MIS0855: Data Science
Assignment 5 (Final Group Project): Original Data Analysis
DUE: December 1, 2015 BEGINNING OF CLASS

Task:
In groups, perform an original analysis on a data set of your choosing. The data set can come from any source as long as it is something you have not already worked on for this course. You also cannot use one of the data sets from the Temple Analytics Challenge. Possible sources of data include:

- Open data sources such as Data.gov, OpenDataPhilly.org, and OpenData by Socrata.
- Data sets from the Pew Research Center (www.pewresearch.org/data)
- Sports statistics, such as those from the Baseball Archive (http://www.seanlahman.com/baseball-archive/statistics)
- A data set from your current employer (Be careful about this one! Get their permission!)
- An original survey conducted by your group.

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, sentiment analysis, developing and visualizing KPIs).

Deliverables:
A seven slide deck, using the following format:

- Slide 1 will list your group members and the title of your presentation.
- Slide 2 will describe the scenario. What question do you want to answer and why is it important?
- Slide 3 will describe the data. What are the key elements and how did you get it?
- Slides 4 and 5 will describe the analysis and the results. Make good use of data visualizations.
- Slide 6 will describe your conclusions. What did you learn? Support your conclusions using the results of the analysis, citing specific evidence!
- Slide 7 will list your references.

Your group will present your work in class through a five-minute presentation, with two minutes for questions. The presentation will take place the day the project is due.

Guidelines:

- You will work in groups of 4 – each member of the group will receive the same grade.
  o Not every group member must present in front of the class, but everybody in the group should contribute to the final deliverables.
- Your group will email (1) the slide deck and (2) your data set to me by the start of class, the day this project is due. Your deliverables should be emailed to me as attachments with the subject: “MIS0855: Assignment 5”
<table>
<thead>
<tr>
<th>TRAIT</th>
<th>4 (A-level)</th>
<th>3 (B-level)</th>
<th>2 (C-level)</th>
<th>1 (D or F-level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario Identification (20%)</td>
<td>Comprehensively and clearly identifies and describes the issue. The central question is clearly stated.</td>
<td>Precisely identifies and describes the issue including the majority of its key components/variables. The central question is somewhat clearly stated.</td>
<td>Correctly identifies issue but certain key components/variables remain unclear or omitted. The central question is somewhat unclear.</td>
<td>Limited ability to clearly identify the issue and its various components/variables. The central question is unclear or missing.</td>
</tr>
<tr>
<td>Data Gathering (20%)</td>
<td>Gathers correct and highly credible data that directly answers the central question.</td>
<td>Gathers correct and highly credible data that mostly answers the central question.</td>
<td>Gathers relevant data that somewhat answers the central question.</td>
<td>Gathers and incorrect, insufficient or unreliable data. Data is not directly related to central question.</td>
</tr>
<tr>
<td>Analysis (20%)</td>
<td>Presents an insightful and thorough analysis. Analysis is driven by logical arguments clearly related to central question.</td>
<td>Presents an effective analysis. Analysis is driven by arguments that, while slightly flawed, are related to central question.</td>
<td>Presents a superficial analysis of central question. Aspects of the question are unanswered.</td>
<td>Presents an incomplete analysis of central question. Analysis is not related to the central question.</td>
</tr>
<tr>
<td>Conclusions (20%)</td>
<td>Clearly identifies and articulates all implications and consequences of analysis. The conclusions are clearly related to the results of the analysis.</td>
<td>Somewhat clearly identifies and articulates all implications and consequences of analysis. The conclusions are mostly related to the results of the analysis.</td>
<td>Some implications and consequences of analysis are unidentified, unaddressed, or unarticulated. The conclusions are weakly related to the results of the analysis.</td>
<td>Implications and consequences of analysis are mostly unaddressed. It is unclear how conclusions follow from the analysis and what was learned from the analysis.</td>
</tr>
<tr>
<td>Visual Appeal (15%)</td>
<td>Visuals display high levels of creativity and strongly enhance the effectiveness of the presentation. Clearly leverages principles of good visualizations.</td>
<td>Visuals display satisfactory levels of creativity and generally enhance the effectiveness of the presentation. Mostly leverages principles of good visualizations.</td>
<td>Visuals display marginal levels of creativity and somewhat enhance the effectiveness of the presentation. Principles of good visualizations are underutilized.</td>
<td>Visuals do not enhance, or get in the way of, the effectiveness of the presentation. Principles of good visualizations are not used.</td>
</tr>
<tr>
<td>References (5%)</td>
<td>References follow a standard for citation. All references are listed in a slide at the end of the deck and includes full citations.</td>
<td>References generally follow a standard for citation with minor errors. All references are listed in a slide at end of the deck and includes full citations.</td>
<td>References don’t follow a standard for citation. One or more references are missing from the slide at the end of the deck.</td>
<td>References are missing from the slide deck.</td>
</tr>
</tbody>
</table>