

Continuous Assessment 3

Due at 11:59pm on Friday 14th of January through Moodle.

Provide the Jupyter Notebook file including the code and all relevant statistics, plots and interpretations. Also upload the final dataset produced.

Devise an automated web scraping strategy for web-based data of your choice. Find a webpage to scrape data from. Provide code that scrapes, cleans, curates and stores the "clean" web-scraped data in a dataset.

When you have finalised your dataset, conduct an exploratory analysis on the data, including summary statistics, plots, and interpretations of your findings. Include the results of your analysis in the Jupyter Notebook file.

This analysis only has to contain a results section, and should not exceed 1,000 words. It should include summary statistics, plots, and interpretations of your findings.

The rubric is on the next page.

Rubric

Component	%	A(70-100)	B(60-70)	C(50-60)	D(40-50)	E-F(0-40)
Web scraping	50	Data scraping was complex. Data is sourced correctly and has been cleaned thoroughly and appropriately. Data has been appropriately structured and imported.	Data scraping was relatively straightforward. Data is sourced correctly and has been cleaned appropriately. Data has been appropriately structured and imported.	Data scraping was relatively straightforward. Data is sourced correctly and has been reasonably cleaned. Data structure is reasonably appropriate.	The data scraping exercise was simple and insufficient cleaning and structuring of the data was conducted.	The data scraping was not performed or was very poor. Poorly structured data. Little or no cleaning done on the data.
Analysis of dataset	50	Analysis of data is accurate, thorough and appropriate. Findings are well presented in both well-presented graphs and summary statistics. Excellent interpretations of the statistics and plots are provided.	Analysis of the data is appropriate, mostly accurate, and fairly thorough. Findings are reasonably well-presented and interpreted.	Analysis of the data is appropriate, but only somewhat accurate and thorough with significant omissions. Findings are reasonably presented but with clear issues.	Analysis of the data lacks rigour and accuracy with significant pieces of analyses missing. Findings are not well clearly presented or related to the analysis.	Analysis of the data is poor and lacks focus and accuracy with major omissions. Findings are poor and there are no clear questions being answered.