Worcester Polytechnic Institute

Semester Project Description

MIS 502 | Managing Data for Analytics

WPI Business School

The semester project for MIS 502 is an opportunity to apply your data management skills learned over the semester to a large real-world data set. Each individual student will perform the following tasks:

1) Select a real-world data set. If the data set has less than 10,000 records, then you will need my approval to use it.
2) Compose a description of the data set, describe the data, and profile the data.
3) Prepare/wrangle the data set using Python.
4) Use data mining and data visualization techniques on the data set using Python.
5) Write a project report describing the data set, documenting the data mining and visualization techniques you have employed, and summarizing the outcomes.
6) Present the data set outcomes to the class.

DATA SET SELECTION

Select a data set that you are familiar with – something from your work experience is usually a good idea. Data sets from your work are encouraged or you may wish to register with Kaggle for a listing of available large data sets (http://www.kaggle.com). Be able to explain the context of the data set and all data elements contained in it.

CONTENTS OF THE PROJECT REPORT

1) Data Set Description

a) Describe the nature and context of the data set.
b) Identify and discuss potential outcomes from analysis of the data set.
c) Identify initial data quality, data integrity, and/or data ethics issues related to the data set.

2) Data Preparation / Data Wrangling

a) Identify all steps and outcomes of the data preparation phase.
b) Provide data profiling information for each data element.
c) Summarize outcomes of the data preparation/wrangling and compare with initial data set potential outcomes.
d) Provide the Python code used.

3) Data Mining Outcomes

a) Use appropriate data mining techniques on the data set, explain and document the results. Attempt to use each of the techniques we covered (e.g., cluster analysis, classification, regression, association, correlation, and dimensionality reduction as well as text mining.) If a technique is not appropriate or relevant, then be sure to explain why.
b) Provide the Python code used in each technique.
c) Summarize the outcomes of the data mining techniques and compare with initial data set potential outcomes.

4) Data Visualization Outcomes

a) Use appropriate data visualization techniques on the data set, explain and document the results. Attempt to use each of the techniques we covered (e.g., column, bar, pie, box-plot, line, multi-line, and scatter). If a technique is not appropriate or relevant, then be sure to explain why.
b) Provide the Python code used in each technique.
c) Summarize the outcomes of the data mining techniques and compare with initial data set potential outcomes.

5) Project Presentation

a) Create a PowerPoint presentation to explain and summarize your project's data set, data preparation/data wrangling, data mining, and data visualization outcomes. Consider the audience for the presentation to be unfamiliar with the context of your data. Presentation duration should be targeted for 10 minutes.
b) Upload a copy of this presentation to the appropriate discussion board forum in our Canvas module prior to the deadline stated in our syllabus under the tentative course schedule.

PROJECT SCHEDULE

Weekly Module	Milestone	Content
Week 5	1	TURN IN: Project data set description (include context of the data set and number of records)
Week 6		Approval/Disapproval of data set selection will be returned.
Week 8	2	TURN IN: Complete draft of your approved data set content (include all the content under data set)
Week 9		Feedback on drafts will be given.
Week 11	3	TURN IN: (1) Revised draft of approved data set content. (2) Draft of data preparation/data wrangling content (no code). (3) List of data mining and visualization techniques to employ.
Week 12		Feedback on drafts and lists will be given.
Week 14	4	Project presentations shared with class on Thursday. Students post comments on project presentations by Sunday. TURN IN: Your complete Project Report