CGT 270 Data Visualization
Module 1 ● Week 2
Lab 2: Parsing Data

Name: **Jeremy Chen**

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. Answer the following questions and complete the table for each dataset.

1. List the name of the Tableau Dataset you selected in the Acquire Lab: The 2014 Inc. 5000
2. How many rows (records) are in the data set? 5001 rows
3. How many columns (variables) are in the data set? 19 columns
4. What assumptions are you making about the data?

Growth/Revenue - are expected to be US currency of how much they earn.
Input – blank/empty column that could represent the act of entering data
Rank - list of integers to show which company has the best revenue
Yrs_on_list - since this is list is for 2014, they must have list that are from the past to show if the company was on the list before.

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

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

Tableau Data Set

In the table below list each variable and its data type (add more rows as needed):

	Variable	**Data type**
1	**_input**	**Void**
2	**_num**	**Integer**
3	**_widgetName**	**String**
4	**_source**	**String**
5	**_resultNumber**	**Integer**
6	**_pageUrl (parsing)**	**String**
	Hostname (eg .com)	**String**
	Path name (eg .javacript, .json	**string**
7	**id**	**Integer**
8	**rank**	**Integer**
9	**workers**	**Integer**

Save this file as: **LastnameFirstInitial-CGT270Fall2021-Lab2Parsing.pdf**

10	**Company**	**Character**
11	**url**	**character**
12	**state_l**	**character**
13	**state_s**	**character**
14	**city**	**String**
15	**metro**	**character**
16	**growth**	**float**
17	**revenue**	**Integer**
18	**industry**	**character**
19	**yrs_on_list**	**integer**

You may add more rows and attach additional pages if needed.

Save this file as: **LastnameFirstInitial-CGT270Fall2021-Lab2Parsing.pdf**

CGT 270 Data Visualization
Module 1 ● Week 2
Lab 2: Parsing Data

Additional Data Set #1

1. List the name of the first (1st) additional data set you acquired in the Acquire Lab: Business and Industry Reports
2. How many rows (records) are in the data set? There are 1048576 rows in this data set
3. How many columns (variables) are in the data set? There are 3 columns in the data set
4. What assumptions are you making about the data?

Time series code – to categorize the different section of the report.
Date – there are some data that are lost which gives us some values that are void.
Value – used to estimate forecast of stock value for future.

In the table below list each variable and its data type (add more rows as needed):

	Variable	Data type
1	Time_series_code	String
2	Date	Date
3	Value	integer
4		

You may add more rows and attach additional pages if needed.
There is no parsing needed in all my data sets. All my values for each variable are in its simplest form.

Additional Data Set #2

1. List the name of the second (2nd) additional data set you acquired in the Acquire Lab: Stock Exchange data
2. How many rows (records) are in the data set? There are 112458 rows
3. How many columns (variables) are in the data set? There are 8 columns
4. What assumptions are you making about the data?

Date- some of the data are missing which cause this data to be void.
High/low – us currency of when the stock is at its highest and lowest return.
Open/close – the price of stock at particular time.

In the table below list each variable and its data type (add more rows as needed):

	Variable	Data type
1	Index	Character
2	date	Date
3	open	Float
4	high	Float
5	low	Float
6	close	Float

Save this file as: **LastnameFirstInitial-CGT270Fall2021-Lab2Parsing.pdf**

7	**adjclose**	**Float**
8	**volume**	**integer**

You may add more rows and attach additional pages if needed.

There is no parsing needed in all my data sets. All my values for each variable are in its simplest form.

Save this file as: **LastnameFirstInitial-CGT270Fall2021-Lab2Parsing.pdf**