ISM 370

Module 7 Project

Requirements

Project Name	Binary Classification – Flight Delays
Project Due Date	Sunday by 11:59pm
Requirements	
In this project, you will perform binary classification on an airline flight delays dataset. Requirements: 1. Read in the airlines_subset.csv file into your Python project. This file includes data on flights arriving in PHX only and whether they were delayed or not. The column headings provide sufficient descriptions of the data. 2. Generate descriptive statistics of the numerical data. Print an interpretation and comment on the mean value for 'delayed'. 3. For this project, use 'flight_length' as your x variable and 'delayed' as your y variable. You do not need to consider any of the other variables or columns—they are not significant predictors of delayed. 4. Use sklearn and logistic regression to develop predicted probabilities for delayed flights and insert these in a new column in your Dataframe. You should use 'flight_length' as your x and 'delayed' as your y. 4. Print the 10 largest predicted probabilities for delayed flights. Print an interpretation and comment on what you see. 5. Calculate and print a confusion matrix. 6. Calculate and print the prediction accuracy, based on the confusion matrix. Print an interpretation and comment on what you learned.	
Completion	
Only successful completion of this program and all the requirements will result in a high marks. You are welcome to add additional functionality and to utilize your creativity in making the program even better.	
Deliverable	
Submit your Python file to Canvas.	