Background:
The 'Flight Delay Prediction' is to predict claim amount for the future flights in hidden data set. Hence I used a binary model for solving this case, as either we pay $800 or 0. In the approach I chose to predict, is the likelihood of having the claim and times it with $800, and then compared it with mean absolute error and mean squared (Brier error) error. Also since the claim is directly come from delay_time, which means we cannot use this as input variable.

Method:
First I found some external data from online about the airline company and the flight, tried to enrich the dataset and give more insight for the model. Also there is some missing on airline company but we have the flight number, so I extracted back the information by the flight number, followed by feature engineering, such as is the airport is an international airport, length of flight numbers.
Then I split the dataset to train and test by 2/3 and 1/3, and train it with different model, such as decision tree, logistic regression, random forest, XG boosting and lightGBM. At the end I decided to use lightGBM because of 2 major reason. First it doesn't need to change categorical variable to dummy and it preserved more information. Second it provided the best result based on Q1 and Q2 error (lowest in Q1 and Q2 is relatively low as well).
After I decided to use lightGBM, I increased the number of boosting iterations to 2000 to improve the result.

What else can do to improve:
1. Dataset
    a. Check for more airport and airline dataset
    b. Get weather data
2. Feature engineering
    a. Create variable which is public holiday in HK or not.
    b. Grouping of location by bigger geographic location, or by how well the countries developed
    c. Create min, max, mean etc. continues variable on Airport and Airline and subtract, divide the row record to them
3. Modeling
    a. Can try SVM, NN others model
    b. Try Cross validation and maybe have more information on the data. As test validation split would exclude out some data for training.
4. Parameter tuning
    a. using GridSearchCV to turn parameter.
5. Get to understand the data more and find out the feature important.
6. Other approach
    a. I can try to predict the time of delay instead of having claim or not, since the logic is after 3 hours delay or cancel customers can get the pay. And we can use the hours of prediction to imply the claim, which may be another method to hanlde with this question
7. Last stuff is I can develop the whole process as a class and can let the user to run it directly.