

Data Analytics

Course title – Intitulé du cours	Data Analytics
Level / Semester – Niveau / semestre	M2 / S2
School – Composante	Ecole d'Economie de Toulouse
Teacher – Enseignant responsable	ARCHIMBAUD Aurore
Other teacher(s) – Autre(s) enseignant(s)	NYAWA Serge
Other teacher(s) – Autre(s) enseignant(s)	
Lecture Hours – Volume Horaire CM	15
TA Hours – Volume horaire TD	
TP Hours – Volume horaire TP	
Course Language – Langue du cours	English
TA and/or TP Language – Langue des TD et/ou TP	

Teaching staff contacts – Coordonnées de l'équipe pédagogique :

aurore.archimbaud@ippon-innov.eu

Serge NYAWA Tél : 0561294730 / s.nyawa@tbs-education.fr

Modes d'interaction privilégiés: sortie de cours et par mail

Course Objectives – Objectifs du cours :

1st part:

In today's data rich world students need to know how to gather data from stakeholder conversations on digital media and processing into structured insights, leading to more information-driven business decisions. Today, an important use of data visualization is to mine customer information to support marketing and customer service activities. Data visualization of non-structured data draws on four fields of study to improve decision making: social science and management science, to help express and model business problems, information systems to collect and manage data, and statistics to process data and visualize results. This class uses readily available computer software. By the end of this course, students should be able to: - collect, clean and manage non-structured datasets; - Understand the key principles of data visualisation; - to optimally interpret information.

2nd part:

Data visualization is a fundamental ingredient of data science as it “forces us to notice what we never expected to see” in a given dataset. In this course, we show through examples and case studies that graphical methods are powerful tools for revealing not only the structure of the data, but also patterns

and (ir)regularities, groups, trends, outliers... Dataviz is relevant both for data analysis, when the analyst wants to study data and, as any statistics, to question the data. It is also a tool for communication and, as such, is a visual language with a theory of the functions of signs and symbols used to encode the visual information. All along the course, we'll focus on methods, tools and strategies to represent simple and then complex or high-dimensional datasets, highlighting the growing development of dynamic and interactive tools.

Prerequisites – Pré requis :

Students should have followed R Programming, Business statistics or modelling classes or its equivalent.

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

Computers are allowed;

- The class is interactive and collaborative;
- Students are not allowed to join the class after 20 minutes from the start.

Bibliography/references – Bibliographie/références :

1st part :

- In Lee (2018), Social media analytics for enterprises: Typology, methods, and processes, Business Horizons, Volume 61, Issue 2, March–April 2018, Pages 199-210.
- Nathan Danneman and Richard Heimann (2014), Social Media Mining with R, Packt Publishing Ltd (Book).
- Matthew Ganis and Avinash Kohirkar (2015), Social Media Analytics, IBM Press Pearson plc. EMC Education Services, Data Science and Big Data Analytics (2015), WILEY (Book).

2nd part:

- Bertin, Jacques. 1983. Semiology of Graphics, translation from *Sémiologie graphique* (1967).
- Cleveland, William S., & McGill, Robert. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.*, 79(387), 531-554.
- Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2 ed. Graphics Press.
- Few, Stephen. 2012. *Show me the numbers: Designing tables and graphs to enlighten*. 2 ed. Burlingame: Analytics Press.
- Munzner, Tamara. 2014. *Visualization Analysis and Design*. (AK Peters Visualization Series). A K Peters/CRC Press.

- Tukey, John W. 1977. Exploratory data analysis. Reading, Mass. Unwin, Antony, Theus, Martin, & Hofmann, Heike. 2006. Graphics of large datasets: visualizing a million. Springer Science & Business Media

Session planning – Planification des séances :

1st part: The class made up of four chapters.

1. Text analytics technics with R
2. Social Media Mining: fundamentals and case studies in Sentiment Analysis
3. Scrapping Non-structured Data
4. Mining twitter with R

2nd part:

1. Data visualization for data sciences
2. Classics in Data visualization
3. Visualizing in many dimensions
4. Interactive and dynamic visualization

Grading system – Modalités d'évaluation :

Evaluation 1st part: Students are evaluated through a group project. No more than 03 students per group.

Evaluation 2nd part: Students are evaluated through a group project. No more than 02 students per group.

Distance learning – Enseignement à distance :

If necessary, online learning will be provided with:

1st part: - Interactive online classes and tutorials; - Multiple Choice Questions and online exercises.

2nd part: - Interactive online classes - Multiple Choice Questions and online exercises. - Remote (online) tutorials - Chatrooms