

CSCI316 Big Data Mining Techniques and Implementation

Laboratory 2 Pandas Tutorial

Objective

- To understand the importance of the data exploration step in Big Data
- To understand how to go about data exploration using Python and Pandas
- To understand the role of pre-processing in Big Data

The lectures explained the role of data exploration and pre-processing in Big Data. You are to perform data explorations on given datasets in order to explain the data and in order to decide on suitable pre-processing procedures.

Material: online Pandas tutorial <https://www.dataquest.io/blog/pandas-python-tutorial/>

Dataset: ign.csv (available on Moodle). A set of video-game rating data.

Follow in the online tutorial step-by-step to get yourself familiarized with data analytics, visualisation and pre-processing with Pandas in Python.