# **Dataset Exploration Project Part 1 (DE pt1) – Dataset Description & FINER Questions**

**Assignment Date:** Week 1

**Due Date:** End of day Sunday after Week 4

**Evaluation Weight:** 8% of course

# **Project Description – Please read the entire document before proceeding!**

This is the first part of a four-part cumulative project you will work on all term. **In each step, you will add new material and update your previous work as appropriate** If possible, you should use the same data set for all parts – otherwise, you will have to re-do previous work.

Think about subjects that interest you, and about what questions you might want to ask about them. In this first step of your project, the most important thing is that you have good data. **You must use MS Excel for all data manipulation, analysis and visualization for the entire project (DE pts 1-4), and your discussions (including appropriate visualizations) must be in MS Word format.** Here are the steps of DE pt 1:

1. Find an appropriate data source: your population or sample must have at least 1,000 records across 8 separate variables, with a good mix of qualitative and quantitative data types including at least two numerical variables.
2. Copy and rearrange your dataset for more effective analysis, with variable labels at the top, index and independent variables at the left, and panes frozen appropriately.
3. Check all variables for ranges and limitations (e.g. missing and/or invalid values), and comment as appropriate.
4. Determine any assumptions that may need to be made about your dataset. For example, do you know when or how the data were acquired? Is it a population or a sample? Do you know the units of measurement for all variables?
5. Describe your dataset in writing and provide a data dictionary (including statistical variable types, descriptions, ranges and limitations as appropriate for every variable).
6. Develop a minimum of three (3) FINER research questions you think might be analyzed with your data set. Your research questions should work together to help you build a coherent and cohesive analysis that examines relationships among the variables in your dataset; questions that can be answered by examining a single variable, or that don’t work together, are not enough. *NOTE that you should not be trying to answer these questions in DE part 1 or 2 – that comes later.*
7. Consider whether more and/or different data may be needed to answer your questions.
8. Begin to track your analysis –record your thoughts about what to study; what data you found and where; what questions you’re asking and what assumptions you’re making, and why. *Tracking is not about recording times or dates, or who said or did what – it’s about enabling other competent data analysts to replicate your work.*

**Your MS-Word discussion must stand on its own as a complete and thorough description of your dataset. You will submit your Excel spreadsheet as well, so your instructor can examine how you got your results and conclusions.**

***Be sure to include appropriate citations and references! Every chart, graph, table or other representation of data in your submitted document must also be found in your submitted spreadsheet or specifically cited and referenced as someone else’s work. If you quote, copy or otherwise borrow material without citation, you may be subject to penalties up to and including Academic Misconduct***

# **Evaluation**

# *Part 1 is marked out of 40, and worth 8% of the course.*

* **Appropriate Data Set (10 marks)** – Does your dataset follow the prescriptions above? Did you do a good job of arranging your data (modified data) for more effective analysis?
* **Dataset Description (10 marks)** – Is your description thorough & complete? Does your data dictionary describe each variable and identify its type, range and limitations (e.g. missing or invalid values)? Did you provide and justify appropriate assumptions and/or extra data requirements as needed?
* **FINER Questions (12 marks)** – Did you list at least three research questions that can be usefully and suitably investigated later on with your dataset? Are they clearly and explicitly stated? Do they work together to examine relationships among variables that will help to build a coherent and cohesive analysis? Did you describe how you developed them?
* **Demonstrated Tracking (8 marks)** – Did you track your activities and record them along with what worked well and what didn’t, so that you or another competent data analyst could repeat and/or improve on this analysis in the future? Do your assumptions and discussion cover potential issues?

# **Submission**

In the ‘Assignments & Tests’ folder in our Blackboard shell for this course there are two places for you to submit materials for each of the four parts of your Dataset Exploration project.

1. Your submission should include three files:
   1. a complete and thorough written discussion of your analysis, in MS-Word format, including charts, graphs and/or tables as appropriate to illustrate all points;
   2. your analysis spreadsheet, in MS-Excel format, including your original raw data in one tab, and all of your arranged and manipulated data and analysis, in other tabs in the same Excel file; and
   3. a Turnitin Report for your discussion file, including Turnitin’s marked-up version of your work, the Originality Report, and the Final Review
2. A day or two before the due date, you should submit your Word file in the **‘DE Turnitin’** assignment. This will not be marked, and does not appear in the Grade Centre – it’s there only to allow you to get Turnitin reports to help with your spelling, grammar and syntax and to see if you’ve cited and referenced everything correctly and whether you’ve used too much borrowed material. Do not submit CSVs or spreadsheets in Turnitin – it doesn’t work out well!

You must submit all of your work in the **‘Dataset Exploration Part 1’** assignment by the given due date and time for your section, and it must be accompanied by your Turnitin report. **If your work is not submitted on time, there is a penalty of 10% of the entire assignment per week late. If your work is submitted without a Turnitin report, there is a penalty of 5% on the entire assignment and your professor may submit your work in Turnitin regardless.**

***NOTE:*** *If your Turnitin report is not ready before the due date and time, you should submit the Turnitin receipt with your other files, then submit ONLY the Turnitin report separately as soon as it is available. If you submit your Word and Excel files more than one time, only the last submission will be graded, and late penalties will be applied as appropriate.*