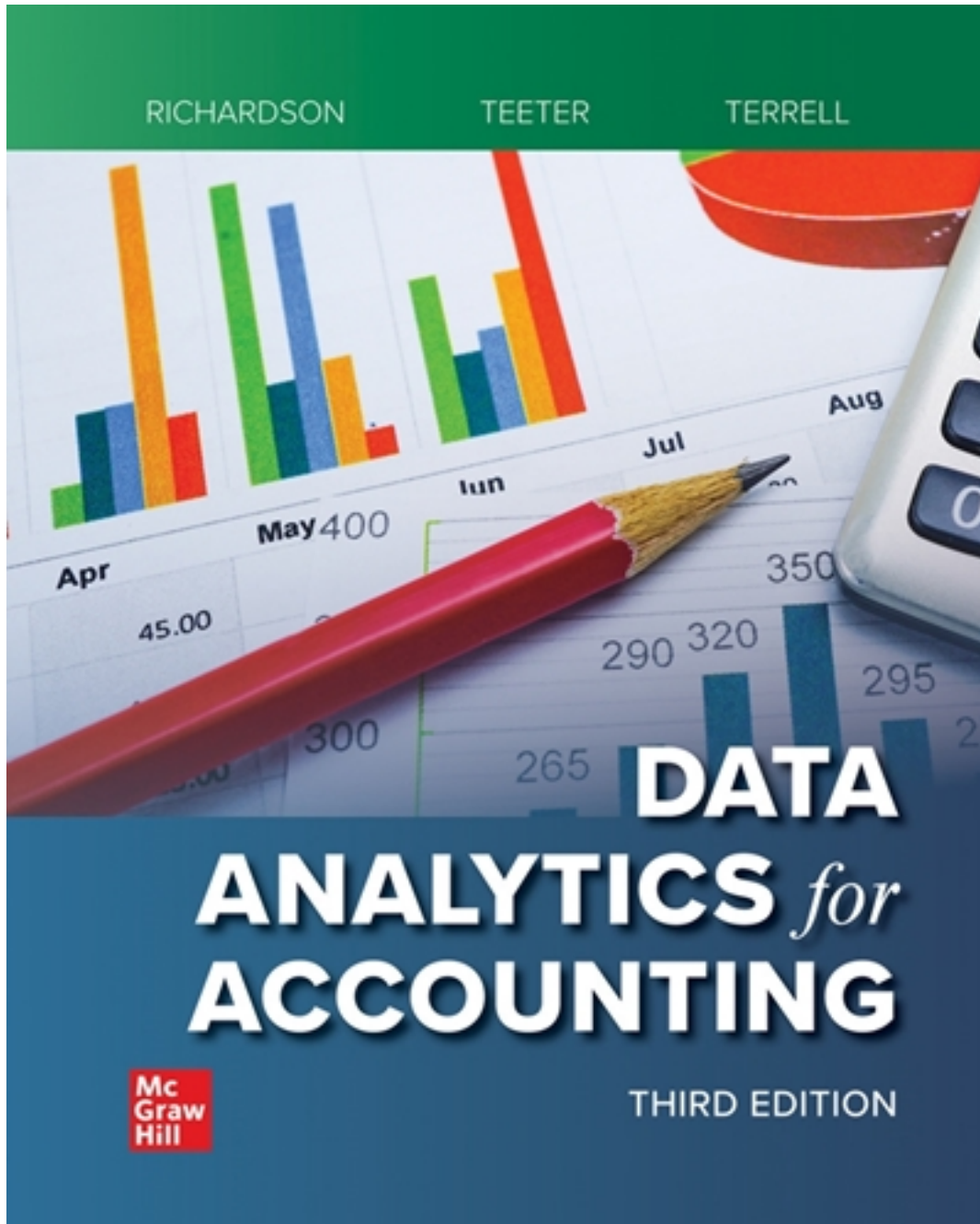


# Solutions for Data Analytics for Accounting 3rd Edition by Richardson

[CLICK HERE TO ACCESS COMPLETE Solutions](#)



# Solutions

## Solutions Manual – Chapter 2

### Solutions to Multiple Choice Questions

1. (LO 2-3) Mastering the data can also be described via the ETL process. The ETL process stands for:

- a. extract, total, and load data.
- b. enter, transform, and load data.
- c. extract, transform, and load data.
- d. enter, total, and load data.

Answer: C

2. (LO 2-3) Which of the following describes part of the goal of the ETL process?

- 1. Identify which approach to Data Analytics should be used.
- 2. Load the data into a relational database for storage.
- 3. Communicate the results and insights found through the analysis.
- 4. Identify and obtain the data needed for solving the problem.

Answer: D

3. (LO 2-2) The advantages of storing data in a relational database include which of the following?

- a. Help in enforcing business rules.
- b. Increased information redundancy.
- c. Integrating business processes.
- d. Help in enforcing business rules and integrating business processes.

Answer: D

4. LO 2-3) The purpose of transforming data is:

- a. to validate the data for completeness and integrity.
- b. to load the data into the appropriate tool for analysis.
- c. to obtain the data from the appropriate source.
- d. to identify which data are necessary to complete the analysis.

Answer: A

5. (LO 2-2) Which attribute is required to exist in each table of a relational database and serves as the “unique identifier” for each record in a table?

- a. Foreign key

- b. Unique identifier
- c. Primary key
- d. Key attribute

Answer: C

6. (LO 2-2) The metadata that describe each attribute in a database is which of the following?

- a. Composite primary key
- b. Data dictionary
- c. Descriptive attributes
- d. Flat file

Answer: B

7. (LO 2-3) As mentioned in the chapter, which of the following is *not* a common way that data will need to be cleaned after extraction and validation?

- a. Remove headings and subtotals.
- b. Format negative numbers.
- c. Clean up trailing zeroes.
- d. Correct inconsistencies across data.

Answer: C

TBEXAM.COM

8. (LO 2-2) Why is Supplier ID considered to be a primary key for a Supplier table?

- a. It contains a unique identifier for each supplier.
- b. It is a 10-digit number.
- c. It can either be for a vendor or miscellaneous provider.
- d. It is used to identify different supplier categories.

Answer: A

9. (LO 2-2) What are attributes that exist in a relational database that are neither primary nor foreign keys?

- a. Nondiscript attributes
- b. Descriptive attributes
- c. Composite keys
- d. Relational table attributes

Answer: B

10. (LO 2-4) Which of the following questions are not suggested by the Institute of Business Ethics to allow a business to create value from data use and analysis, and still protect the privacy of stakeholders?

- How does the company use data, and to what extent are they integrated into firm strategy?
- Does the company send a privacy notice to individuals when their personal data are collected?
- Does the data used by the company include personally identifiable information?
- Does the company have the appropriate tools to mitigate the risks of data misuse?

Answer: C

### Solutions to Discussion and Analysis Questions

- The information needs only to be entered once and changes or edits only need to be done in one file versus multiple files. It won't take up unnecessary space (which is expensive), take up unnecessary processing to run reports to ensure that there aren't multiple versions of the truth, and will not increase the risk of data entry errors.
- Relational databases are designed to support business processes across the organization, which results in improved communication across functional areas and more integrated business processes. Greater integration of data and business processes typically means that there will be less duplication since functional areas of the organization rely on a lot of the same information. Further integrating data will usually result in a unified/consistent data design strategy so that users do not need to learn many different data schemas and approaches (e.g., same attributes named differently or in different formats).
- Relational databases all connect with each other by use of the primary and foreign key. That makes data analysis very easy to do since you can readily join the tables and run the requested data analysis. There are a variety of reasons that it is beneficial to go to the trouble to store data in a relational database:
  - Ensuring data integrity from a standpoint of completeness of data and reduced redundancy of data
  - Enforcing business rules through internal controls. Relational databases allow for automating preventive controls and maintaining data so that detective controls can be run.
  - Integrating business processes by having all of the data stored in one centralized location, which can improve communication across the organization.
- Relational databases can be designed to aid in the placement and enforcement of internal controls and business rules in ways that flat files cannot. Due to the nature of the primary key/foreign key, both a primary key and a foreign key must line up with each other before any business can be transacted. For example, to enforce the business rule that all purchase orders must be processed with an approved supplier, in a relational database the purchase order table

and the suppliers table would be related, thus creating a preventive control to keep a purchase order from being created with a supplier that is not in the approved supplier table.

5. The data dictionary is a centralized repository of descriptions for all of the data attributes of the data set. Attributes of a data dictionary for each field might include a variable name, a brief description, whether the field is made up of numbers or text or alphanumerics, the size (or number of digits) of the field, whether it serves as a primary or foreign key and notes, etc.

It is particularly useful during the “Master the Data” steps because it describes what data is stored, how it is stored, and how different tables are related. This information is critical when extracting data from a database.

6. Before extracting the data, it is important to be able to answer these questions:
  - a. What is the purpose of the data request? What do you need the data to solve? What business problem will it address?
  - b. What risk exists in data integrity (e.g., reliability, usefulness)? What is the mitigation plan?
  - c. What other information will impact the nature, timing and extent of the data analysis?

When you begin to extract the data, you need to do the following:

- a. Identify the tables that contain the information you need.
  - b. Identify which attributes hold the information you need in each table.
  - c. Identify how those tables are related to each other.
7. The analyst needs to know what data is available, how it is organized, what is possible to include, and how reliable the data is to be able to answer the central question which was the reason for the analysis. The better that the analysts understands the raw data and how it is organized, the better they can formulate questions to expedite the data request and communication process with those that will do the extraction
8. The more frequent the requested report, the more the database administrator will set it up for automatic extraction and delivery. It may also be a question of how often the data changes. If the data is updated weekly and the data is extracted daily, that may not make any sense.
9. The database administrator is most familiar with the data and may be able to help the analyst get the data needed to address the question. There also might be some sensitivities to who gets what data to ensure that the data gets to the intended analyst and audience. The analyst may suggest other data or data sources that the data requestor is not aware of to help answer that question.

10. The impact of transforming data to work with NULL, N/A and zero values in the dataset might have an impact on programs like Excel.
- Transforming NULL and N/A values into blanks.
    - The COUNT and AVERAGE functions would not include these fields in their computation for these variables.
  - Transforming NULL and N/A values into zeroes.
    - The COUNT and AVERAGE functions would incorporate these zeroes and would be included in their computation for these variables. It would have an impact particularly on the computation of the average since it would have the value of zero.
  - Deleting records that have NULL and N/A values from your dataset.
    - The COUNT and AVERAGE functions would not include these fields in their computation for these variables. If they are deleted all of the other fields and variables would be deleted as well, thus having a bigger impact on the overall dataset.
11. The first of the six questions suggested by the Institute of Business Ethics looks to assess whether the purpose of the data is to enhance firm strategy. Questions 2-5 assess the company's data collection and data use and whether potentially sensitive data is adequately protected from misuse or privacy breach. The final question asks whether the company's third party data providers follow similar ethical standards.

As a reminder, the six questions introduced in the chapter are as follows:

- How does the company use data, and to what extent is it integrated into firm strategy?** What is the purpose of the data? Is it accurate or reliable? Will it benefit the customer or the employee?
- Does the company send a privacy notice to individuals when their personal data is collected?** Is the request to use the data clear to the user? Do they agree to the terms and conditions of use of their personal data?
- Does the company assess the risks linked to the specific type of data the company uses?** Have the risks of data use or data breach of potentially sensitive data been considered?

- **Does the company have safeguards in place to mitigate these risks of misuse?** Are preventive controls on data access in place and are they effective? Are penalties established and enforced for data misuse?
- **Do the company have the appropriate tools to manage the risks of misuse?** Is the feedback from these tools evaluated and measured? Does internal audit regularly evaluate these tools?
- **Does our company conduct appropriate due diligence when sharing with or acquiring data from third parties?** Do third party data providers follow similar ethical standards in the acquisition and transmission of the data?

One way to mitigate the risks associated with data use and collection is to understand if there is a purpose of the data use it will reduce the likelihood of collecting data that doesn't have a known business use and reduce the extent of an impact should the data be breached."

#### Solutions to Problems

1. (LO 2-2) Match the relational database function to the appropriate relational database term:

- Composite primary key
- Descriptive attribute
- Foreign key
- Primary key
- Relational database

TBEXAM.COM

Relational Database Function	Relational Database Term
1. Serves as a unique identifier in a database table.	Primary key
2. Creates a relationship between two tables	Foreign key
3. Two foreign keys from the tables that it is linking combine to make up a unique identifier	Composite primary key
4. Describe each record with characteristics with actual business information	Descriptive attributes
5. A means of storing data to ensure data are complete, redundant, to help enforce business rules.	Relational database

2. (LO 2-3) Identify the order sequence in the ETL process as part of mastering the data (i.e., 1 is first; 5 is last).

Steps of the ETL Process	Sequence Order (1 to 5)
Validate the data for completeness and integrity.	3
Sanitize the data.	4
Obtain the data.	2
Load the data in preparation for data analysis.	5
Determine the purpose and scope of the data request.	1

3. (LO 2-3) Identify which ETL tasks would be considered “Validate” the data, and which would be considered “Cleaning” the data.

ETL Task	Validate or Clean?
1. Compare the number of records that were extracted to the number of records in the source database	Validate
2. Remove headings or subtotals	Clean
3. Remove leading zeroes and nonprintable characters	Clean
4. Compare descriptive statistics for numeric fields	Validate
5. Format negative numbers	Clean
6. Compare string limits for text fields	Validate
7. Correct inconsistencies across data, in general	Clean

4. (LO 2-3) Match each ETL task to the stage of the ETL process:

- determine purpose
- obtain
- validate
- clean
- load

ETL Task	Stage of ETL Process
1. Use SQL to extract data from the source database	Obtain
2. Remove headings or subtotals	Clean
3. Choose which database and specific data will be needed to address the accounting question	Determine purpose



4. Compare the number of records extracted to the number of records in the source database.	Validating
5. Make sure all data formats start with two capital letters. Fix inconsistencies.	Clean
6. Input the data into the analysis tool.	Load

5. (LO 2-4) For each of the six questions suggested by the Institute of Business Ethics to evaluate data privacy, categorize each question into one of these three types:
1. Evaluate the company's purpose of the data
  2. Evaluate the company's use or misuse of the data
  3. Evaluate the due diligence of the company's data vendors in preventing misuse of the data.

Institute of Business Ethics Questions Regarding Data Use and Privacy	Types of Data Privacy/Use Question By Number
Does the company assess the risks linked to the specific type of data the company uses?	2
Does the company send a privacy notice to individuals when their personal data are collected?	2
How does the company use data, and to what extent are they integrated into firm strategy?	1
Does our company conduct appropriate due diligence when sharing with or acquiring data from third parties?	3
Do the company have the appropriate tools to manage the risks of misuse?	2
Does the company have safeguards in place to mitigate these risks of misuse?	2

6. (LO 2-2) Which of the following are useful, established characteristics of using a relational database?

Institute of Business Ethics Questions Regarding Data Use and Privacy	Useful, Established Characteristics of Relational Databases (Y)es/(N)o
Completeness	Y
Reliable	N
No redundancy	Y

Communication and integration of business processes	Y
Less costly to purchase	N
Less effort to maintain	N
Business rules are enforced	Y

7. (LO 2-3) As part of master the data, analyst must make certain trade-offs when they consider which data to use.

Consider these three different scenarios:

- a) **Analysis:** What are the trade-offs of using data that is highly relevant to the question but have a lot of missing data?

Answers will vary, but here is a representative response: It is critically important to get data that will address the question. However, if the data has missing or erroneous data, it may not be possible to adequately address the question. Why is the data missing – is it due to collection error or errors in the ETL process? Is it possible to answer the question with the data you have, with the expectation that you can apply the findings to those that had missing observations? By addressing these questions we can determine if it is useful to analyze this data.

- b) **Analysis:** What are the trade-offs an analyst should consider between data that is very expensive to acquire and analyze, but will most directly address the question at hand? How would you assess whether they are worth the extra cost? [TBEXAM.COM](http://TBEXAM.COM)

In general, it is easy to quantify the cost of data, but much more challenging to assess the value that data provides. To the extent it is possible to quantify, the value of the data acquisition and analysis must be greater than the cost of the data acquisition and analysis.

- c) **Analysis:** What are the trade-offs between extracting needed data by yourself, or asking a data scientist to get access to the data?

Sometimes it is easier for the accountant to access the data herself, rather than to adequately communicate to a data scientist exactly what data is needed. On the other hand, to the extent that the data scientist knows the data more intimately than the accountant or is more efficient at extracting the data, the task of extracting data might be better left to the data scientist.

8. (LO 2-4) The Institute of Business Ethics proposes that a company protect the privacy of stakeholders by considering these questions of its third-party data providers:

- Does our company conduct appropriate due diligence when sharing with or acquiring data from third parties?
- Do third-party data providers follow similar ethical standards in the acquisition and transmission of the data?

**a) Analysis:** What type of due diligence with regards to a third party sharing and acquiring data would be appropriate for the company (or company accountant or data scientist) to perform? An audit? A questionnaire? Standards written in to a contract?

The accountant or data scientist may begin the conversation by talking through the process with the third-party data provider. Asking the data provider specifically what they do to maintain privacy throughout the process of acquisition and transmission of data will allow the company to determine if more specific steps, such as a data audit or privacy concerns written into a contract, are necessary.

**b) Analysis:** How would you assess whether the third-party data provider follows ethical standards in the acquisition and transmission of the data? An audit? A questionnaire? Standards written in to a contract?

The accountant or data scientist may begin the conversation by talking (such as interviewing, observation, questionnaires, etc.) through the process with the third-party data provider. Asking the data provider specifically how they acquire and transmit data and how they work to keep ethical standards to determine if more specific steps, such as a data audit or privacy concerns written into a contract, are necessary.

TBEXAM.COM

## Solutions Manual – Chapter 2 Labs

### Lab 2-1 Request Data from IT—Sláinte

#### Lab 2-1 Part 1 Identify the Questions and Generate a Data Request

##### Lab 2-1 Part 1 Analysis Questions (LO 2-2)

AQ1. Given that you are new and trying to get a grasp on Sláinte's operations, list three questions related to sales that would help you begin your analysis. For example, how many products were sold in each state?

*Answer: Answers will vary. Some examples may include:*

- "How many products were sold in each state?"
- "What was the sales volume in the first quarter?"

AQ2. Now hypothesize the answers to each of the questions. Remember, your answers don't have to be correct at this point. They will help you understand what type of data you are looking for. For example: 500 in Missouri, 6,000 in Pennsylvania, 4,000 in New York, and so on.

*Answer: Answers will vary. Some examples may include:*

- 500 in Missouri, 6,000 in Pennsylvania, 4,000 in New York
- 100,000 in Q1

TBEXAM.COM

AQ3. Finally, for each question, identify the specific tables and fields that are needed to answer your questions. Use the data dictionary and ER Diagram provided in Appendix J for guidance on what tables and attributes are available. For example, to answer the question about state sales, you would need the Customer\_State attribute that is located in the Customer master table as well as the Sales\_Order\_Quantity\_Sold attribute in the Sales table. If you had access to store or distribution center location data, you may also look for a State field there, as well.

*Answer: Answers will vary. Some examples may include:*

- Customer\_State attribute that is in the Customer master table as well as the Sales\_Order\_Quantity\_Sold attribute in the Sales table.
- Customer\_State attribute that is in the Customer master table as well as the Sales\_Order\_Quantity\_Sold and Sales\_Order\_Date attributes in the Sales table.

#### Lab 2-1 Part 2 Evaluate the Data Extract

##### Lab 2-1 Part 2 Objective Questions (LO 2-2)

OQ1. Which tables and attributes are missing from the data extract that would be necessary to answer the question "How many products were sold in each state?"

- A) Customer\_State
- B) Product\_Description

## Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

C) *Sales\_Order\_Quantity\_Sold*

D) *Product\_Sale\_Price*

Answer: A

OQ2. What new question can you answer using the data extract?

A) *Which product categories are the most popular by sales volume?*

B) *Which customers spent the most money or orders?*

C) *What was to total shipping cost of the orders?*

D) *How many products were sold in each state?*

Answer: C

### Lab 2-1 Part 2 Analysis Questions (LO 2-2)

AQ1. Evaluate your original questions and responses from Part 1. Can you still answer the original questions that you identified in step 1 with the data provided?

*Answers may be yes or no, depending on the attributes selected by the students.*

AQ2. What additional tables and attributes would you need to answer your questions?

*Answers will vary but may include:*

- *Customer data – name, address, state, zip*
- *Product data – categories, size, weight*
- *Employee data – name, position*

### Lab 2-1 Submit Your Screenshot Lab Document

Verify that you have answered any questions your instructor has assigned, then upload your screenshot lab document to Connect or to the location indicated by your instructor.

## Lab 2-2 Prepare Data for Analysis—Sláinte

### Lab 2-2 Part 1 Prepare a Data Model

#### Lab 2-2 Part 1 Objective Questions (LO 2-3)

OQ1. How many tables did you just load?

- A) 2
- B) 3
- C) 4
- D) 5

Answer: B

OQ2. How many rows were loaded for the Sales\_Order query?

Answer: 569

OQ3. How many rows were loaded for the Finished\_Goods\_Products query?

- A) 4
- B) 5
- C) 6
- D) 7

Answer: C

### Lab 2-2 Part 1 Analysis Questions (LO 2-3)

AQ1. Have you used the Microsoft or Tableau tools before this class?

Answer: Answers may be yes or no, depending on students' previous experiences.

AQ2. Compare and Contrast: If you completed this part with multiple tools, which tool options do you think will be most useful for preparing future data for analysis?

Answer: Answers will vary depending on students' perceptions of the software tools and other external influences, such as marketing and discussion with recruiters.

### Lab 2-2 Part 2 Validate the Data

#### Lab 2-2 Part 2 Objective Questions (LO 2-3)

OQ1. What was the total sales volume for Imperial Stout in January 2020?

Answer: 2411

OQ2. What was the total sales volume for all products in January 2020?

Answer: 2295

## Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

OQ3. Which product is experiencing the worst sales performance in January 2020?

- A) *Imperial IPA*
- B) *IPA*
- C) *Pale Ale*
- D) *Wheat*

Answer: A

### Lab 2-2 Part 2 Analysis Questions (LO 2-3)

Now that you've completed a basic analysis to answer management's question, take a moment to think about how you could improve the report and anticipate questions your manager might have.

AQ1. If the owner of Sláinte wishes to identify which product sold the most, how would you make this report more useful?

Answer: *Sort the data in descending order by sales volume.*

AQ2. If you wanted to provide more detail, what other attributes would be useful to add as additional rows or columns to your report, or what other reports would you create?

Answer: *Answers will vary but may include discussion of location data or aggregate data, such as average order size.*

AQ3. Write a brief paragraph about how you would interpret the results of your analysis in plain English. For example, which data points stand out?

Answer: *Answers will vary but should focus on the Imperial Stout as being a top performer, or that sales dropped in April overall.*

AQ4. In Chapter 4, we'll discuss some visualization techniques. Describe a way you could present these data as a chart or graph.

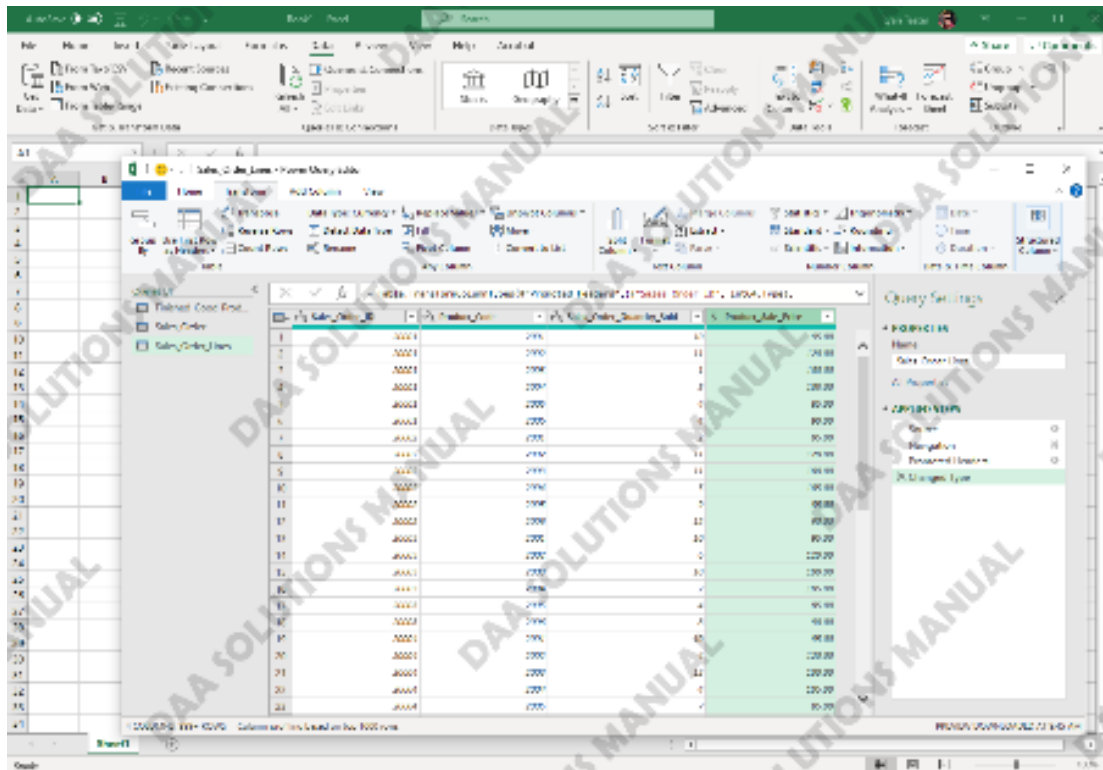
Answer: *Answers will vary but may include a description of bar charts to show relative size, or line graphs to show trends over time.*

### Lab 2-2 Submit Your Screenshot Lab Document

Verify that you have answered any questions your instructor has assigned, then upload your screenshot lab document to Connect or to the location indicated by your instructor.

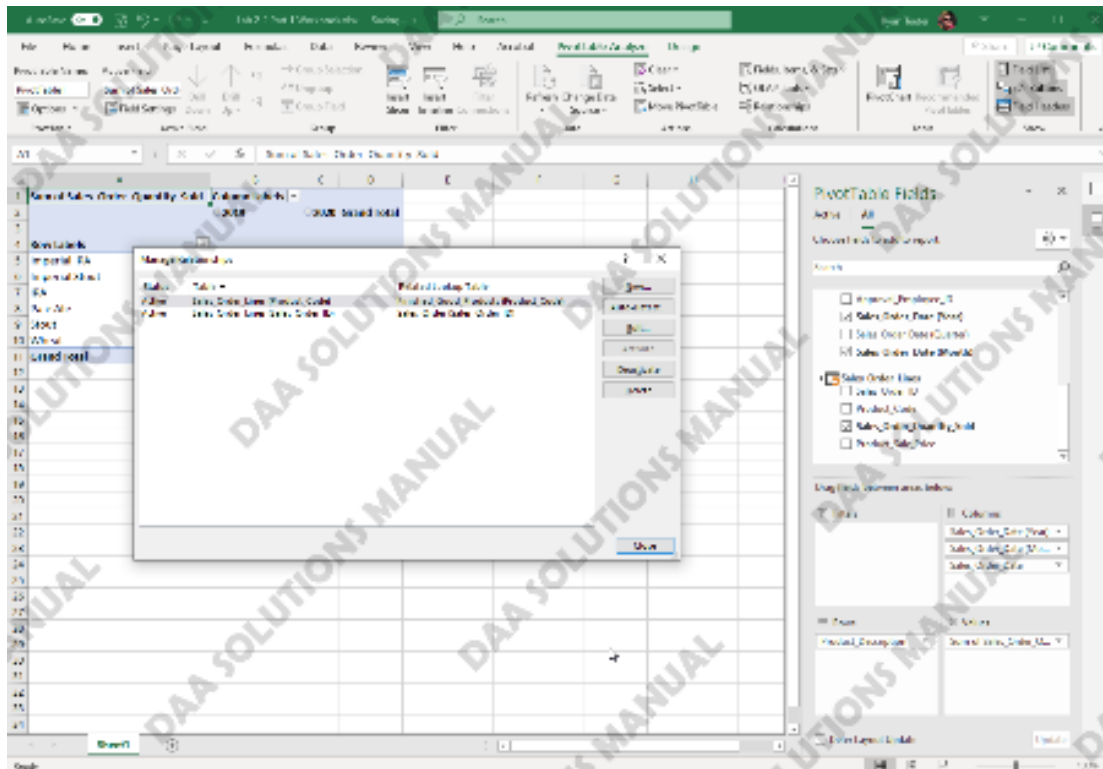
# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-2MA



2-2MB

TBEXAM.COM





# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-2MC

The screenshot shows an Excel spreadsheet with a PivotTable titled 'Sales by month'. The PivotTable is structured with 'Month' as the row labels and 'Sales' as the column labels. The data is summarized by month, showing sales figures for each month. On the right side, the 'PivotTable Task' pane is open, showing options for 'Analyze Data Using This Report's Data' and 'Show Data as a Table'. The 'Show Data as a Table' section is checked, and the 'Table Name' is set to 'SalesTable'. The 'Table Range' is set to '\$A\$1:\$D\$11'. The 'Table Options' section is also visible, showing options for 'Table Style' and 'Table Name'.

Month	Sales
Jan	100
Feb	150
Mar	200
Apr	250
May	300
Jun	350
Jul	400
Aug	450
Sep	500
Oct	550
Nov	600
Dec	650

2-2TA

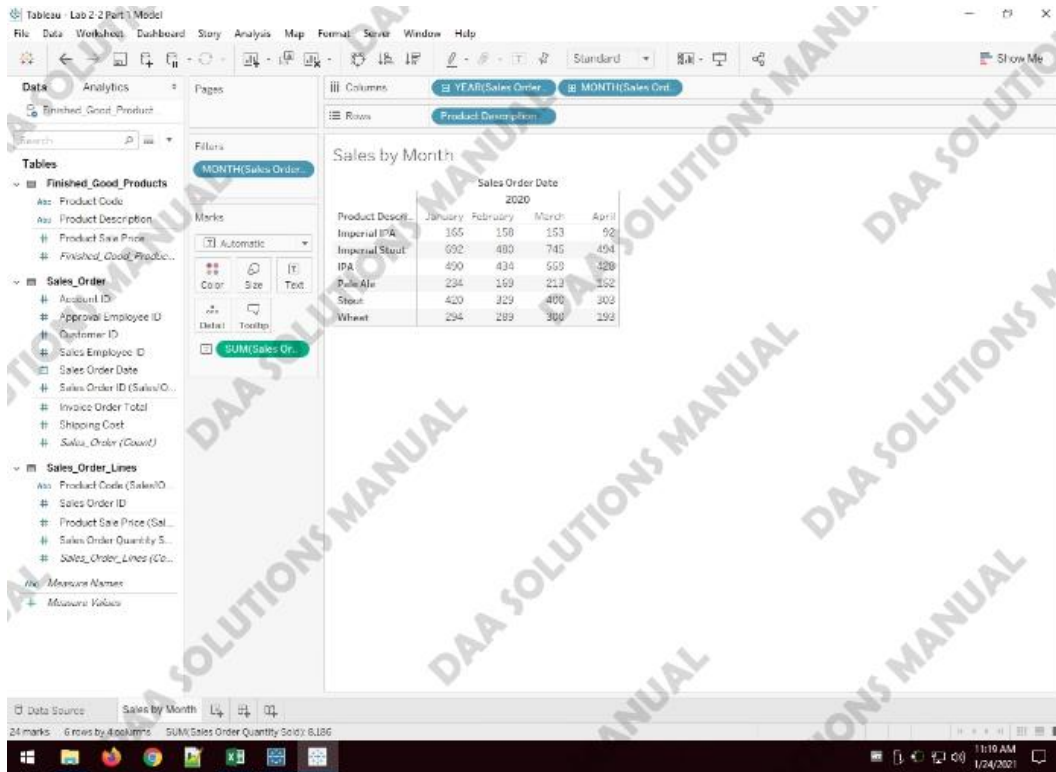
TBEXAM.COM

The screenshot shows the Tableau Desktop interface. The 'Connections' pane on the right lists the data source 'Finished\_Good\_Products'. The main view displays a table of data with columns for 'Sales Order ID', 'Sales Order Date', 'Sales Order Total', 'Customer ID', 'Account ID', and 'Shipping Cost'. The data is sorted by 'Sales Order ID' in descending order.

Sales Order ID	Sales Order Date	Sales Order Total	Customer ID	Account ID	Shipping Cost
50002	11/15/2019	312	1000	10000004	150.050
50001	11/15/2019	490	1000	10000004	25.400
50000	11/15/2019	3730	1000	10000004	140.930
50000	11/15/2019	1790	1000	10000004	82.930
50000	11/15/2019	5090	1000	10000004	113.840
50000	11/15/2019	1710	1000	10000004	128.840
50000	11/15/2019	8982	1000	10000004	101.560
50000	11/15/2019	4750	1000	10000004	21.500
50000	11/15/2019	3940	1000	10000004	78.830
50000	11/15/2019	4790	1000	10000004	144.630
50000	11/15/2019	3798	1000	10000004	81.760

## Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-2TB



TBEXAM.COM

## Lab 2-3 Resolve Common Data Problems—LendingClub

### Lab 2-3 Part 1 Identify Relevant Attributes

#### Lab 2-3 Part 1 Objective Questions (LO 2-3)

OQ1. Which attributes would you expect to contain date values?

- A) *term*
- B) *emp\_length*
- C) *earliest\_cr\_line*
- D) *delinq\_2y*

Answer: C

OQ2. Which attributes would you expect to contain text values?

- A) *grade*
- B) *emp\_length*
- C) *delinq\_2y*
- D) *total\_acc*

Answer: A

OQ3. Which attributes would you expect to contain numerical values?

- A) *grade*
- B) *emp\_length*
- C) *issue\_d*
- D) *earliest\_cr\_line*

Answer: B

OQ4. Which attribute most directly impacts a borrower's cost of capital?

- A) *loan\_amnt*
- B) *revol\_bal*
- C) *total\_acc*
- D) *int\_rate*

Answer: D

### Lab 2-3 Part 1 Analysis Questions (LO 2-3)

AQ1. What do you expect will be major data quality issues with LendingClub's data?

**Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e**

*Answer: Answers will vary, but some examples might be missing data, incorrect text values, errors in the data, etc.*

AQ2. Given this list of attributes, what types of questions do you think you could answer regarding approved loans? (If you worked through Lab 1-2, what concerns do you have with the data's ability to predict answers to the questions you identified in Chapter 1)?

*Answer: Answers will vary, let's some example questions might be "What is the average number of outstanding accounts?", "Where are borrowers located?", "What is the average interest rate?", etc.*

**Lab 2-3 Part 2 Transform and Clean the Data****Lab 2-3 Part 2 Objective Questions (LO 2-3)**

OQ1. How many records or rows appear in your cleaned dataset?

*Answer: 235,633*

OQ2. How many attribute or columns appear in your cleaned dataset?

*Answer: 20*

**Lab 2-3 Part 2 Analysis Questions (LO 2-3)**

AQ1. Why do you think it is important to remove text values from your data before you conduct your analysis?

*Answer: Answers will vary but text can cause problems with numerical data that we want to summarize.*

AQ2. What do you think would happen in your analysis if you didn't remove the text values?

*Answer: Aggregate values will not be calculated correctly.*

AQ3. Did you run into any major issues when you attempted to clean the data? How did you resolve those?

*Answer: Answers will vary but may include incomplete data or missing attributes. Some solutions may include replacing missing attributes with a value or importing the file again with a different interpreter.*

AQ4. What are some steps you could take to clean the data and resolve the difficulties you identified?

*Answer: Answers will vary but may include replacing values, summarizing data, or removing attributes.*

**Lab 2-3 Submit Your Screenshot Lab Document**

Verify that you have answered any questions your instructor has assigned, then upload your screenshot lab document to Connect or to the location indicated by your instructor.

# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-3MA

id	addr_state	delinq_12m	delinq_24m	mths_since_last_delinq	mths_since_last_inquiry
1	VA	22.05	0	8/2/2014	
2	CA	14.92	0	9/2/1989	
3	CA	8.21	0	10/2/2000	
4	MD	28.40	0	8/2/2005	
5	NY	25.81	0	11/2/1990	
6	AZ	34.81	0	8/2/2007	
7	MM	15.77	0	10/2/2003	
8	NY	23.63	0	1/2/2001	
9	IL	24.5	0	11/2/1987	
10	FL	22.42	0	1/2/2009	
11	MD	14.05	0	5/2/2009	
12	MI	8.49	0	9/2/2003	
13	WA	27.05	1	10/2/1998	
14	VA	11.62	1	11/2/2000	
15	KY	0.87	1	7/2/1988	
16	CA	0.54	0	8/2/2000	
17	AZ	18.39	2	10/2/1999	
18	NY	27.14	0	9/2/1987	
19	FL	3.98	0	2/2/2010	
20	TX	18.79	0	6/2/2006	
21	GA	10.23	0	7/2/2007	
22	TX	35.68	0	11/2/1999	

2-3MB

TBEXAM.COM

id	addr_state	delinq_12m	delinq_24m	mths_since_last_delinq	mths_since_last_inquiry
1	VA	22.05	0	8/2/2014	
2	CA	14.92	0	9/2/1989	
3	CA	8.21	0	10/2/2000	
4	MD	28.40	0	8/2/2005	
5	NY	25.81	0	11/2/1990	
6	AZ	34.81	0	8/2/2007	
7	MM	15.77	0	10/2/2003	
8	NY	23.63	0	1/2/2001	
9	IL	24.5	0	11/2/1987	
10	FL	22.42	0	1/2/2009	
11	MD	14.05	0	5/2/2009	
12	MI	8.49	0	9/2/2003	
13	WA	27.05	1	10/2/1998	
14	VA	11.62	1	11/2/2000	
15	KY	0.87	1	7/2/1988	
16	CA	0.54	0	8/2/2000	
17	AZ	18.39	2	10/2/1999	
18	NY	27.14	0	9/2/1987	
19	FL	3.98	0	2/2/2010	
20	TX	18.79	0	6/2/2006	
21	GA	10.23	0	7/2/2007	
22	TX	35.68	0	11/2/1999	

## Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-3TA

Tableau Desktop interface showing the 'LoanStats3c' data source. The 'Columns' shelf contains 'term', 'int\_rate', 'grade', 'emp\_length', and 'home\_ownership'. The 'Rows' shelf contains 'loan\_status'. The 'Marks' card is set to 'Count'. The view displays a horizontal bar chart showing the count of loans for each term, interest rate, grade, employee length, and home ownership status, categorized by loan status.

2-3TB

TBEXAM.COM

Tableau Desktop interface showing the 'LoanStats3c' data source. The 'Columns' shelf contains 'term', 'int\_rate', 'grade', 'emp\_length', and 'home\_ownership'. The 'Rows' shelf contains 'loan\_status'. The 'Marks' card is set to 'Count'. The view displays a horizontal bar chart showing the count of loans for each term, interest rate, grade, employee length, and home ownership status, categorized by loan status.



## Lab 2-4 Generate Summary Statistics—LendingClub

### Lab 2-4 Calculate Summary Statistics

#### Lab 2-4 Objective Questions (LO 2-3)

OQ1. What is the maximum loan amount that was approved for borrowers from PA?

*Answer: \$35,000*

OQ2. What is the average interest rate assigned to a loan to an approved borrower from PA?

*Answer: should be between 12% and 15%*

OQ3. What is the average annual income of an approved borrower from PA?

*Answer: should be between \$70,000 and \$80,000*

#### Lab 2-4 Analysis Questions (LO 2-3)

AQ1. Compare the loan amounts to the validation given by LendingClub for borrowers from PA:

Funded loans: \$123,262. 53

Number of approved loans: 8,427

Do the numbers in your analysis match the numbers provided by LendingClub? What explains the discrepancy, if any?

*Answer: Numbers should match. If there is a discrepancy, it is because the students are missing a cleaning step. They should re-perform the import and cleaning process.*

AQ2. Does the Numerical Count provide a more useful/accurate value for validating your data? Why or why not do you think that is the case?

*Answer: Answers will vary, but a numerical count should help understand whether the data is complete or missing values.*

AQ3. Compare and contrast: Why does Power Query and Tableau Desktop return different values for their summary statistics?

*Answer: Depending on the tool used for analysis, summary statistics may be calculated using a sample of data, for example the first 1000 records. The different tools will select each sample differently.*

AQ4. Compare and contrast: What are some of the summary statistics measures that are unique to Power Query? To Tableau Desktop?

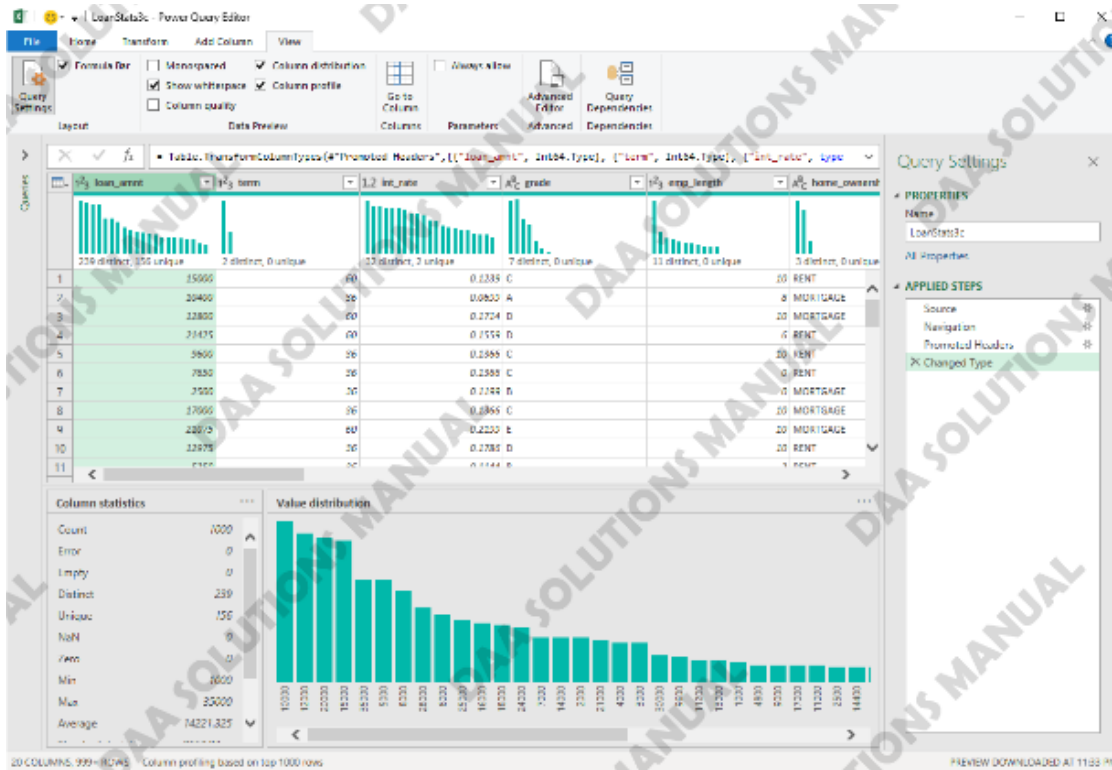
*Answer: Power query calculates the number of distinct and unique values. Tableau desktop calculates the median value.*

#### Lab 2-4 Submit Your Screenshot Lab Document

Verify that you have answered any questions your instructor has assigned, then upload your screenshot lab document to Connect or to the location indicated by your instructor.

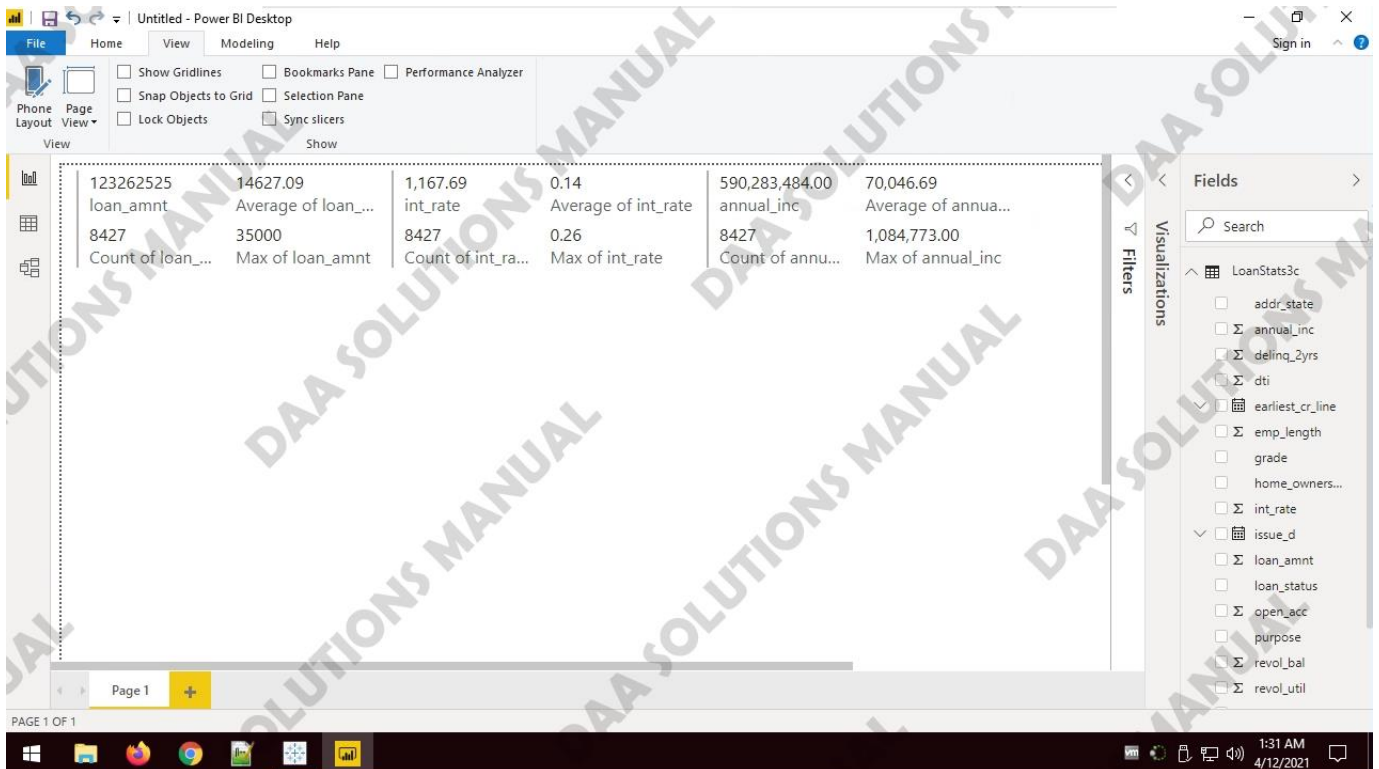
# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-4MA



2-4MB

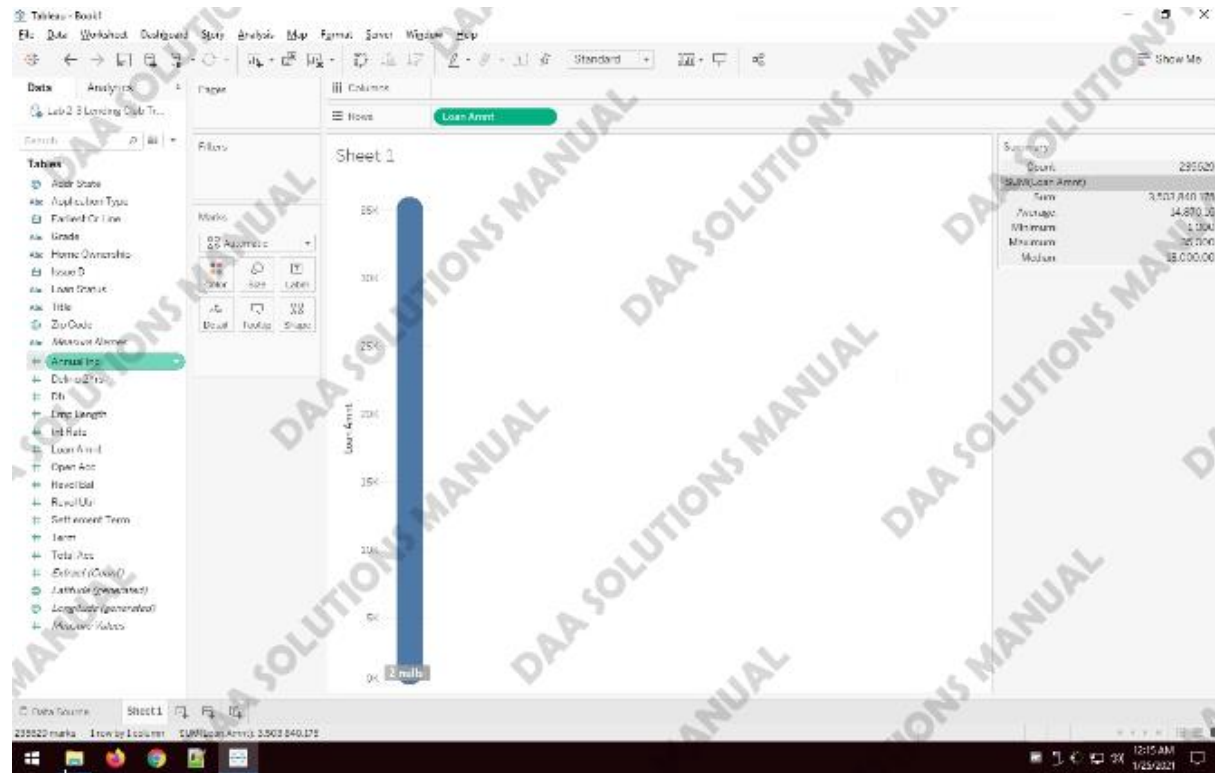
TBEXAM.COM





# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-4TA



## Lab 2-5 Validate and Transform Data—College Scorecard

### Lab 2-5 Load and Clean Data

#### Lab 2-5 Objective Questions (LO 2-3)

OQ1. How many schools report average SAT scores?

*Answer: 1304*

OQ2. What is the average completion rate (C150 4) of all of the schools?

*Answer: 0.478*

OQ3. How many schools report data to the U. S. Department of Education?

*Answer: 7703*

#### Lab 2-5 Analysis Questions (LO 2-3)

AQ1. In the checksums, you validated that the average SAT score for all of the records is 1,059.07. When we work with the data more rigorously, several tests will require us to transform NULL or blank values. If you were to transform the NULL SAT values into 0, what would happen to the average (would it stay the same, decrease, or increase)?

*Answer: Decrease. The extra zeros would skew the average downward.*

AQ2. How would that change to the average affect the way you would interpret the data?

*Answer: Answers will vary but a common explanation might be that colleges accept students with lower SAT scores.*

AQ3. What would happen if we excluded all schools that don't report an average SAT score?

*Answer: Answers will vary but there will be a bias toward schools that report scores. We would have to make assumptions about the schools or gather additional data.*

#### Lab 2-5 Submit Your Screenshot Lab Document

Verify that you have answered any questions your instructor has assigned, then upload your screenshot lab document to Connect or to the location indicated by your instructor.

# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-5MA

Table1

	SAT_AVG	UNHS	UNHS_WHITE	UNHS_BLACK	UNHS_HISP	UNHS_ASIAN	UNHS_PAC
1	0.5256	827	4206	0.0933	0.0035	0.0018	0.0024
2	0.6360	1107	11383	0.5922	0.020	0.0518	0.0022
3	na	na	na	0.25	0.4332	0.0004	0
4	0.8202	1219	5451	0.6988	0.1225	0.0376	0.0148
5	0.5326	851	4811	0.0158	0.9208	0.0121	0.0015
6	0.5326	1185	20851	0.7225	0.2613	0.0044	0.0025
7	na	na	1592	0.7225	0.2613	0.0044	0.0025
8	na	na	2991	0.7023	0.12	0.0151	0.0053
9	0.8017	1099	4304	0.5338	0.3376	0.0024	0.0044
10	0.8347	1213	20514	0.8507	0.0704	0.0348	0.0227
11	0.5339	1181	1180	0.7983	0.1103	0.0135	0.0102
12	na	na	1727	0.4651	0.4372	0.0492	0.0023
13	0.5328	842	322	0.028	0.3756	0.0373	0.0091
14	na	na	522	0.3046	0.8094	0.0153	0.0153
15	na	na	1729	0.4408	0.2435	0.0509	0.0302
16	na	na	4206	0.0979	0.2295	0.032	0.0084
17	na	na	2367	0.3839	0.5137	0.0298	0.0083
18	na	na	4917	0.6921	0.2076	0.0306	0.0047
19	na	na	115	0.8957	0.0345	0.0056	0
20	na	na	4273	0.8273	0.324	0.0241	0.0072
21	na	na	4820	0.863	0.0612	0.0362	0.0065
22	na	na	4820	0.863	0.0612	0.0362	0.0065
23	na	na	1513	0.1956	0.7445	0.0026	0.004
24	0.5889	1076	1149	0.6988	0.1891	0.0212	0.0078
25	na	na	82	0.7019	0.3125	0.0384	0
26	na	na	976	0.5337	0.5965	0.0246	0.0082
27	na	na	1041	0.7020	0.6834	0.2382	0.0132
28	na	na	1013	0.5897	0.7983	0.0077	0.0252
29	na	na	7443	0.6896	0.2246	0.0486	0.0381
30	na	na	5830	0.0528	0.184	0.0435	0.0257
31	na	na	176	0.7953	0.1603	0.0108	0.0053
32	na	na	2955	0.1511	0.7584	0.0053	0.0055
33	na	na	1916	0.4421	0.4055	0.0224	0.0016

2-5MB

TBEXAM.COM

AutoSaveLab 2-5 College Scorecard TransformationsManual

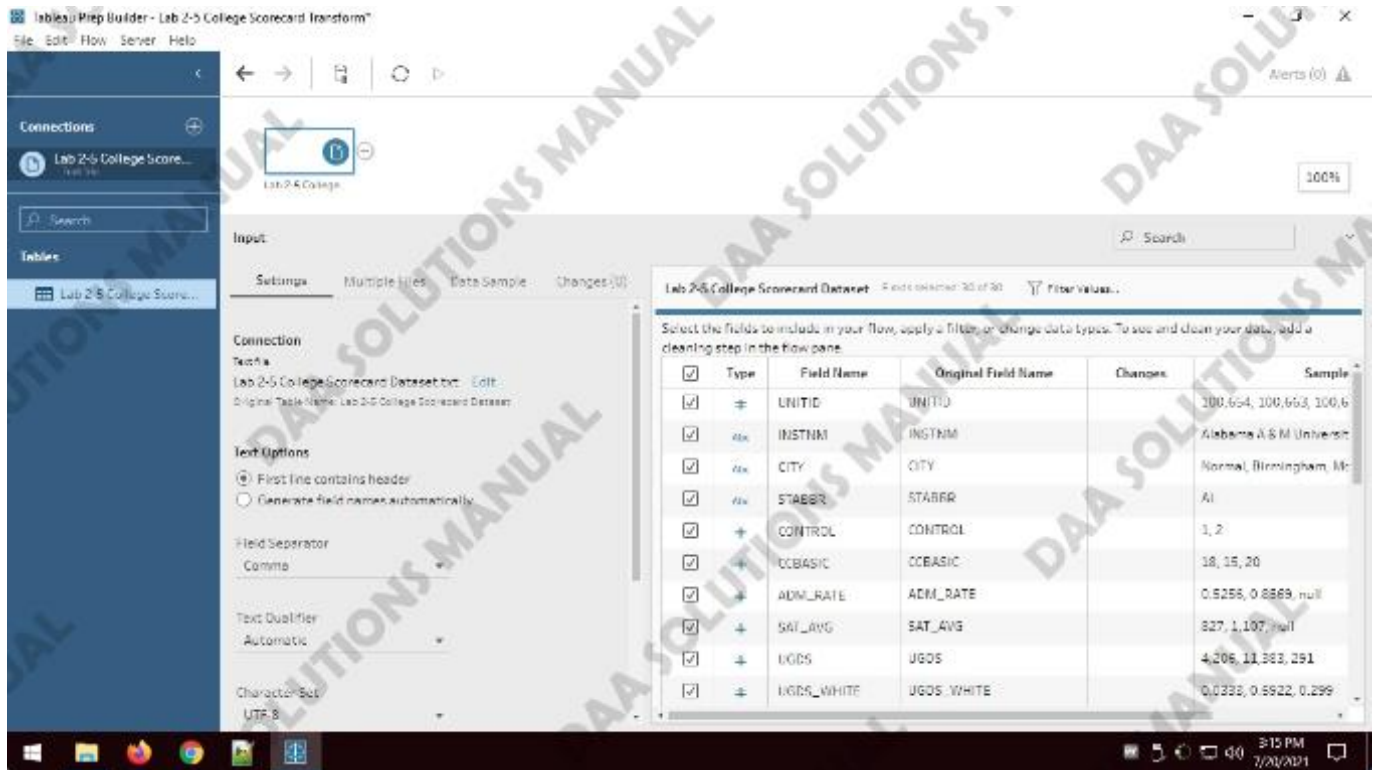
FileHomeInsertPage LayoutFormulasDataReviewViewHelpAcrobatTable DesignQuery

Table Name: Lab 2-5 College Scorecard Dataset

Summary with PivotTable, Pivot

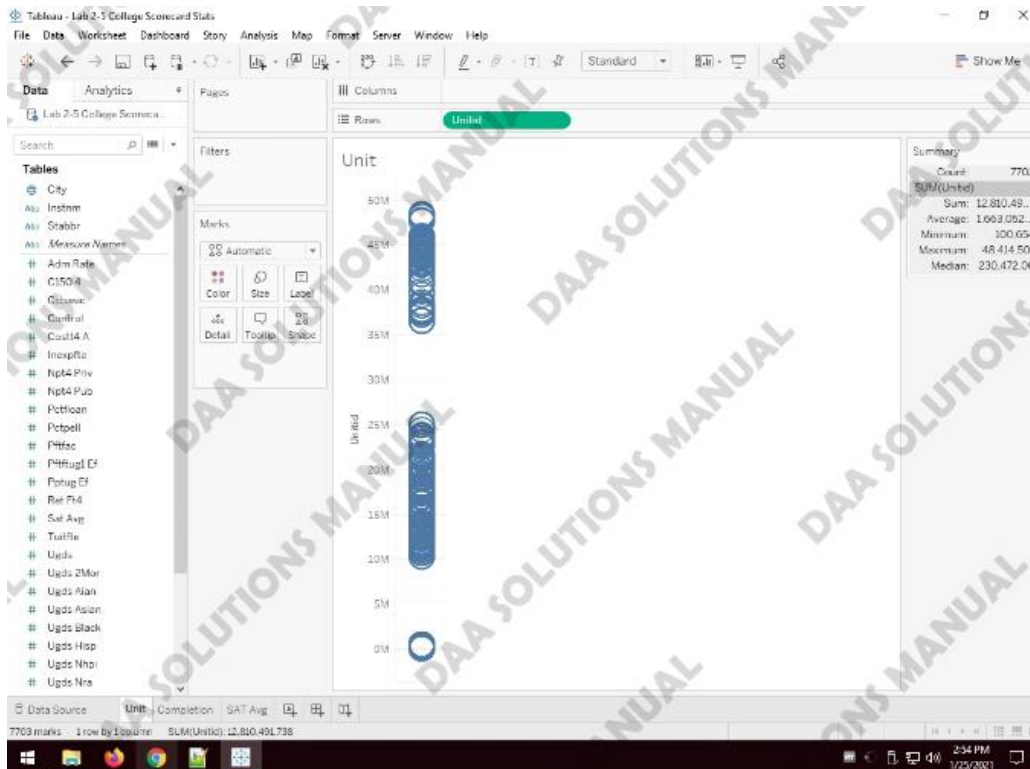
# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-5TA



2-5TB

TBEXAM.COM



## Lab 2-6 Comprehensive Case: Build Relationships Among Database Tables— Dillard's

### Lab 2-6 Build Relationships between Tables

#### Lab 2-6 Objective Questions (LO 2-2)

OQ1. How many tables relate directly to the TRANSACT table?

*Answer: 3*

OQ2. Which table does the DEPARTMENT table relate to?

- A) *Customer*
- B) *Store*
- C) *Transact*
- D) *SKU*

*Answer: D*

OQ3. What is the name of the key that relates the TRANSACT and CUSTOMER tables?

- A) *SKU*
- B) *CUST\_ID*
- C) *DEPT*
- D) *REGISTER*

*Answer: B*

#### Lab 2-6 Analysis Questions (LO 2-2)

AQ1. How would a view of the entire database or certain tables out of that database allow us to get a feel for the data?

*Answer: Answers will vary but the entire database might provide additional context to the sales data whereas a limited set of tables show only a partial picture.*

AQ2. What types of data would you guess that Dillard's, a retail store, gathers that might be useful beyond the scope of the sales data available on the remote desktop? How could Dillard's suppliers use these data to predict future purchases?

*Answer: Answers will vary but may include elements such as the timestamp of purchases, available promotions, weather conditions, store traffic, etc.*

AQ3. Compare and Contrast: Compare the methods for connecting to data in Tableau versus Power BI. Which is more intuitive?

*Answer: Answers will depend on the student's opinion.*

## Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

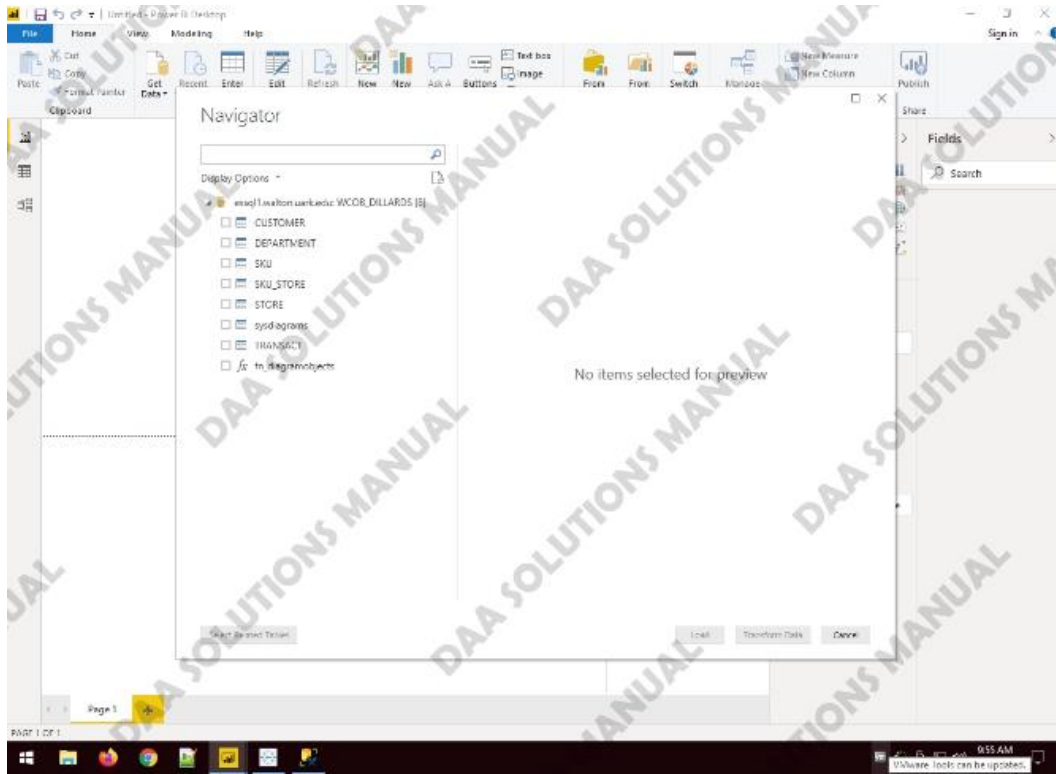
AQ4. Compare and Contrast: Compare the methods for viewing (and creating) relationships in Tableau versus Power BI. Which is easier to work with? Which provides more insight and flexibility?

*Answer: Have answers will depend on the student's opinion.*

## Lab 2-6 Submit Your Screenshot Lab Document

Verify that you have answered any questions your instructor has assigned, then upload your screenshot lab document to Connect or to the location indicated by your instructor.

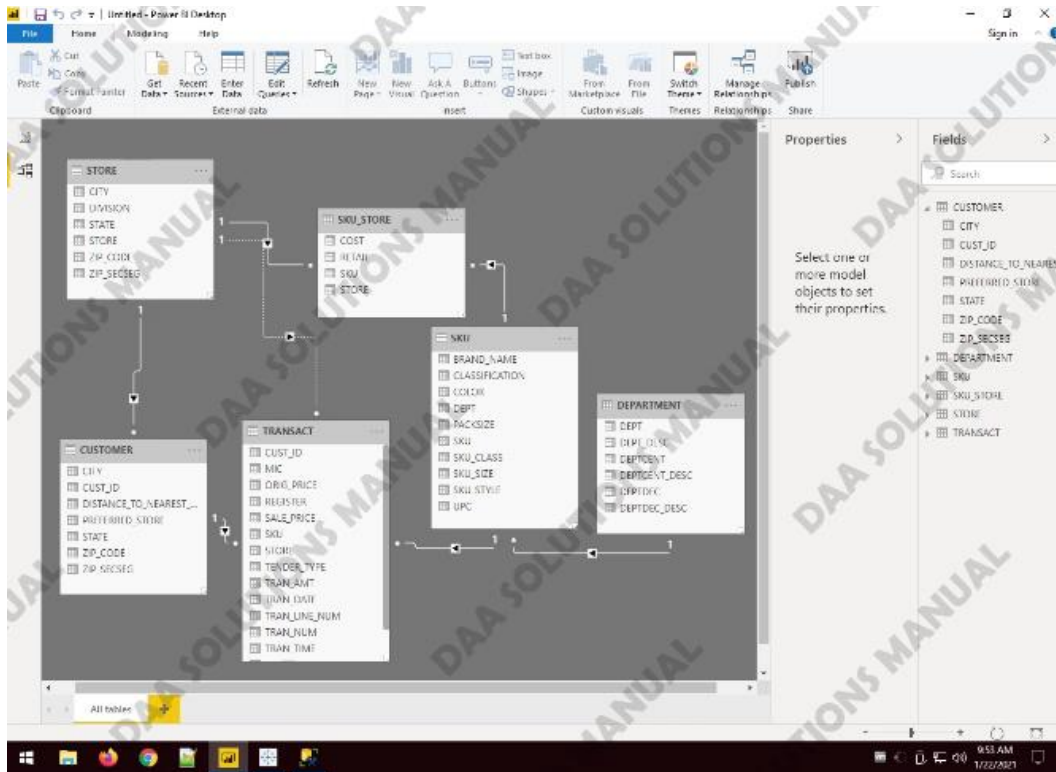
2-6MA





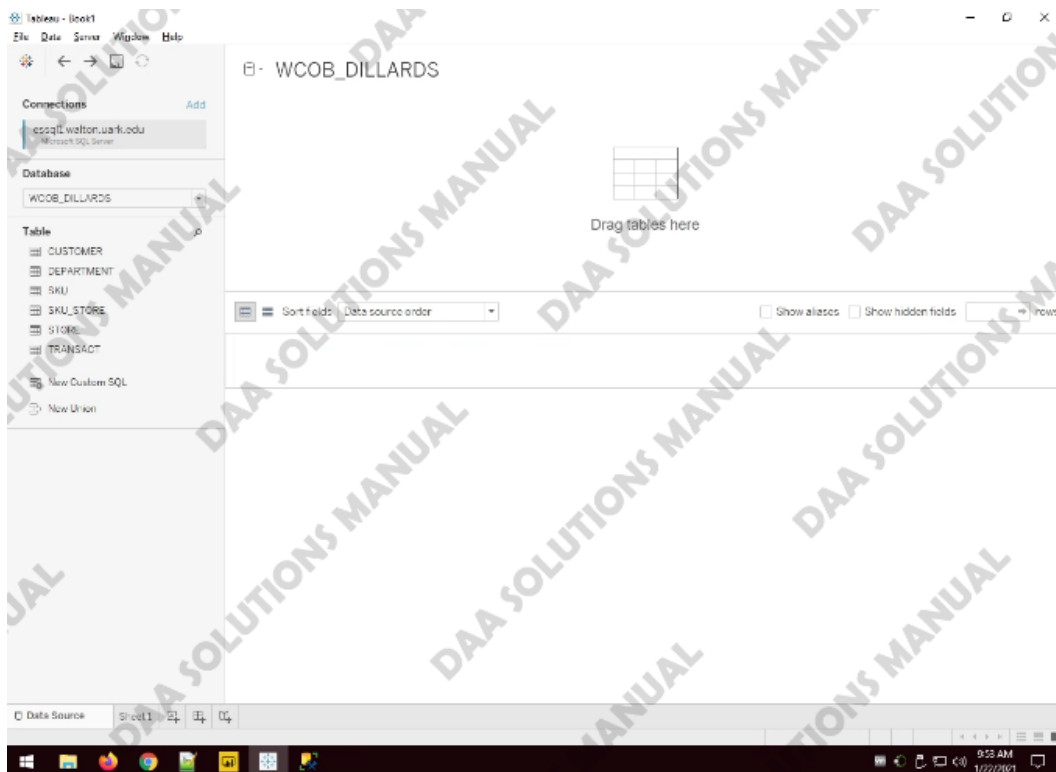
# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-6MB



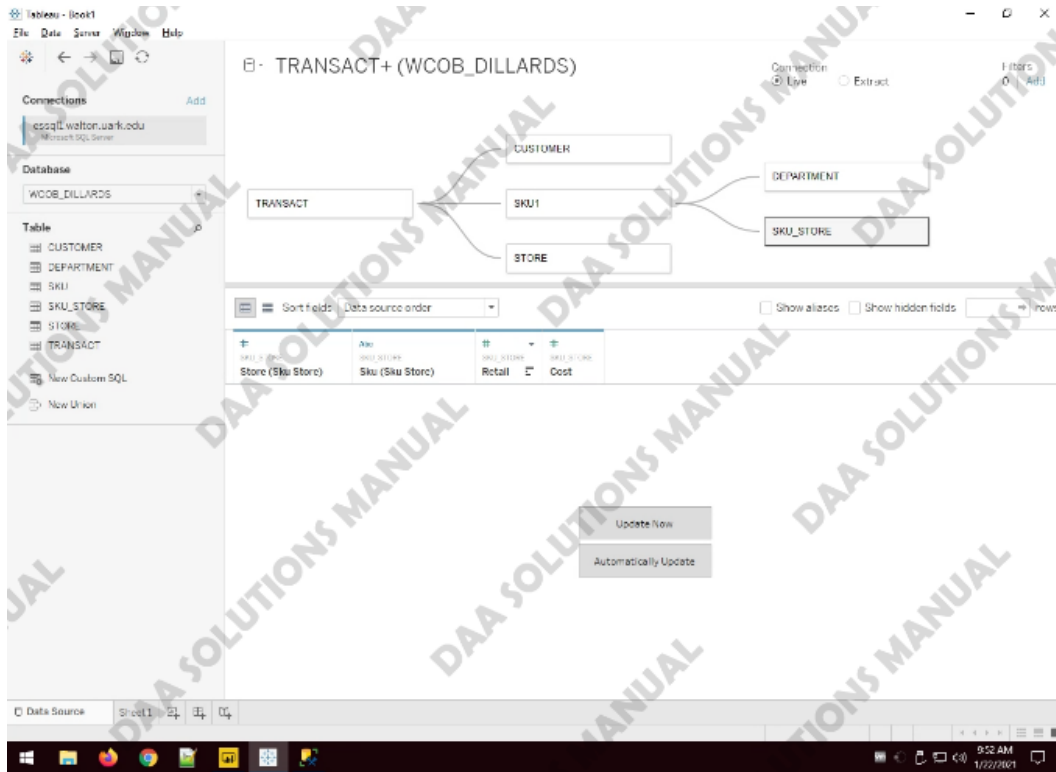
2-6TA

TBEXAM.COM



## Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-6TB



TBEXAM.COM



## Lab 2-7 Comprehensive Case: Preview Data from Tables—Dillard’s

### Lab 2-7 Part 1 Preview Dillard’s Attributes and Datatypes

#### Lab 2-7 Part 1 Objective Questions (LO 2-2, 2-3)

OQ1. What is the primary key for the CUSTOMER table?

- A) *CUST\_ID*
- B) *STATE*
- C) *PREFERRED\_STORE*
- D) *ZIP\_CODE*

Answer: A

OQ2. What is the primary key for the SKU table?

- A) *UPC*
- B) *SKU*
- C) *SKU\_CLASS*
- D) *DEPT*

Answer: B

OQ3. Which tables are related to the Customer table? (Hint: Do not forget the foreign keys that you discovered in the Transact table)

- A) *STORE and SKU*
- B) *SKU and CUSTOMER*
- C) *CUSTOMER and SKU*
- D) *TRANSACT and STORE*

Answer: D

### Lab 2-7 Part 2 Explore Dillard’s Data More In-Depth

#### Lab 2-7 Part 2 Objective Questions (LO 2-2, 2-3)

OQ1. What do you notice about the TRAN\_AMT for transactions with TRAN\_TYPE “P”?

- A) *Most of the transaction amounts are negative.*
- B) *Most of the transaction amounts exceed \$1000.*
- C) *Most of the transaction amounts are less than \$1000.*
- D) *There are not very many “P” transactions.*

Answer: C

**Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e**

OQ2. What do you notice about the TRAN\_AMT for transactions with TRAN\_TYPE “R”?

- A) *Most of the transaction amounts are negative.*
- B) *Most of the transaction amounts exceed \$1000.*
- C) *Most of the transaction amounts are between \$0 and \$1000.*
- D) *There are more “R” transactions than others.*

Answer: A

OQ3. What do “P” type transactions and “R” type transactions represent?

- A) *Refunds and promotions, respectively.*
- B) *Promotions and refunds, respectively.*
- C) *Returns and purchases, respectively.*
- D) *Purchases and returns, respectively.*

Answer: D

**Lab 2-7 Part 2 Analysis Questions (LO 2-2, 2-3)**

AQ1. Compare and Contrast: Compare the methods for previewing data in the tables in Tableau Prep versus Microsoft Power BI. Which method is easier to interact with?

Answer: Student responses will vary. [TBEXAM.COM](https://www.tbexam.com)

AQ2. Compare and Contrast: Compare the methods for identifying datatypes in each table in Tableau Prep versus Microsoft Power BI. Which method is easier to interact with?

Answer: Student responses will vary.

AQ3. Compare and Contrast: Compare viewing the data distribution and filtering data in Tableau Prep’s Clean step to Microsoft Power BI’s Data Profiling options. Which method is easier to interact with?

Answer: Student responses will vary.

**Lab 2-7 Submit Your Screenshot Lab Document**

Verify that you have answered any questions your instructor has assigned, then upload your screenshot lab document to Connect or to the location indicated by your instructor.

# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-7MA

Tableau Desktop interface showing the CUSTOMER table in the Navigator pane. The table contains the following data:

CUST_ID	CITY	STATE	ZIP_CODE	ZIP_SEQUENCE	DISTANCE
115003840	ADVANCE	NC	27006	2940	
115003841	CABO RAY	PR	00523	0078	
115003842	MOHAWK	OK	73059	5518	
115003843	ST THOMAS	VI	00804	1069	
115003844	BUCKSVILLE	NY	13038	2028	
115003845	SCOTTSDALE	AZ	85219	5245	
115003846	SANIT PETERSBURG	FL	33712	4855	
115003847	SANIT PETERSBURG	FL	33712	4256	
115003848	JONESBORO	AR	72404	8384	
115003849	TAYLOR	AZ	85219	1059	
115003850	HOUSTON	TX	77057	2144	
115003851	LAUREL	MA	02854	2010	
115003852	N MYRTLE BCH	SC	29582	3936	
115003853	SANIT	MA	01470	1474	
115003854	MARCO ISLAND	FL	34445	2125	
115003855	OKLAHOMA CITY	OK	73121		
115003857	OWING	MA	02777	3419	
115003858	SOMERSET	MA	02726	4333	
115003859	MUSKOGEE	AL	05255	3346	
115003862	MICHAEL	VA	22110	2021	
115003863	ROSELLE	NY	13357	4109	
115003864	BATAVIA	NY	14021	0021	
115003865	CHUMLEY PARK	NY	14127	8900	

2-7MB

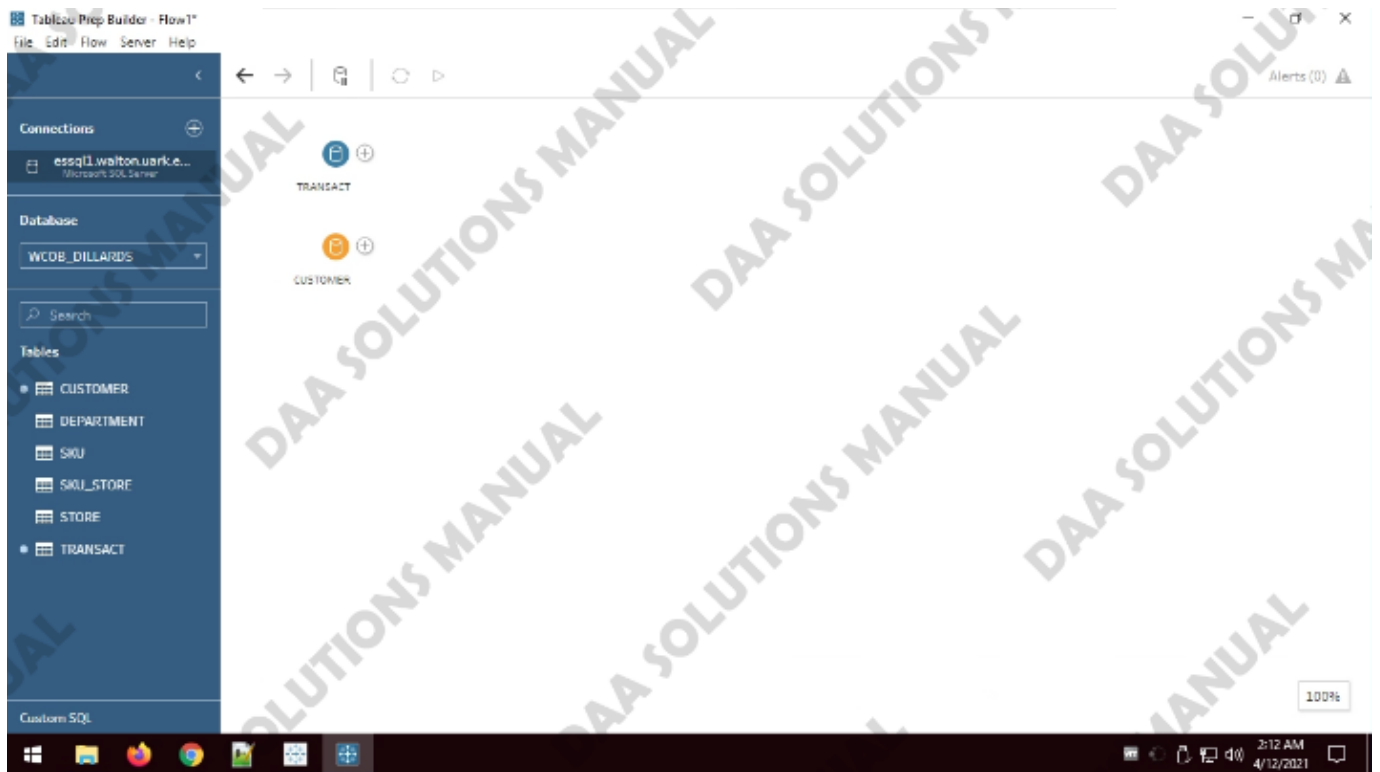
TBEXAM.COM

Tableau Desktop interface showing a query result for the TRANSACTIONS table. The table contains the following data:

ID	DATE	AMOUNT	TYPE
1	1/1/2010	120	DEBIT
2	1/1/2010	0	DEBIT
3	1/1/2010	0	DEBIT
4	1/1/2010	105	DEBIT
5	1/1/2010	0	DEBIT
6	1/1/2010	95	DEBIT
7	1/1/2010	128	DEBIT
8	1/1/2010	48	DEBIT
9	1/1/2010	235	DEBIT
10	1/1/2010	0	DEBIT
11	1/1/2010	80	DEBIT
12	1/1/2010	0	DEBIT
13	1/1/2010	27.95	DEBIT
14	1/1/2010	0	DEBIT
15	1/1/2010	35	DEBIT
16	1/1/2010	0	DEBIT
17	1/1/2010	213	DEBIT
18	1/1/2010	230	DEBIT
19	1/1/2010	0	DEBIT
20	1/1/2010	0	DEBIT
21	1/1/2010	270	DEBIT
22	1/1/2010	0	DEBIT

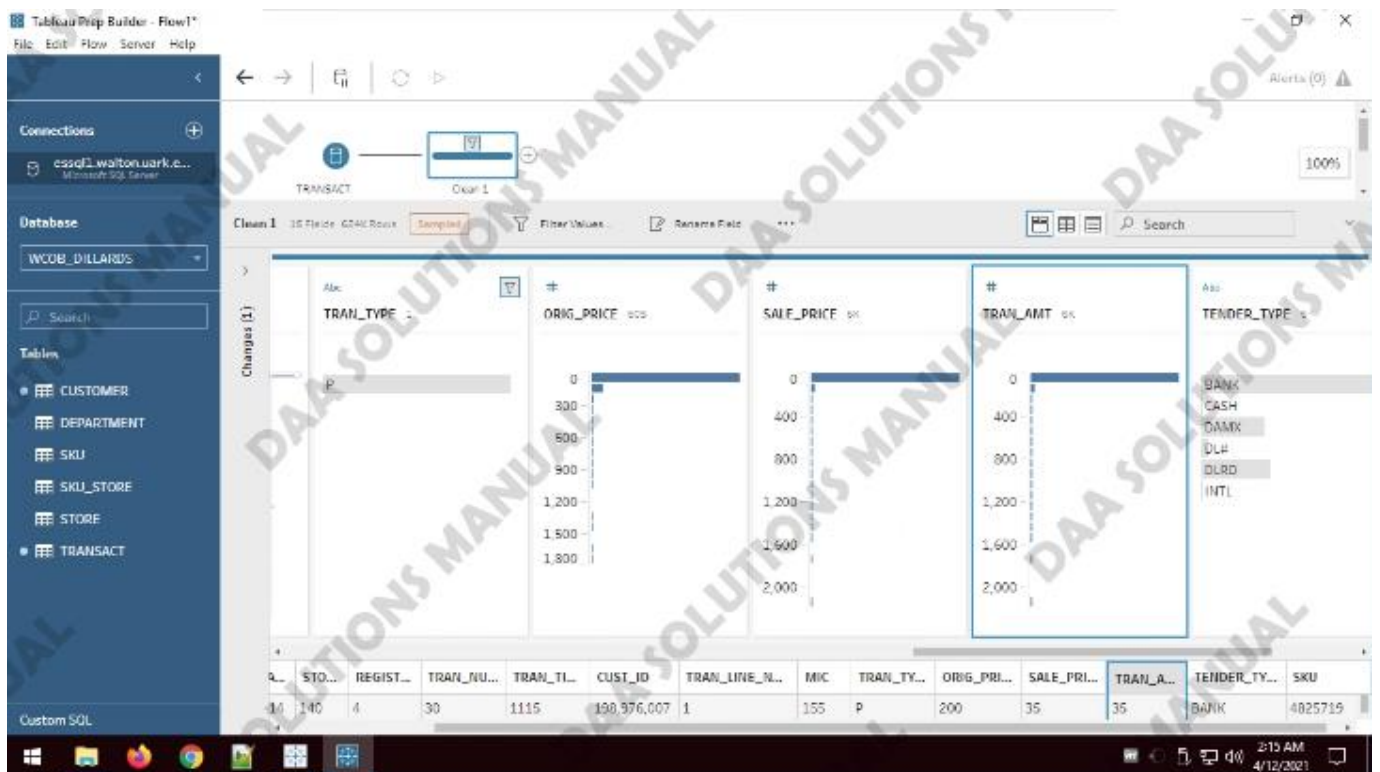
# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-7TA



2-7TB

TBEXAM.COM



## Lab 2-8 Comprehensive Case: Preview a Subset of Data in Excel, Tableau Using a SQL Query—Dillard’s

### Lab 2-8 Part 1 Connect to the Data with a SQL Query

No questions

### Lab 2-8 Part 2 View the Distribution of Transaction Amounts across States

#### Lab 2-8 Part 2 Objective Questions

OQ1. Which state has the highest average transaction amount?

- A) Texas
- B) Arkansas
- C) Florida
- D) Louisiana

Answer: B

OQ2. What is the average of transactions for North Carolina? Round your answer to the nearest dollar.

Answer: 16.57

#### Lab 2-8 Part 2 Analysis Questions

AQ1. How does creating a query to connect to the data allow quicker and more efficient access and analysis of the data than connecting to entire tables?

Answer: Queries are simple tools that are built for efficiently pulling in data from large databases. They can be used in tools like Tableau and Power BI to summarize data without all of the computational overhead needed.

AQ2. Is 5 days of data sufficient to capture the statistical relationship among and between different variables? What will Excel do if you have more than 1 million rows? How might a query help?

Answer: Answers will vary, but five days is likely insufficient to capture a statistical relationship. Excel does not allow more than a million rows so is not always the most appropriate tool. Queries can help poll specific pieces of data without the file size limitation in Excel.

If you have completed BOTH tracks,

AQ3. Compare and Contrast: Compare the methods for analyzing transactions across states in Excel versus Tableau. Which tool was more intuitive for you to work with? Which provides more interesting results?

Answer: Student responses will vary.

### Lab 2-8 Submit Your Screenshot Lab Document

Verify that you have answered any questions your instructor has assigned, then upload your screenshot lab document to Connect or to the location indicated by your instructor.



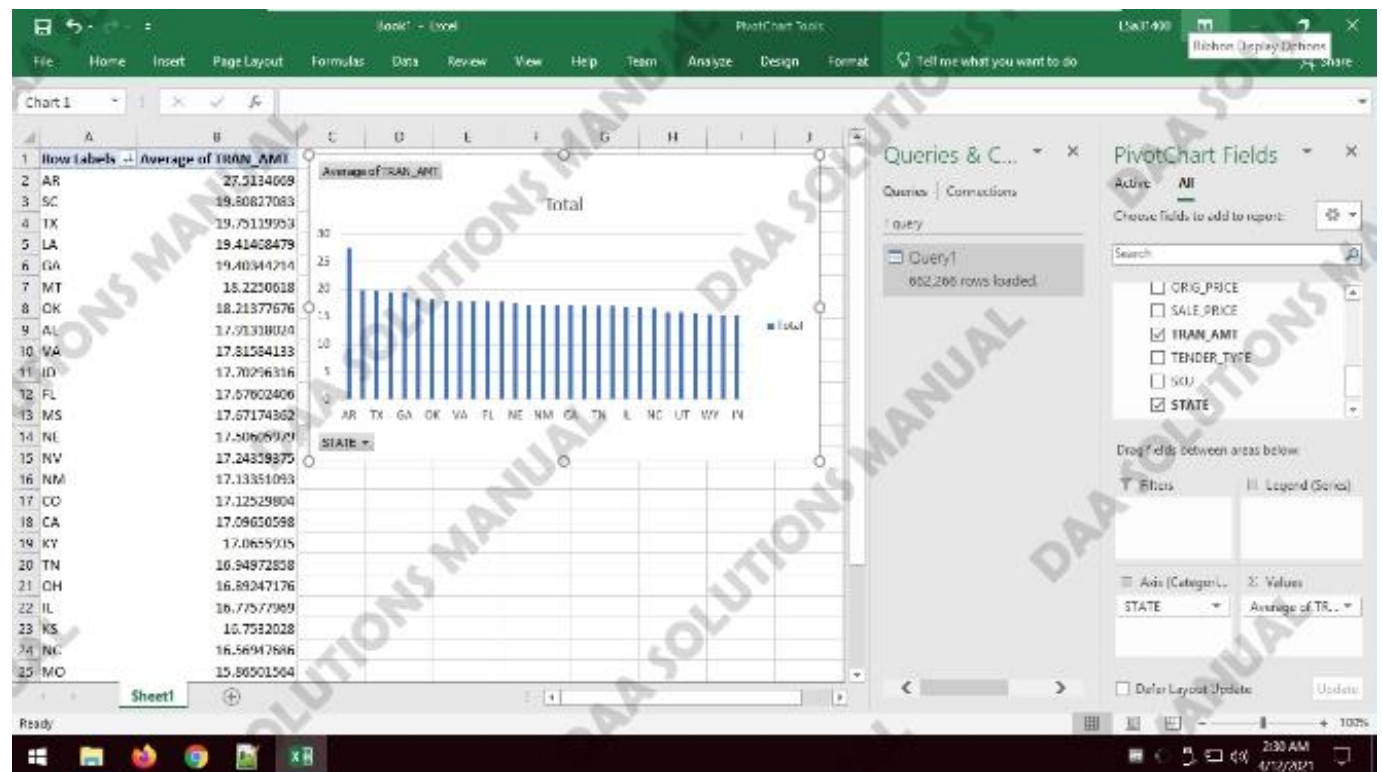
# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-8MA

TRANSACTION_ID	TRAN_DATE	L2_STORE	L2_REGISTER	TRAN_NUM	TRAN_TIME	CUST_ID	TRAN_LINE_NUM
101082188	5/1/2015 12:00:00 AM	321	6	14 1413	209221852		
101082405	5/2/2015 12:00:00 AM	102	103	1047 0425	152176227		
101082507	5/2/2015 12:00:00 AM	321	45	44 1526	280000625		
101082175	5/1/2015 12:00:00 AM	258	64	5 1152	209707256		
101082555	5/3/2015 12:00:00 AM	427	34	18 1405	217115962		
101082638	5/3/2015 12:00:00 AM	411	21	21 1523	268722535		
101082657	5/2/2015 12:00:00 AM	282	42	27 1240	215741002		
101082075	5/2/2015 12:00:00 AM	427	320	3353 0618	242178454		
101082414	5/1/2015 12:00:00 AM	329	30	11 1153	219114017		
101082507	5/2/2015 12:00:00 AM	427	320	3353 0618	242178454		
101082561	5/2/2015 12:00:00 AM	258	26	18 1318	209401947		
101091038	5/2/2015 12:00:00 AM	351	121	2384 0054	135218700		
101091044	5/1/2015 12:00:00 AM	218	6	5 1154	218725799		
101091577	5/2/2015 12:00:00 AM	258	322	5402 0450	224501014		
101091805	5/2/2015 12:00:00 AM	318	322	3583 0652	204900547		
101091831	5/1/2015 12:00:00 AM	213	59	6 1236	200982817		
101091970	5/2/2015 12:00:00 AM	427	54	62 1930	225011052		
101092174	5/5/2015 12:00:00 AM	218	53	5 1122	216900303		
101092412	5/2/2015 12:00:00 AM	427	54	67 1400	210011431		
101092813	5/2/2015 12:00:00 AM	321	22	24 1328	219546088		
101092702	5/1/2015 12:00:00 AM	229	26	17 1858	247571200		

2-8MB

TBEXAM.COM



# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-8MC

2-8MD

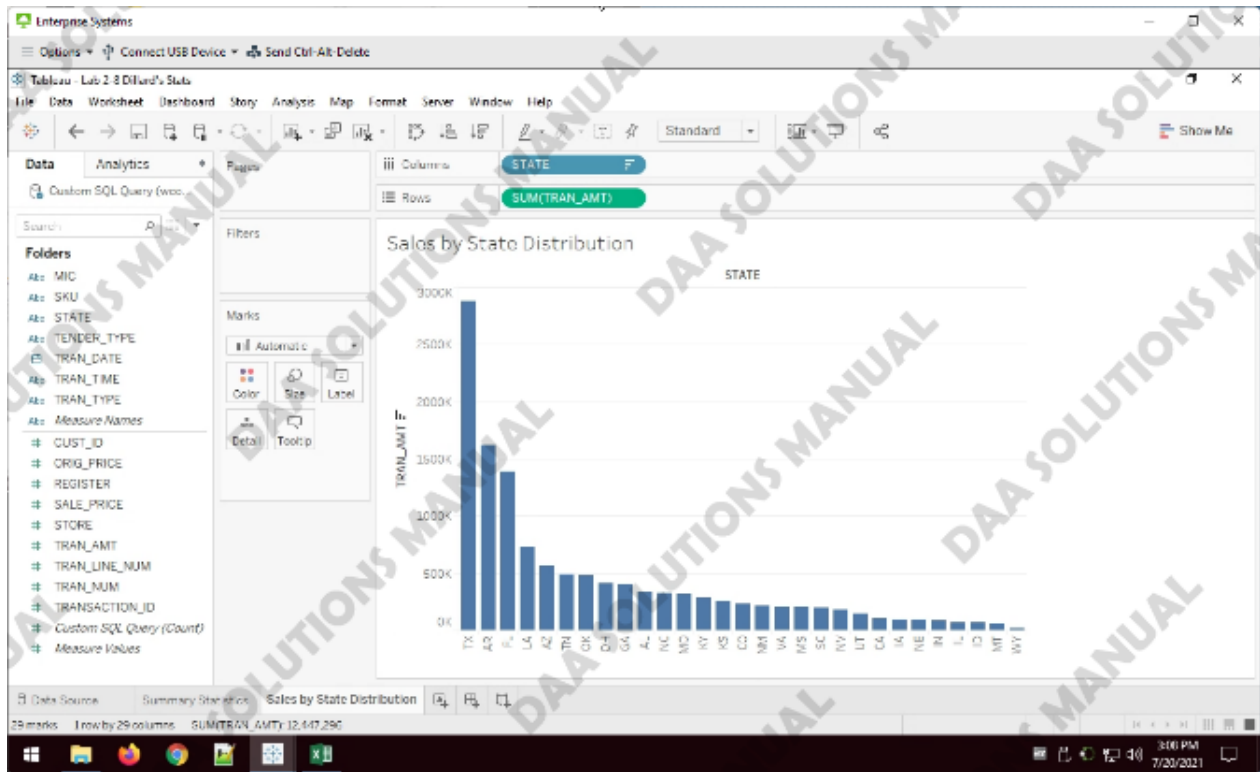
2-8TA

The screenshot shows the Tableau Desktop interface. The main view is a 'Custom SQL Query (wcob\_dillards)' with a table of transaction data. The table has columns: TRANSACTION\_ID, TRAN\_DATE, STORE, REGISTER, TRAN\_NUM, TRAN\_TIME, CUST\_ID, and TRAN\_LINE\_NUM. The data is sorted by TRAN\_NUM in descending order. The interface also shows a 'Connections' pane on the left with a list of tables: CUSTOMER, DEPARTMENT, SKU, SKU\_STORE, STORE, TRANSCRIPT, New Custom SQL, and New Union. The bottom status bar shows 'Data Source: Summary Statistics: Filter by State Distribution'.

TRANSACTION_ID	TRAN_DATE	STORE	REGISTER	TRAN_NUM	TRAN_TIME	CUST_ID	TRAN_LINE_NUM
99,994,963	9/1/2016	292	13	4	1102	132,233,775	4
99,995,556	9/1/2016	435	6	1295	1295	215,499,127	1
99,995,744	9/1/2016	788	11	12	1300	170,847,261	1
99,995,793	9/1/2016	161	8	9	1226	210,667,139	1
99,996,176	9/1/2016	767	7	19	1540	120,006,130	1
99,996,225	9/1/2016	161	21	14	1759	185,755,062	1
99,996,262	9/1/2016	924	54	41	1927	118,127,133	1

2-8TB

# Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e



2-8TC

TBEXAM.COM