Business Analytics Project
Analysis of airport On-Time Performance

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 these causes would 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 destination when you take off from your assigned destination airport. The project is divided in two tasks.

Task 1 : Data collection and preparation

Each group received a data set for one airport and one specific period. It is compulsory to use this file as the main one. However, you may - and are encouraged to - try to improve the quality of your predictions by collecting additional information (weather, company information, aircraft information...). Note that the dataset provided contains real data collected from a US official database. You may have a look at http://www.transtats.bts.gov for more (but you cannot collect data for another period or additional features within the flight table 'Airline On Time Performance Data').

You are asked to select the variables that seem relevant to you. The description of all the variables are given on the US website. 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) 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.

(C) Follow the steps that were recommended during the lectures. Be extremely careful with the data types.

Data preparation will be done with Excel. All the operations performed must be clearly explained in the report. We also ask you to create a small dashboard with Power BI. It is up to you to decide what should be presented and how to do it. Obvioulsy, it must be related to the main question: 'Analysis of airport On-Time Performance'

Task 2 : SAS SEM - Predictive modeling

Deadline: May, 23rd 2022

Can you predict **before departure** if a flight will be late (at least 15')? How to do it? How confident are you in your predictions and why? Do you predict as easily delayed or on-time flights? Could you say why a flight might be late?
Think twice! Do not conclude too fast. If a reason for the delays is too obvious, dig deeper and be clever.

The outcome of this task must be a report (max 15 pages) 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 python; **interpretation is the key**.