CONTINUOUS ASSESSMENT

Business Analytics 2

CODEBOOKS

- Mars crater study, presents global database that includes over 300,000 Mars craters 1 km or larger that were created between 4.2 and 3.8 billion years ago.
- The National Longitudinal Study of Adolescent Health (AddHealth) is a representative school-based survey of adolescents in grades 7-12 in the U.S.
- GapMinder includes one year of numerous country-level indicators of health, wealth and development.

RESEARCH QUESTION

- Once you select a data set of interest to you then you can identify a topic of interest and print those pages of the codebook that include the variable or variables that measure your selected topic.
- Many codebooks are too large to print so it can be important to create your own code book with only those pages that include the variables that you want to examine.

DATA ANALYSIS IN PYTHON

- Load a dataset, run frequency distributions for your chosen variables and select the columns (variables) and possibly rows (observations).
- Start to focus on your selected data in relation to your research question.
- Conduct data management on the variables you have identified.
  - Missing values, coding data, recoding variables, creating secondary variables, binning or grouping variables.
- Run new frequency distributions for your chosen variables after data management and select columns and possibly rows.

VISUALISATION IN PYTHON

- Create graphs of your variables one at a time (univariate graph)
  - Interpret the graph results (centre, spread)
- Create a graph showing the association between your explanatory and response variables (bivariate graph)
  - Interpret the graph results.

MARKING CRITERIA

- Is the research question and identified variables valid and supported in documentation?
- Was the program output interpretable? (labels, organized)
- Does the program output display three data managed variables as frequency tables?
- Did the summary describe the frequency distributions in terms of the values the variables take, how often they take them, the presence of missing data etc?
- Did the summary describe the data management conducted?
- Were the new frequency tables displayed after data management?
- Was a univariate graph created for each variable?
- Was a bivariate graph created for the selected variables?
- Did the summary describe what the graphs revealed in terms of the relationship?

Deliverable	Marks breakdown
Research question and identification of variables	10
Initial frequency distributions	5
Data management tasks	10
New frequency distributions	5
Univariate graphs	10
Bivariate graphs	10
Interpretation of graphs/frequencies/associations	10
Summary report (RQ's, background, codebook)	10
Total	70/100

DELIVERABLE

- Submit python file to Moodle with a summary report of your work.
- Due on Sunday March 17th 2019 11pm.
- Late submission has a penalty.
- Must be your own work, reference any code you used not written by yourself.