

Lab 2: Parsing DataName: <**Ria Rajan**>

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: Employment Changes in Great Britain by Industry
2. How many rows (records) are in the data set? 1332 rows
3. How many columns (variables) are in the data set? 10
4. What assumptions are you making about the data?
 - SIC-1: The SIC codes come as a character (1 letter) and do not follow regular SIC code structure (4-digit codes) and therefore probably have their own correlation to the SIC-1 name category that doesn't match up with regular SIC codes.
 - SIC-2: These SIC codes come as an integer (2-digit integers) and also don't follow regular SIC code structure, so I think they have their own correlation to the Industry category that doesn't match up with regular SIC codes.

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.

Tableau Data Set

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

	Variable	Data type
1	City	String
2	Country	String
3	SIC-1	Character
4	SIC-1 name	String
5	SIC-2	Integer
6	Industry	String

7	Jobs 2011	Integer
8	Jobs 2014	Integer
9	Change	Integer
10	% Change	Float/double

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: US Unemployment Rate by County, 1990-2016
2. How many rows (records) are in the data set? 885,548 rows
3. How many columns (variables) are in the data set? 5 columns
4. What assumptions are you making about the data?
 - State: most of the states are very heavily focused on Southern states and that makes it an inaccurate representation of the unemployment in the US.

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

	Variable	Data type
1	#Year	Integer
2	Month	String
3	State	String
4	County	String
5	Count name	String
6	County category	String
7	#Rate	Double

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

Additional Data Set #2

1. List the name of the second (2nd) additional data set you acquired in the Acquire Lab: French employment, salaries, population per town
2. How many rows (records) are in the data set? 36,681 rows
3. How many columns (variables) are in the data set? 14 columns
4. What assumptions are you making about the data?
 - #of firms with x number of employees: The testing group is either too small or too spread out by number of employees because the data they gather especially for larger groups of employees are really small.

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

	Variable	Data type
1	CODGEO (geographic code)	Integer
2	LIBGEO (town name)	String
3	REG (region number)	Integer
4	DEP (department number)	Integer
5	E14TST (total number of firms)	Integer
6	E14TS0ND (# unknown/null size firms)	Integer
7	E14TS1 (# firms with 1-5 employees)	Integer
8	E14TS6 (# firms with 6-9 employees)	Integer
9	E14TS10 (# firms with 10-19 employees)	Integer
10	E14TS20 (# firms with 20-49 employees)	Integer
11	E14TS50 (# firms with 50-99 employees)	Integer
12	E14TS100 (# firms with 100-199 employees)	Integer
13	E14TS200 (# with 200-499 employees)	Integer
14	E14TS500 (# firms with more than 500 employees)	Integer

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

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