

Lab 2: Parsing DataName: **Tejasvi Bhagwatkar**

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: Forbes Richest Athletes, Global Sport Finances, and NBA player salaries from the year 2000-2019.
2. How many rows (records) are in the data set?
3. How many columns (variables) are in the data set?
4. What assumptions are you making about the data?

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

	Variable	Data type
1	First name	String
2	2014 earnings	Float
3	Net salary/winnings	Float
4	Earnings from endorsements	Float
5	Type of sport	String
6	Last Name	String

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

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: Forbes Richest Athletes
2. How many rows (records) are in the data set? 301
3. How many columns (variables) are in the data set? 7
4. What assumptions are you making about the data? Assumptions that I can make about the earnings is that this data took into consideration every athlete in the world and their earnings while recording the data. From looking at the names of the athlete I can assume that the data set only recorded the earnings of male athletes and not female athletes. From looking at the sports variable I can assume that the data set only took the popular sports athletes into account and not the least popular sports.

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

	Variable	Data type
1	First Name	String
2	Nationality	String
3	Rank	Integer
4	Type of Sport	String
5	Year	Integer
6	Earnings	Float
7	Last Name	String
8	Middle Initial/Suffix	String

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: NBA Players' Salaries from the year 2000-2019
2. How many rows (records) are in the data set? 37420
3. How many columns (variables) are in the data set? 5
4. What assumptions are you making about the data? Assumptions that I can make about the dataset is that while recording the salaries of the athletes the data set took into consideration their whole net income, which included any additional sources of income and not just their earnings from NBA. Another assumption I can make is that the rank of the athlete is based on how much they earned and not how good they are at their respectable sport.

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

	Variable	Data type
--	----------	-----------

1	First Name	String
2	Last Name	String
3	Salaries	Integer
4	Rank	Integer
5	Year	Integer

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

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