**Intitulé du cours**

|  |  |
| --- | --- |
| Course title – Intitulé du cours | Statistical Softwares for data scientists (R Python, SAS) |
| Level / Semester – Niveau /semestre | M1 / S1 |
| School – Composante | Ecole d'Economie de Toulouse |
| Teacher – Enseignant responsable | Benoit Gaudou (R, Python) |
| Other teacher(s) – Autre(s) enseignant(s) | Kevin Godin-Dubois |
| Other teacher(s) – Autre(s) enseignant(s) |  |
| Other teacher(s) – Autre(s) enseignant(s) |  |
| Other teacher(s) – Autre(s) enseignant(s) |  |
| Other teacher(s) – Autre(s) enseignant(s) |  |
| Lecture Hours – Volume Horaire CM | 0 |
| TA Hours – Volume horaire TD | 0 |
| TP Hours – Volume horaire TP | 36h |
| Course Language – Langue du cours | English (international track) French (standard track) |
| TA and/or TP Language – Langue des TD et/ou TP | Anglais |

**Teaching staff contacts – Coordonnées de l’équipe pédagogique :**

Benoit Gaudou (R and Python)(benoit.gaudou@ut-capitole.fr, MF205)

Teaching assistant:

Kevin Godin-Dubois (R and Python)(Kevin.Dubois@irit.fr )

**Course Objectives – Objectifs du cours :**

The objective is to give students in economics bases in three reference softwares and programming language necessary for data scientists: R, Python and SAS. Through practical sessions, students should be able to manipulate datasets (import, clean, compute indicators, and visualise them).

**Prerequisites – Pré requis :**

Basis of descriptive statistics.

**Practical information about the sessions – Modalités pratiques de gestion du cours :**

Practical tutorial: 36 hours for R and Python (3h every week). 5 first tutorials are dedicated to practical introduction to R, programming in R and data importation, manipulation and graphical visualisation. The 6th tutorial is dedicated to the evaluation related to the R part. The next 5 tutorials are dedicated to practical introduction to thePython language for programming, data manipulation and web scrapping. The last tutorial (the 12th one) is dedicated to the Python evaluation.

Personal computer allowed

**Grading system – Modalités d’évaluation :**

**R and Python:** Midterm evaluation (50%) and final evaluation (50%). Each evaluation will be a practical evaluation on computer.

**Bibliography/references – Bibliographie/références :**

There is no compulsory textbooks. Below are some references we would recommend:

An Introduction to R  
http://cran.r-project.org/doc/manuals/R-intro.html

R pour les débutants - Emmanuel Paradis   
(https://cran.r-project.org/doc/contrib/Paradis-rdebuts\_fr.pdf)

Python in a nutshell  
Alex Martelli, O'Reilly Media, 2017.

**Session planning – Planification des séances**

**R and Python**

Session 1: (R) Introduction to R and R language. Importation of datasets.

Session 2: (R) Indexing and filtering in R vectors and data frame

Session 3: (R) Programing in R (loops, conditionals, functions)

Session 4: (R) The family of apply functions

Session 5: (R) Data visualisation

Session 6: (R) Evaluation about the R part

Session 7: (Python) Introduction to Python andPython language. Importation of datasets. Syntax of Python language (loops, conditionals, functions)

Session 8: (Python) Introduction to Numpy library

Session 9: (Python) Introduction to Pandas library

Session 10: (Python) Introduction to Matplotlib library

Session 11: (Python) Webscrapping withPython

Session 12: (Python) Evaluation about the Python part

**Distance learning – *Enseignement à distance :***

*Distance learning can be provided when necessary by implementing:*

* *Interactive virtual classrooms*
* *Recorded lectures (videos)*
* *MCQ tests and other online exercises / assignments*
* *Remote (online) tutorials (classes)*
* *Chatrooms*

*En cas de nécessité, un enseignement à distance sera assuré en mobilisant:*

* *Classe en ligne interactive*
* *Vidéo enregistrée de la présentation du matériel pédagogique*
* *QCM et exercices en ligne*
* *TP/TD à distance*
* *Forum...*