

2019-0703 IST 707 Data Analytics

Homework Assignment 1 (week 1)

Ryan Timbrook

NetID: RTIMBROO

Course: IST 707 Data Analytics

Term: Summer, 2019

Homework Assignment 1 (week 1)

Table of Contents

1	Introduction	3
1.1	Purpose	3
1.2	Scope	3
2	Analysis and Models	4
2.1	About the Data	4
2.1.1	Dataset Info	4
2.1.2	Data Exploration & Cleaning	4
2.1.3	Data Transformations	4
2.1.4	Data Tables & Visualizations	4
2.2	Models	4
2.2.1	Model Details	4
3	Results	6
3.1	Model x Results	6
3.1.1	Model x Visualizations	6
4	Conclusions	7

Homework Assignment 1 (week 1)

1 Introduction

Introduction: (3 paragraphs)

General: An Introduction is about the area or topic, not about the data or models. The introduction helps the reader to understand what the assignment area is about. For example, support the assignment is about schools. In this case, the introduction is about school systems, why schools are measured and ranked, who might be concerned with school measures and rankings (such as students, parents, states, governments, and funding agencies), and the value of comparing schools.

An introduction is a like a warm-up or like dating. It allows the reader to “get to know” the area of interest.

The Intro should not contain any information about the dataset or the data cleaning, prep, processing, etc. Everything about the dataset goes into the Analysis section under the “About the Data” subsection.

Introductions can and should include basis, background, history, the state-of-the-art, images, references, etc.

1.1 Purpose

1.2 Scope

Homework Assignment 1 (week 1)

2 Analysis and Models

The Analysis section contains **subsections**.

The second and remaining subsections of Analysis are the model(s).

In some cases, there may only be one model. A model is any method used to analyze the data. Each Assignment specifies which models to use. Always include model details and parameter values when applicable.

***** Have Visualizations throughout the assignment.**

Include measures and comparisons.

Tables are great for comparing.

2.1 About the Data

Contains all the information about the dataset, the variables, the cleaning and prep, checking for and dealing with missing values, checking for and dealing with incorrect values, checking for and dealing with outliers, feature generation, normalization (if needed), etc. In this subsection, you will also “explore” the data.

This means that you write about each variable, **visualize** each variable (as feasible), and talk about what the variable represents. Tables are great for this as well.

2.1.1 Dataset Info

2.1.2 Data Exploration & Cleaning

2.1.3 Data Transformations

2.1.4 Data Tables & Visualizations

2.2 Models

The second and remaining subsections of Analysis are the model(s).

In some cases, there may only be one model. A model is any method used to analyze the data. Each Assignment specifies which models to use. Always include model details and parameter values when applicable.

2.2.1 Model x1 Details

Homework Assignment 1 (week 1)

2.2.2 *Model x1 Parameters*

2.2.3 *Model x2 Details*

2.2.4 *Model x2 Parameters*

Homework Assignment 1 (week 1)

3 Results

The Results section of the Assignment will have a subsection for results for each model (assuming that you have more than one).

Results are technical.

They offer technical information about what was found in the analysis. For example, if you performed a correlation in the analysis between all pairs of numeric variables, then your results would discuss the r-value and relationship of each pair. Similarly, if you looked at measures of center and variation, the results talk about what those measures are and what they reveal. For example, if the mean is less than the median, the data is skewed, which means....

Each model we will use in this class has results and parameters associated with it. For example, association rule mining will offer the top ten rules for sup, conf, and/or lift if you code it to do so. These would go into the results along with the sup, conf, and lift for each rule. The meaning would also be discussed.

**** Always have visualizations**

3.1 Model x1 Results

Technical Analysis, discoveries found...

3.1.1 Model x1 Visualizations

3.2 Model x2 Results

Technical Analysis, discoveries found...

3.2.1 Model x2 Visualizations

Homework Assignment 1 (week 1)

4 Conclusions

General : 3 paragraphs.

This area is not technical at all.

This area explains what was actually found in a way that would make sense to anyone. For example, if you discovered in the analysis that association rule mining with a conf of .2 and a sup of .3 offered 10 rules, you would talk about the measures and values and rules in the *results*. In the Conclusions, you would talk about what it all means. So you would not include the rules themselves or mention of technical measures such as conf or sup. Rather, you would say that you found (as a random example) that people who buy diapers are very likely to buy beer and that this means that a store should consider placing these items "near" to each other.