

Exploratory Data Analysis

----- EDA Resources -----

Remember that in this program (beginner) we will see the basics of ML. We will start with very basic concepts and classifications, we will go back to regressions, model developments and data preprocessing. Later we will deepen the unsupervised learning, an introduction to deep learning and the NLP.



(60 min) Video EDA:

https://youtu.be/fwWCw_cE5al

The goal of this week is to become familiar with Pandas and Numpy if you are not already there, such as cleaning, analyzing and exploring data. On the other hand, we will see the classification problem.

Remember: *the course starts strong, week by week you will understand everything.*

While you are following the video, we recommend that you open the *topic1-exploratory-data-analysis-with-pandas* notebook and follow it carefully by executing and understanding what is being done. It's this same one you find in [Kaggle Kernel](#).

Reading

EDA <https://towardsdatascience.com/exploratory-data-analysis-with-pandas-508a5e8a5964>

----- **Optional** -----

(5 min - Optional) Article:

Article : <https://ourcodingclub.github.io/tutorials/pandas-python-intro/>

(Optional) Data Cleaning

<https://youtu.be/eMOA1pPVUc4>

----- Useful resources for this week -----

- ["Merging DataFrames with pandas"](#) - Tutorial
- ["Handle different dataset with dask and trying a little dask ML"](#) - Tutorial
- Official Pandas [Documentation](#)
- [Pandas in 10 minutes](#)
- [Pandas cheatsheet PDF](#)

RESOURCES

topic1_pandas_data_analysis.ipynb

telecom_churn.csv

Classification

----- Classification Features -----

We introduce ourselves to classification problems and learn to solve them using supervised learning techniques.



For the classification and decision trees part, open the *topic-3-decision-trees-and-knn notebook* while you follow the videos to run and implement it.

(60 min) Theoretical video:

<https://youtu.be/H4XIBTPv5rQ>

(60 min) Practical video:

<https://youtu.be/RrVYO6Td9Js>

----- Useful resources for this week -----

- If you want to follow the readings for an article [here](#)
- Sci-kit learn library [documentation](#)
- If you need any [tutorial](#) on sci-kit learn

RESOURCES

[topic3_decision_trees_kNN.ipynb](#)

#1 Practice!

Exploratory Data Analysis

To practice with Pandas and EDA, you can complete the following challenge where you will be analyzing socio-demographic data.

<https://medium.com/datadriveninvestor/data-preprocessing-for-machine-learning-188e9eef1d2c>

Exercises with solution:

https://drive.google.com/drive/folders/1Ox75wLVQcGP4cQ9iAGBvuwbEeWDfn_Rv

Classification

<https://medium.com/edureka/machine-learning-classifier-c02fbd8400c9>

<https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>

<https://www.aprendemachinelearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/>

RESOURCES

adult.data.csv

assignment03_decision_trees.ipynb

assignment01_pandas_uci_adult.ipynb