

Lab 2: Parsing DataName: **Zhenghao Fang**

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 parse 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:
Hollywoods Most Profitable story
2. How many rows (records) are in the data set?
74
3. How many columns (variables) are in the data set?
7
4. What assumptions are you making about the data?
The profitability of the movies would be the income from the movie theater, online platform.

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:
Hollywoods Most Profitable story
<https://github.com/gchan/hollywood-budgets/blob/master/public/data/Most%20Profitable%20Hollywood%20Stories.csv>
2. How many rows (records) are in the data set?
640
3. How many columns (variables) are in the data set?

Lab 2: Parsing Data

15

4. What assumptions are you making about the data?
There is not a lot of difference between comedy, so anything with “comedy” will be assumed as comedy.

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:
Hollywoods Most Profitable story
https://sit.instructure.com/courses/22632/files/2992196?module_item_id=462843
2. How many rows (records) are in the data set?
471
3. How many columns (variables) are in the data set?
19
4. What assumptions are you making about the data?
The unit of the foreign gross is dollar.

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 1st additional data set.
4. Parse Activity Worksheet (.pdf) for the 2nd additional data set.