Fox School of Business
TEMPLE UNIVERSITY ®

MIS 0855 Data Science (Section 006) – Fall 2017
Assignment #5 – Original Data Analysis (15% of the Final Grade)

Plan Due by Monday, November 20th, 11:59 PM EST (optional)
Final Due by <u>Monday, December 11th, 11:59 PM EST</u>

Task:

This assignment is to perform an original analysis of your own with a dataset of your choosing. A dataset can come from any source as long as it is something you have not already worked on for this course.

You can complete this assignment either individually or as a group of no more than four. Individual and group work will be graded with the same criteria.

Possible sources of data include

- Open government data sources such as Census.gov, Data.gov, OpenDataPhilly.org, Washington DC Open Data (http://opendata.dc.gov), NYC Open Data (https://nycopendata.socrata.com/) or Socrata (https://opendata.socrata.com/).
- Datasets from Pew Research Center (http://www.pewresearch.org/data/)
- Sports statistics, such as those for Major League Baseball (http://www.seanlahman.com/baseball-archive/statistics) or National Football League (http://nflsavant.com/)
- Election data such as FEC (http://www.fec.gov/pubrec/electionresults.shtml)
- Healthcare datasets from https://data.medicare.gov, https://www.healthdata.gov/, or any other source
- Datasets from World Bank (http://data.worldbank.org/)
- A data set from your current employer (Be careful about this one! Get their permission!)
- Any data that you can find on the Web!

It is not allowed to use any dataset that was already used for in-class exercises. You have to collect raw data (most likely in an Excel or a table format) by yourself and conduct your own analyses. While not required, integrating data from more than one data source will result in a higher grade.

Your analysis should clearly demonstrate the tools and techniques you've been exposed to in this course. This can take any form you'd like (i.e. comparison of averages across categories, mapping geographic data, developing and visualizing KPIs).

Deliverables:

Submit every file that you use for your analysis, including datasets and analysis files, into Canvas. If you use Tableau for analysis, submit your Tableau file as well.

In addition, submit a PowerPoint file that consists of the following. Please refer to the sample files posted on the class site.

1. Title of your analysis
2. Scenario or questions – What question do you want to answer and why is it important? What are your hypotheses?
3. Description of the data – What are the key elements and where/how did you get it?
4. Analysis and results with visualization – The more, the merrier. Visualize so that they look informative and visually appealing.
5. Conclusions – What did you learn? What insights or decisions can we make based on your findings?
6. References

Data Analysis Plan Submission Instruction (Optional, due by November 20)

- If you can submit your data analysis plan by November 20, the instructor will provide feedback to your plan.

- Your plan should include a title, a scenario, and data sources (#1, #2, and #3 in the above list). Analysis results or data are not needed.

- Submit your file in Word or Powerpoint into Canvas by Monday, November 20th, 11:59PM EST.

Final Submission Instruction (due by December 11)

- Submit all of your files (PowerPoint, Excel, and Tableau, if used) into Canvas by Monday, December 11th, 11:59PM EST. This deadline is firm, and the instructor will not take any extraneous circumstance into consideration that occurs to you such as a PC malfunction or network outages.

- Please make sure to list your teammates in your PowerPoint file.

- Late submission is allowed, but there will be 10% penalty per each 12 hours. For example, if you submit in the morning of Dec 13th, a 30% penalty is imposed on your submission. Thus, any submission after noon of Dec 16th will not be graded.

EVALUATION CRITERIA FOR ASSIGNMENT #5

	4 (A-level)	3 (B-level)	2 (C-level)	1 (D or F-level)
Scenario Identification (20%)	Comprehensively and clearly identifies and describes the issue. The central question is clearly stated.	Precisely identifies and describes the issue including the majority of its key components/variables. The central question is somewhat clearly stated.	Correctly identifies issue but certain key components/variables remain unclear or omitted. The central question is somewhat unclear.	Limited ability to clearly identify the issue and its various components/variables. The central question is unclear or missing.
Data Gathering (20%)	Gathers correct and highly credible data that directly answers the central question.	Gathers correct and highly credible data that mostly answers the central question.	Gathers relevant data that somewhat answers the central question.	Gathers and incorrect, insufficient or unreliable data. Data is not directly related to central question.
Analysis (20%)	Presents an insightful and thorough analysis. Analysis is driven by logical arguments clearly related to central question.	Presents an effective analysis. Analysis is driven by arguments that, while slightly flawed, are related to central question.	Presents a superficial analysis of central question. Aspects of the question are unanswered.	Presents an incomplete analysis of central question. Analysis is not related to the central question.
Conclusions (20%)	Clearly identifies and articulates all implications and consequences of analysis. The conclusions are clearly related to the results of the analysis.	Somewhat clearly identifies and articulates all implications and consequences of analysis. The conclusions are mostly related to the results of the analysis.	Some implications and consequences of analysis are unidentified, unaddressed, or unarticulated. The conclusions are weakly related to the results of the analysis.	Implications and consequences of analysis are mostly unaddressed. It is unclear how conclusions follow from the analysis and what was learned from the analysis.
Visual Appeal (20%)	Visuals display high levels of creativity and strongly enhance the effectiveness of the presentation. Clearly leverages principles of good visualizations.	Visuals display satisfactory levels of creativity and generally enhance the effectiveness of the presentation. Mostly leverages principles of good visualizations.	Visuals display marginal levels of creativity and somewhat enhance the effectiveness of the presentation. Principles of good visualizations are underutilized.	Visuals do not enhance, or get in the way of, the effectiveness of the presentation. Principles of good visualizations are not used.