

COMP3162
Lab 1
Due Feb 4 11:59PM

You've been given a dataset (csv file) on university rankings. You will create a notebook and use the pandas library to do some exploratory data analysis. The following are the steps to be completed. Submit the notebook file .ipynb file to the course container.

Section 1 (Reading Data)

1. Read the csv file into a data frame.
2. Display the first five rows of the dataset.
3. Display the last five rows of the dataset.

Section 2 (Cleaning Data)

1. Show how many rows have missing values.
2. Remove any row that has missing values.
3. Rename the column **Student Population** to **Population**

Section 3 (Data Filtering and Selection)

1. Display the country and name of all the universities that have more than 20,000 students
2. Filter all the rows that have Student to Staff ratio higher than 10 and an overall score greater than 60.

Section 4 (Grouping and Aggregation)

1. Show the average student population grouped by country
2. Show the total number of students grouped by country
3. Find the university with the max overall score by country.
4. Show the average student to staff ratio grouped by country

Submission

Submit the notebook file .ipynb file on VLE by the due date. Ensure that in the notebook you show you put your id number. Also name the file your id number as well.