94-881: Managing Analytics Projects (Fall 2023)
HW#3: Choices around Analytics and Visualizations
Due October 4, 2023

David Steier

The focus of this assignment is making choices around analytics and visualizations for your team project (it does not include infrastructure).

1. **What analytic techniques are matches for improving the decision?**
    - What analytics problem type matches this decision?
    - What analytics techniques can be considered for this problem type?
    - What techniques satisfy the decision and process constraints (accuracy, timeliness, etc.)?
2. **How can the analytic results best be communicated to stakeholders?**
    - What information needs to be communicated to stakeholders to take the necessary action(s)?
    - Which data visualization techniques best communicate the necessary information for each stakeholder?

Be as specific in your answers as possible: classify the problem type, recommend analytics techniques and your reasoning, and show a prototype of the data visualizations.

There is no standard list of analytics problem types, so you may use the list of task types discussed in class. Another possible reference is J. Han, M. Kamber and J. Pei, *Data Mining Concepts and Techniques*, Third Edition, Morgan Kaufmann, 2011 (e.g. class/concept discrimination, mining frequent patterns, associations and correlations, classification and regression for predictive analysis, cluster analysis, outlier analysis).
These types are not cast in stone, and you can feel free to add others.

Analytic techniques map onto these problem types: e.g. classification can be implemented by decision trees, neural nets, naïve Bayes, support vector machines, etc. For lists of analytics techniques, the *Booz Allen Field Guide to Data Science* is recommended. Your choice of analytics techniques should satisfy the constraints you gathered in the problem framing step on the decision and the decision process, and take into account what you've learned about the data sources for your project

There is also no definitive list of visualization types, but a good list is in the visualization e-book assigned in the syllabus and in the Data Visualization catalogue at https://datavizcatalogue.com/index.html

Rubric

This assignment will be graded based on how well the submission answers the questions under analytics and visualization (four points for each question, for a maximum of 20 points for the assignment)

1. What analytics problem type matches this decision? Problem types should be at the level of classification, regression, clustering etc., not too high level (e.g. analytics) or specific analytics technique (e.g. random forest)
2. What analytics techniques can be considered for this problem type? Full answers should list more than one technique and be specific. (not classification – too high-level but algorithms like logistic regression and random forests)
3. What techniques satisfy the decision and process constraints (accuracy, timeliness, etc.)? The full answer should list the decision and process constraints considered and explain the rationale for why a technique satisfies the constraint.
4. What information needs to be communicated to stakeholders to take the necessary action(s)? The answer should describe the result of the analytics that the decision-maker uses to make an improved decision, and perhaps information relevant to other stakeholders as well (say, cost or accuracy of the decision)

5. Which data visualization techniques best communicate the necessary information for each stakeholder? Full answers should be specific visualization techniques, like bar charts, or Sankey diagrams. The answers include an example applied to your project, even in prototype form if actual data is not available.