SAS EM Certification Project
Analysis of airport On-Time Performance

To be done individually.
Read carefully all the instructions.

Flight delays are obviously frustrating to air travelers and costly to airline companies. On-Time performance of airline schedule is key factor in maintaining customer satisfaction. Even if flight delay is complex to explain and often occurs because of a combination of several factors, some data mining tools may help finding some recurrent patterns. A better understanding of theses causes and help decision-makers to improve the on-time performance and thus save huge amounts of money. In this project, you will study the factors influencing arrival delay at your assigned destination airport in order to predict future delays under specific conditions. The project is divided in two tasks.

Task 1 : Data collection and preparation

Deadline : March, 15th 2020

In this project, you will focus on the analysis of factors influencing the arrival delay at a specific airport on the basis of several data sources. You will be asked to gather and process the data sets needed for this project using the data preparation software of your choice. You will need to choose the relevant time period to gather your data from. Here are the data sources you can use as a starting point :

1. **Airline On Time Performance Data**
   This database contains scheduled and actual departure and arrival times reported by certified U.S. air carriers that account for at least one percent of domestic scheduled passenger revenues. The data is collected by the Office of Airline Information, Bureau of Transportation Statistics (BTS). It provides such additional items as departure and arrival delays, origin and destination airports, flight numbers, scheduled and actual departure and arrival times, and cancelled or diverted flights, taxi-out and taxi-in times, air time, and non-stop distance.

2. **Carrier information**
   Information on each carrier.

3. **Aircraft information**
   Information on each aircraft.

You can (and are encouraged to) add any other relevant data sests you might find useful to your project. The instructions as well as the links to download the starting data sets can be found in the Appendix. You are asked to select the variables that seem relevant to you. The description of all the variables are given on the corresponding websites. As data analyst, it is your first job to understand the meaning of each variable and the units in which they are expressed in order to draw some appropriate conclusions.

After having collected the desired information for the period under consideration, your first task consists in preparing the On-Time Performance data file in order to analyze it later. Here are some (rough) preparation steps you can follow :

(A) The analysis will only focus on non-canceled and non-diverted flights with the assigned airports as origin or destination. Filter the monthly data files and then, gather the desired data for the entire period in a single file.

(B) In order to join the tables correctly,you will need to transform the dates accordingly. You might also need to create new observation variables for the arrival and departure times.

(C) Filter the datasets according to your assigned destination airport.

(D) Group all the data sets (prepared On-Time Performance with the assigned airport as destination, weather data, carrier data, aircraft data) in a single file.

Task 2 : SAS SEM - Predictive modeling

Deadline: May, 8th 2020

For this part of the project, you will need to import your previously generated data set into SAS EM in order to analyze it.

Using SAS EM, can you predict if a flight will be late? How to do it? How confident are you in your predictions and why? Could you say why a flight might be late?

The outcome of this task must be a report explaining the methods you used, the decisions you had to take and why, the choice of parameters and the interpretation of the results you obtained. The diagram of your prediction project and your results must be reproduced in this report. Conversely, the report CANNOT just be a copy/paste of values from SAS EM; interpretation is the key.

Appendix

In this appendix, you will find all the information to download the mentioned data sets.

1. **Airline On Time Performance Data**
   http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time

2. **Carrier information**
   https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=312&DB_Short_Name=Air%20Carrier%20Financial

3. **Aircraft information**
   https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=314&DB_Short_Name=Air%20Carrier%20Financial

4. **Another database...**
   ...