# **Massive Data Processing**

## Big Data Project – Bets Exploring

Students: Alex Ghenghiu and David Sánchez Marín

In this document we explain the structure of the code and other files result of the development of our Big Data Project: Bets Exploring

The structure of folders and files is:

* Data
  + Raw: Originals files downloaded from web page.
    - Main: Files from main competitions
    - Extra: Files from extra competitions
  + Interim: Result of filtering and unification of originals files. Base to Data analysys.
  + Processed: Results from the data analysis. Not used.
* Scripts: Python scripts, Jupyter notes and complementary files.
  + 1\_download.py: Python script that take care of downloading files from web page, of validating format and of saving information in Raw folder.
  + 2\_initialFileAnalysis.ipynb: Jupyter notebook were analyse the downloaded files and the fields we must use in the next step.
  + 3\_initialAnalysisEngland.ipynb: Jupyter notebook where we can take a first look to content of some original files. We show fields and rang of values
  + 4\_unifyingFiles.ipynb: Jupyter notebook to prepare unification of original files
  + 4\_unifyingFiles.py: Python script to unify contents from main original files. It creates the final file where it will done the analysis.
  + 5\_matchAnalysis.ipynb: Quantification of main variables and analysis of team’s results
  + 5\_betStatistics.ipynb: Analysis of bets houses quotas and correlation with match results.
  + File\_index.csv: Index of countries, competitions and season to download
  + Notes.txt: Description of fields from original files