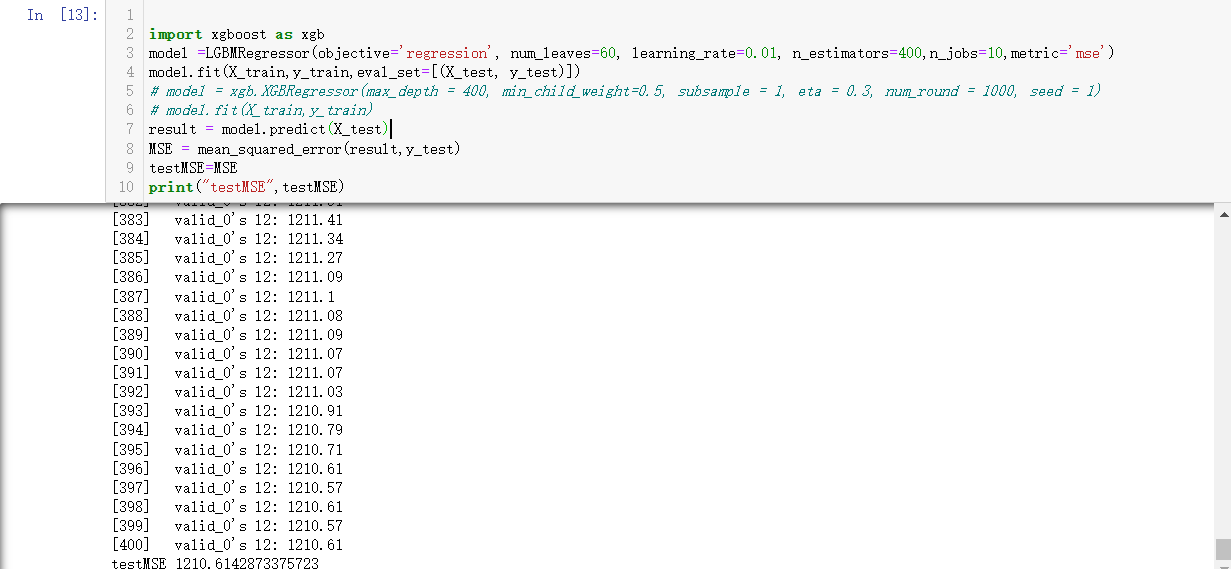
Change data astype



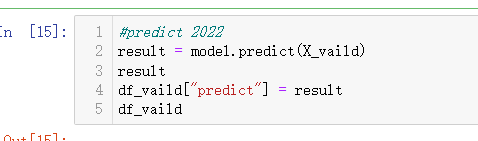
Split dataset



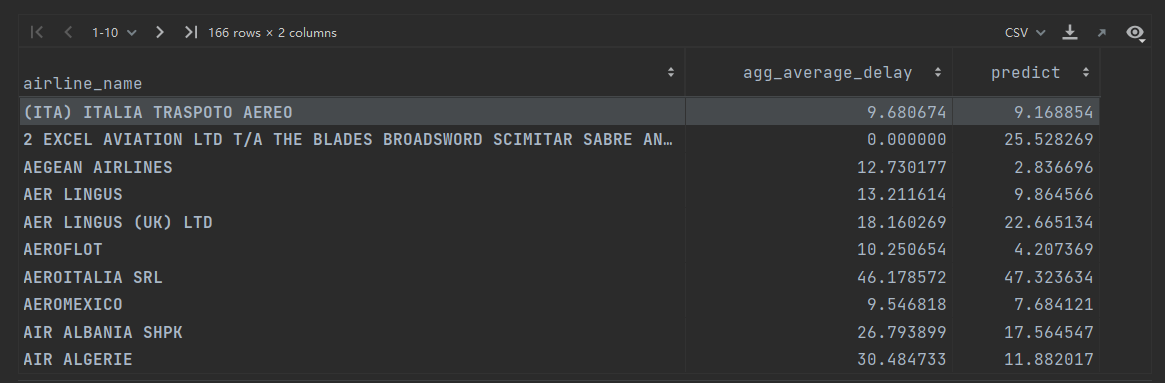
Train model use lgbmregressor



Predict 2022



Groupby airline\_name



Taking the airlines as a group, calculate an average of the delay time of all months. It can be seen that some of the predicted results and the actual values are relatively close, but some of them are far apart, which may be caused by the relatively small data of some airlines.