

**Lab 2: Parsing Data**

Name: &lt;Nathan Rusk&gt;

The goal of this lab is to understand the structure of data. In this lab you will change data into a format that tags each part of the data with its intended use. After completing this lab every element of the data, you selected (Tableau dataset) and the two (2) additional datasets you acquired in lab last week will be broken into its individual parts.

**What you should be able to do (at the end of this lab):**

Remember	<i>Describe</i> what happens in the <b>parse</b> stage.
Understand	<i>Describe</i> the data in detail according to the parsing specifications.
Apply	<i>Demonstrate</i> the ability to change data into a useful format for future processing.
Evaluate	<i>Categorize</i> the data according to parsing specs.
Analysis	<i>Identify</i> specific features about the data.
Create	<i>Generate</i> a parsed listing of the data.

**Part I: Tableau Data Set****In this document:**

Answer the following questions about your Tableau data:

1. List the name of the Tableau Dataset you selected in the Acquire Lab: **netflix\_titles.xlsx**
2. How many rows (records) are in the data set? **76,248**
3. How many columns (variables) are in the data set? **13, 17 including show\_id duplicates in the separate tables**
4. What assumptions are you making about the data?
  - Not all entries have a value in column duration\_seasons (some are movies rather than series).
  - There is a wide range of maturity ratings.
  - There is a wide range of genres.

Use the following link to provide a parsing of the Tableau Data.

<https://tinyurl.com/Parse-Worksheet>

For Visualization Assignment select: **Training Data**

Then provide the parsed data and list your assumptions.

**Part II: Additional Data Set #1****In this document:**

Answer the following questions about your Tableau data:

1. List the name of the Tableau Dataset you selected in the Acquire Lab: **disney\_plus\_shows.csv**
2. How many rows (records) are in the data set? **993**
3. How many columns (variables) are in the data set? **19**
4. What assumptions are you making about the data?
  - Most maturity rating are below PG-13.
  - Content includes well known/popular titles considering awards and number of imdb\_votes.
  - Some titles do not have values for released\_at and or added\_at.

**Lab 2: Parsing Data**

Use the following link to provide a parsing of the first additional data set from Lab 1.

<https://tinyurl.com/Parse-Worksheet>

For Visualization Assignment select: **Lab assignment**

Then provide the parsed data and list your assumptions.

**Part II: Additional Data Set #2**

**In this document:**

Answer the following questions about your Tableau data:

1. List the name of the Tableau Dataset you selected in the Acquire Lab: **Prime TV Show Data set.csv**
2. How many rows (records) are in the data set? **393**
3. How many columns (variables) are in the data set? **8**
4. What assumptions are you making about the data?
  - Viewer age range is varying.
  - Genre of the shows vary.
  - Shows of different language are available (different demographics).

Use the following link to provide a parsing of the second additional data set from Lab 1.

<https://tinyurl.com/Parse-Worksheet>

For Visualization Assignment select: **Lab assignment**

Then provide the parsed data and list your assumptions.

**What to submit:**

1. This document saved as a PDF file: **LastnameFirstInitial-CGT270Fall2021-Lab2Parsing.pdf**
2. Parse Activity Worksheet (.pdf) for the Tableau data set.
3. Parse Activity Worksheet (.pdf) for 1<sup>st</sup> additional data set.
4. Parse Activity Worksheet (.pdf) for the 2<sup>nd</sup> additional data set.