**DESCRIPTION**

You will work in groups to get insights from the Flights dataset, using your Spark skills. Every group must arrange the analysis and conclusions into a Jupyter notebook which mixes code and text. Each group will be assigned a particular topic and must answer a set of concrete business questions based on the data. Each group will need to download or build an additional dataset, for a purpose described in your notebook.

**GOAL**

You’re supposed to put everything you’ve learned during the course in practice, work as a team and share your conclusions. In addition to the class slides, you are supposed to use the official API documentation in case you need to use other functions/methods.

**CONTENTS ORGANIZATION**

Organize your notebook as follows:

* Describe steps followed to get the data for the analysis and issues found along the way.
* PySpark code answering your questions.
* Business/non-technical conclusions (insights)
* Add a brief intro on why your topic might be relevant to the air transportation industry

**DELIVERABLES**

You should deliver the following before deadline:

* Jupyter Notebook with the analysis and explanations. It has to be downloaded from the OVA to your PC with all executions and then uploaded to the Campus Portal.
* Additional data set used. Provide the URL to the external files (e.g. CSV) you’ve used for the assignment, or upload the file if you’ve built it yourselves and is moderately small (as a guideline, not larger than 10 MB)

**EVALUATION CRITERIA**

Assignment’s score will rely on the following evaluation points, all equally important:

* Code correctness.- simple code that works!
* Insights.- must be non-technical conclusions based on the result of the performed analysis.
* True teamwork.- the most important criteria so, please please please, work as a team.

**DEADLINE**

Everything has to be provided via the Campus Portal before due date: March13th, 2020