Predicting Flight Delays

The data for this project comes from a subset of data that was compiled from the Bureau of Transportation statistics and National Centers for Environmental Information (NOAA) that contains detailed airline, weather, airport and employment information. The goal is to predict whether or not a flight will be delayed by more than 15 minutes (`DEP_DEL15`). The variables included in the dataset are shown in the table below.

Feature	**Description**
MONTH	Month
DAY_OF_WEEK	Day of Week
DEP_DEL15	TARGET Binary of a departure delay over 15 minutes (1 is yes)
DEP_TIME_BLK	Departure time block (generally the hour except 0001–0559)
DISTANCE_GROUP	Distance group to be flown by departing aircraft
SEGMENT_NUMBER	The segment that this tail number is on for the day (a *segment* is a single leg from one airport to another)
CONCURRENT_FLIGHTS	Concurrent flights leaving from the airport in the same departure time block
NUMBER_OF_SEATS	Number of seats on the aircraft
CARRIER_NAME	Carrier
AIRPORT_FLIGHTS_MONTH	Avg Airport Flights per Month
AIRLINE_FLIGHTS_MONTH	Avg Airline Flights per Month
AIRLINE_AIRPORT_FLIGHTS_MONTH	Avg Flights per month for Airline AND Airport
AVG_MONTHLY_PASS_AIRPORT	Avg Passengers for the departing airport for the month
AVG_MONTHLY_PASS_AIRLINE	Avg Passengers for airline for month
FLT_ATTENDANTS_PER_PASS	Flight attendants per passenger for airline
GROUND_SERV_PER_PASS	Ground service employees (service desk) per passenger for airline
PLANE_AGE	Age of departing aircraft
DEPARTING_AIRPORT	Departing Airport
LATITUDE	Latitude of departing airport
LONGITUDE	Longitude of departing airport
PREVIOUS_AIRPORT	Previous airport that aircraft departed from
PRCP	Inches of precipitation for day
SNOW	Inches of snowfall for day
SNWD	Inches of snow on ground for day
TMAX	Max temperature for day
AWND	Max wind speed for day

The data is my Data GitHub repository in the subfolder "Delays". You can load `.csv.zip` files with `pandas.read_csv` without doing anything special (for example, you don't need to decompress the files first).

Project Description:

I would like you to work in groups of size 2 or 3. As a group, find a model that you believe will have the best predictive power on new data. You can clean, transform, perform feature engineering, etc. any way that you want. All members of the group should participate in all of the steps of the project.

Serialize your trained model into a Pickle file:

```
model = MyClassifier()
model.fit(X_train, y_train)

import pickle
with open('my_model.pkl', 'wb') as f:
    pickle.dump(model, f)
```

If you do any data processing outside of the model (including renaming variables), then write a function that will process the data. The goal is for me to use your model directly to make predictions on a similar dataset. If all processing steps are performed inside a Pipeline that includes your final estimator, then just your serialized model is sufficent.

Use your model to predict the probability of a flight delay for the test data. Submit your predicted probabilities in a csv file called `predictions.csv`. Your csv file should one column called "predictions".

Project deliverables (Group deliverables upload to the assignment GitHub repository):

1. (Group) Executive Summary: A concise 1-2 page report that summarizes:
    - **the project objectives**
    - **methodology:** why/how you chose your model and why you think it has the best predictive power
    - **results:** your final model, including hyperparameters, performance metrics, and variables that are important for the prediction (including *how* they affect the prediction. For example, if the variable "SNOW" turns out to be important, say if higher values of SNOW are associated with higher or lower probabilities of delay)
    - **conclusions**
2. (Group) `predictions.csv` file. Be sure that your file contains only one column named "predictions" and that your values are in the same order as the test data.
3. (Group) Your serialized trained model in a file called `my_model.pkl`. If applicable, also your processing function in a `.py` file.
4. (**Individual – submit to LearningSuite NOT GitHub**) A 1-2 paragraph summary of your individual contributions. Include specific tasks that you completed and any challenges faced.