



Universidad Central de Venezuela  
Facultad de Ciencias

## Data Science 101

- Teacher : Alfredo Quintana
- Credit Units : 5 UC
- Theory : 4 hours/week
- Practice : 2 hours/week

### Content

#### **Unit 1 : Introduction**

Motivation; Data Driven Culture; Data Scientist Toolbox : R, Python, Project Jupyter, command-line tools, Github.

#### **Tema 2 : Getting Data**

Data Structures; SQL Databases; NoSQL Databases; APIs; Preprocessing Data; Pandas and Dplyr.

#### **Tema 3 : Exploratory Data Analysis**

Probability distributions; Summary Statistics; Data Visualization; Statistical Tests; Outlier Detection.

#### **Tema 4 : Machine Learning**

Supervised Learning. Unsupervised Learning. Reinforcement Learning.

#### **Tema 5 :Communication of Results**

Data visualization Tools: Plotly, Matplotlib; R Markdown.

#### **Unit 6: Big Data and Cloud Computing**

# Bibliography

- [1] Hadley Wickham, Garrett Grolemund, (2016) *R for Data Science*. O'Reilly Media.
- [2] Jake VanderPlas, (2016) *Python Data Science Handbook*. O'Reilly Media.
- [3] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, (2013) *An Introduction to Statistical Learning: with Applications in R*. Springer