

## Assignment #2 (Group Assignment)

Solve a prediction problem using an unsupervised learning method. Work in groups of 4-5. Have ONE person from your group submit the following: problem statement, approach, summary of findings/recommendation, data, visualizations, model/code, output. Make sure to include the group name and names of all group members. You can source the data from here (the link will open in new window):

<https://blog.bigml.com/list-of-public-data-sources-fit-for-machine-learning>

You are expected to submit at the due date. **Sunday, 3 November 2019, 11:55 PM**

- 1) R Markdown
- 2) Deployed Shiny app link
- 3) Shiny app code + data used

**Assignment Rubric:** [https://learn.continue.yorku.ca/pluginfile.php/282419/mod\\_page/content/4/Rubric%20for%20Assignments.pdf](https://learn.continue.yorku.ca/pluginfile.php/282419/mod_page/content/4/Rubric%20for%20Assignments.pdf)

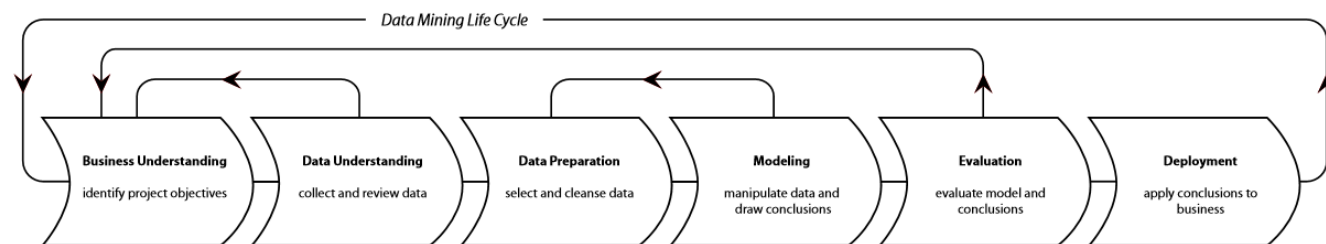
### Rubric for Assignments # 1 and # 2

#### ML 1000

Assessment Criteria	Not Good Enough (0 ≤ score < 2)	Good (3 ≤ score < 4)	Very Good (score 5)
<b>Interpretation of Data (qualitative)</b>	Little or no attempt to interpret data; or there are significant errors; or some data are over- or under-interpreted.	Interpret most data correctly; part of conclusions may be suspect; suggestions on future implementation are sound.	Data are completely and appropriately interpreted; there is no over- or under-interpretation; draw convincing conclusions.
<b>Analysis (quantitative)</b>	Methods are completely misapplied or applied but with significant errors or omissions. Choose inappropriate methods and make wrong predictions.	Most statistical methods are correctly applied but more could have been done with the data. Predictions are sensible but may deviate from the true results in a large range.	Statistical methods are fully and correctly applied; demonstrate superior data analysis skills; deeply mine the data and obtain useful insights for decision making.
<b>Critical evaluation of findings;</b>	Blindly accept defective results; or recognize defective results but does not know how to fix them.	Recognize defective results and figure out the causes; understand the main sources of errors.	Show deep understanding for the sources of errors; recognize defective results and eliminates the causes.
<b>Ability to draw proper conclusions and make effective suggestions</b>	No drawn conclusions; draw incorrect conclusions; suggestions are not acceptable.	Draw correct conclusion; suggestions may have potential impact on the future business.	Demonstrate substantial understanding of the problem; conduct deep data analytics using correct methods; draw correct conclusions with sufficient explanation and elaboration.
<b>Writing</b>	Report is inadequately written and poorly organized. Analysis is insufficient. Conclusions are unconvincing.	Report is concise and clearly written. Analyze problems following scientific strategies; provide useful suggestions with detailed explanation.	Report is well organized and insightfully written, includes thorough and thoughtful details. Conclusions are convincing.

# Cross Industry Standardized Procedure for Data Mining:

## Phases



### Determine Business Objectives

Background  
Business Objectives  
Business Success Criteria  
(Log and Report Process)

### Assess Situation

Inventory of Resources,  
Requirements, Assumptions,  
and Constraints  
Risks and Contingencies  
Terminology  
Costs and Benefits  
(Log and Report Process)

### Determine Data Mining Goals

Data Mining Goals  
Data Mining Success Criteria  
(Log and Report Process)

### Produce Project Plan

Project Plan  
Initial Assessment of Tools and  
Techniques  
(Log and Report Process)

### Collect Initial Data

Initial Data Collection Report  
(Log and Report Process)

### Describe Data

Data Description Report  
(Log and Report Process)

### Explore Data

Data Exploration Report  
(Log and Report Process)

### Verify Data Quality

Data Quality Report  
(Log and Report Process)

### Data Set

Data Set Description  
(Log and Report Process)

### Select Data

Rationale for Inclusion/  
Exclusion  
(Log and Report Process)

### Clean Data

Data Cleaning Report  
(Log and Report Process)

### Construct Data

Derived Attributes  
Generated Records  
(Log and Report Process)

### Integrate Data

Merged Data  
(Log and Report Process)

### Format Data

Reformatted Data  
(Log and Report Process)

### Select Modeling Technique

Modeling Technique  
Modeling Assumptions  
(Log and Report Process)

### Generate Test Design

Test Design  
(Log and Report Process)

### Build Model Parameter Settings

Models  
Model Description  
(Log and Report Process)

### Assess Model

Model Assessment  
Revised Parameter  
(Log and Report Process)

### Evaluate Results

Align Assessment of Data  
Mining Results with  
Business Success Criteria  
(Log and Report Process)

### Approved Models

Review Process  
Review of Process  
(Log and Report Process)

### Determine Next Steps

List of Possible Actions  
Decision  
(Log and Report Process)

### Plan Deployment

Deployment Plan  
(Log and Report Process)

### Plan Monitoring and Maintenance

Monitoring and  
Maintenance Plan  
(Log and Report Process)

### Produce Final Report

Final Report  
Final Presentation  
(Log and Report Process)

### Review Project

Experience  
Documentation  
(Log and Report Process)

## Generic Tasks

## Specialized Tasks

(Process Instances)

## a visual guide to CRISP-DM methodology

SOURCE CRISP-DM 1.0

<http://www.crisp-dm.org/download.htm>

DESIGN Nicole Leaper

<http://www.nicoleleaper.com>

