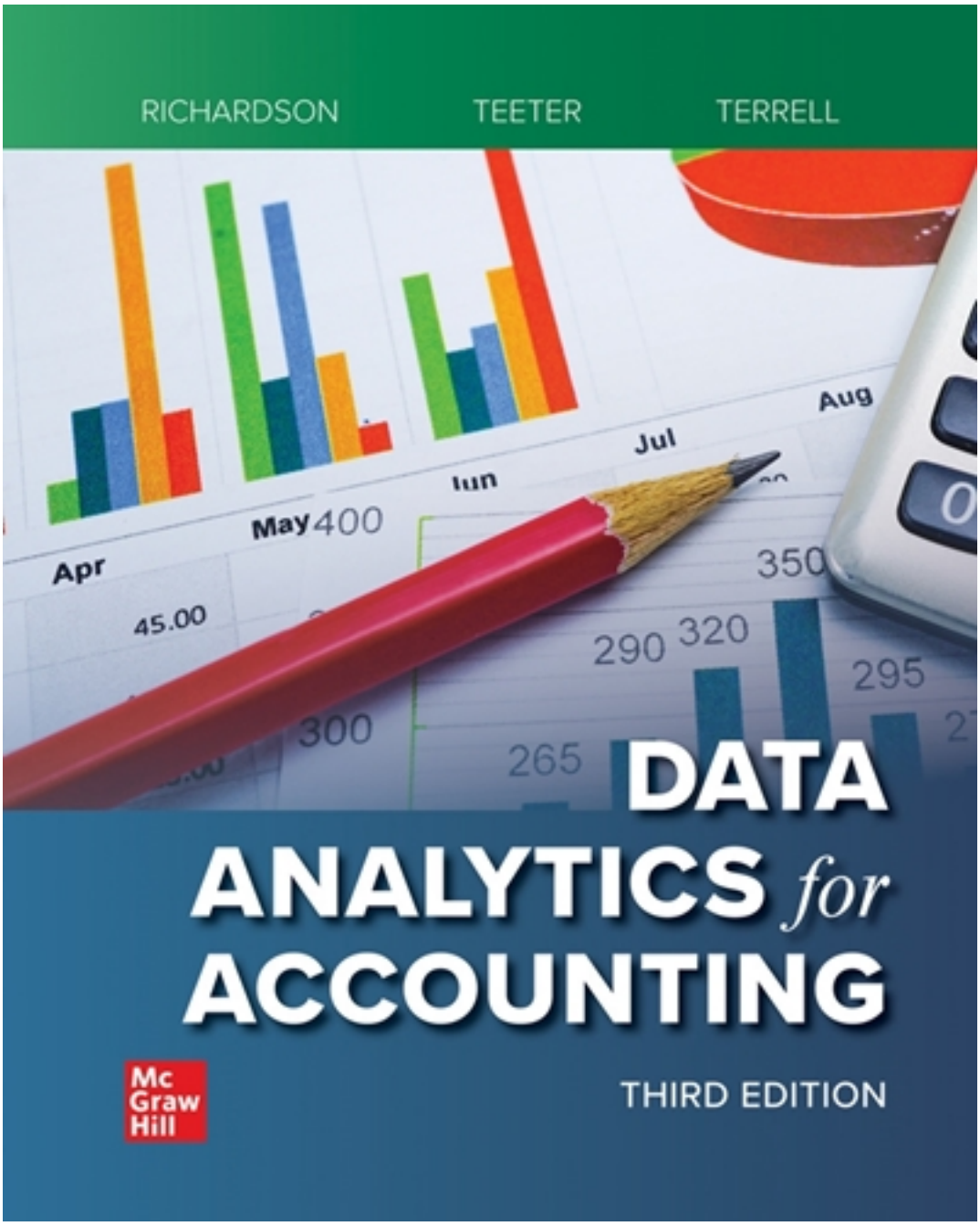
Solutions for Data Analytics for Accounting 3rd Edition by Richardson

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

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. Nondescript 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?

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

Answer: C

Solutions to Discussion and Analysis Questions

1. 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.

2. 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).

3. 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.

4. 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.

    a. Transforming NULL and N/A values into blanks.

        i. The COUNT and AVERAGE functions would not include these fields in their computation for these variables.

    b. Transforming NULL and N/A values into zeroes.

        i. 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.

    c. Deleting records that have NULL and N/A values from your dataset.

        i. 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

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?

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.

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*

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*

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

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.

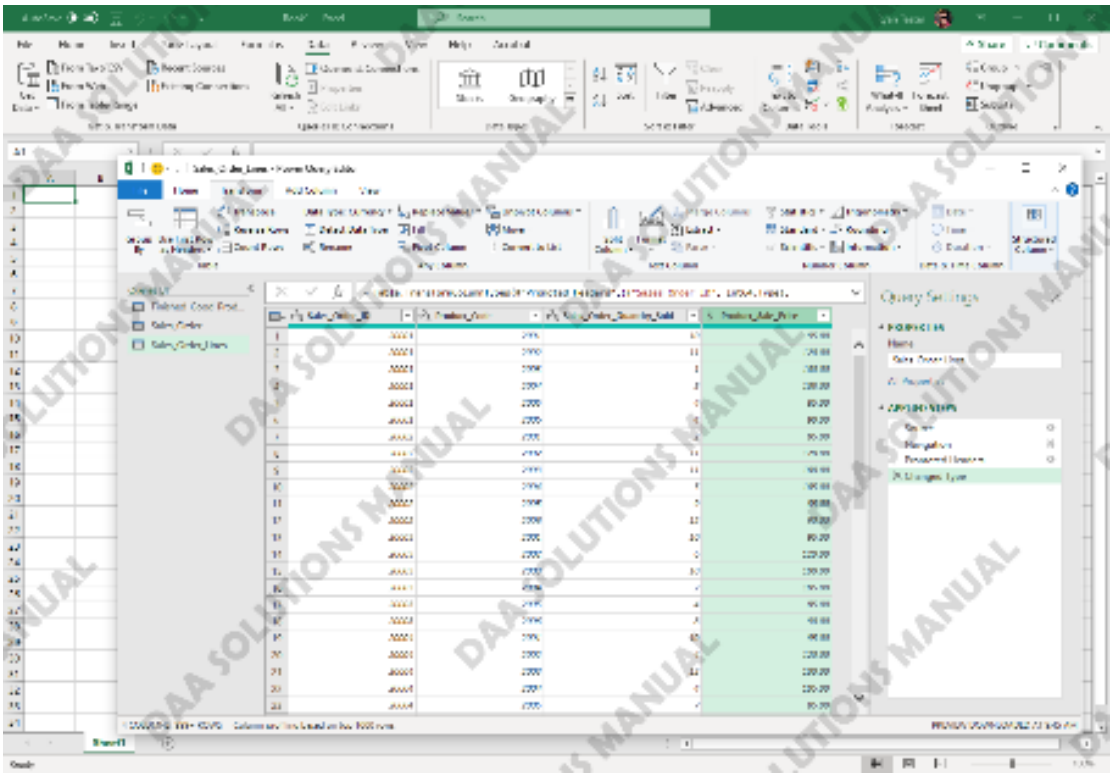
AQ4. In Chapter 4, we'll discuss some visualization techniques. Describe a way you could present these data as a chart or graph.

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

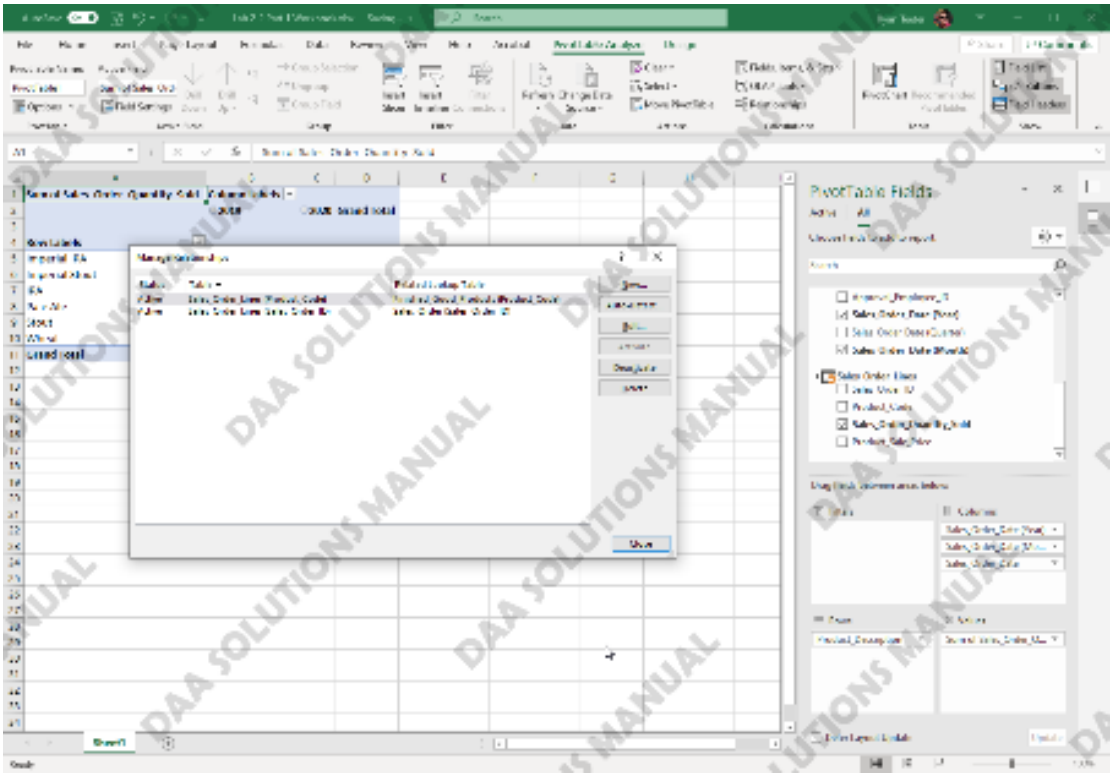
Lab 2-2 Submit Your Screenshot Lab Document

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.

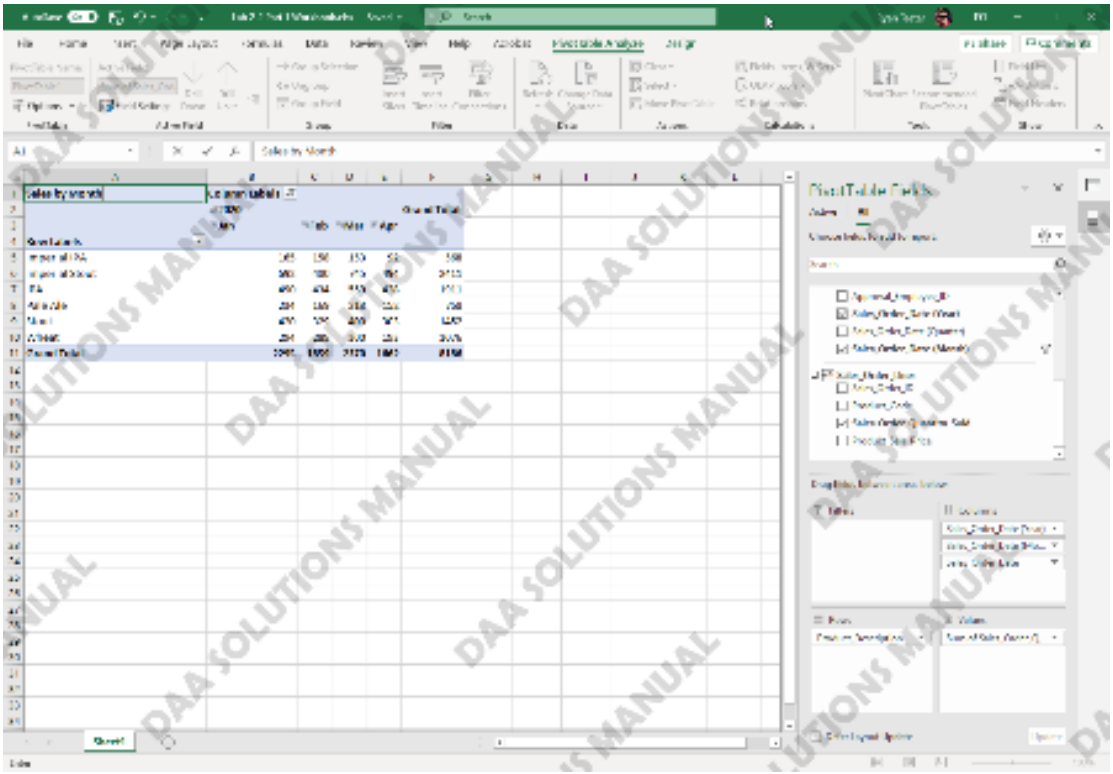
2-2MA



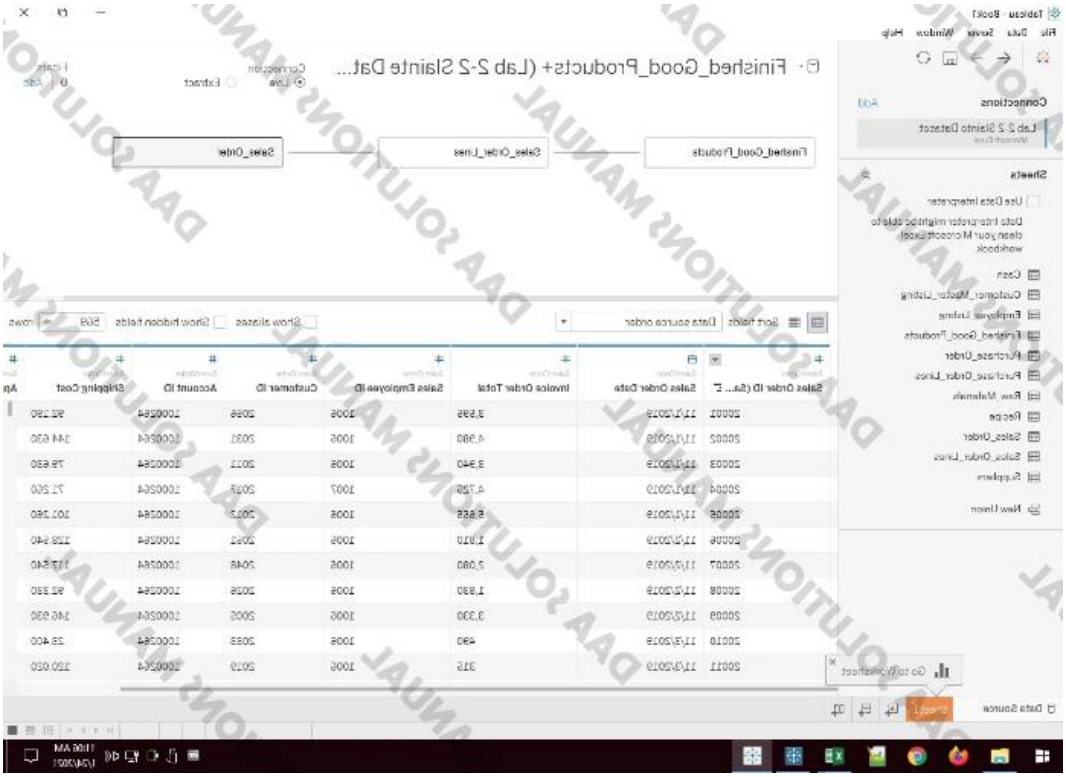
2-2MB



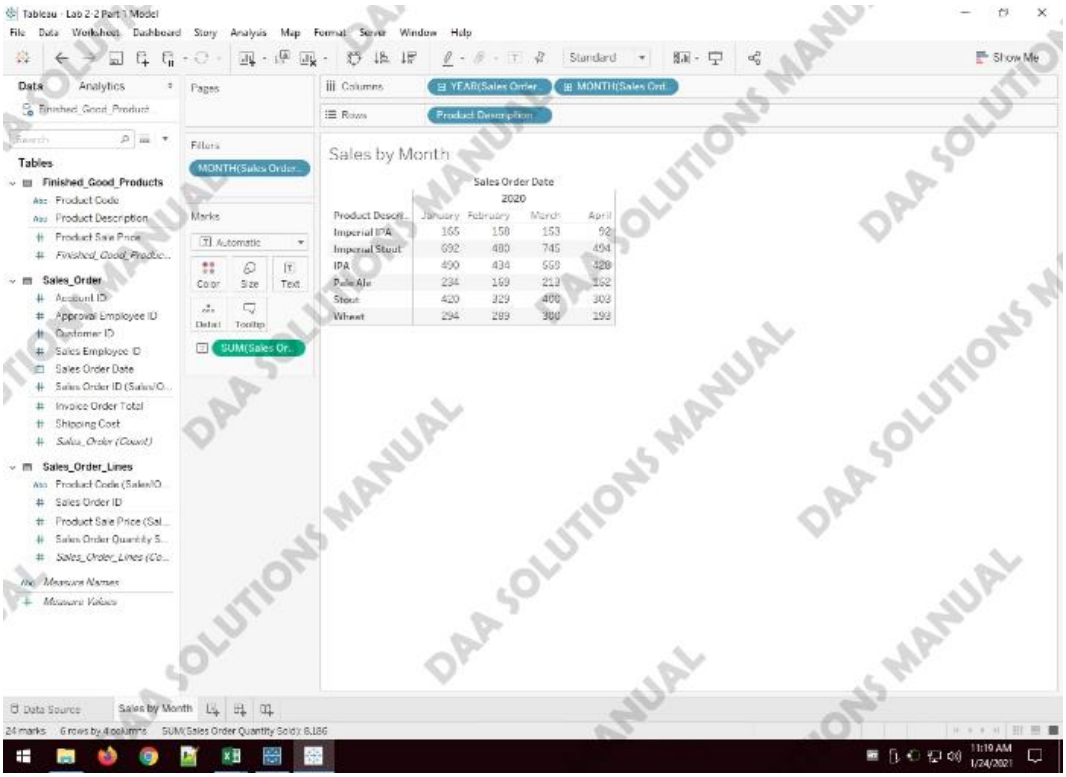
2-2MC



2-2TA



2-2TB



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?

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.

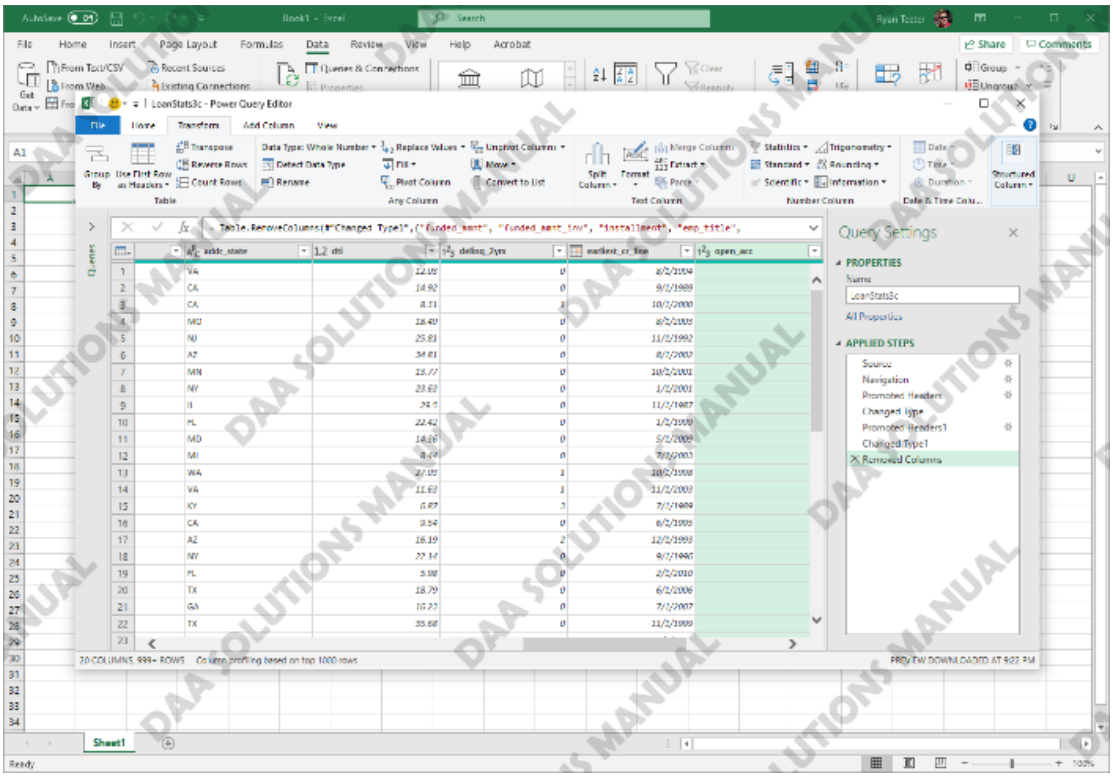
AQ4. What are some steps you could take to clean the data and resolve the difficulties you identified?

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

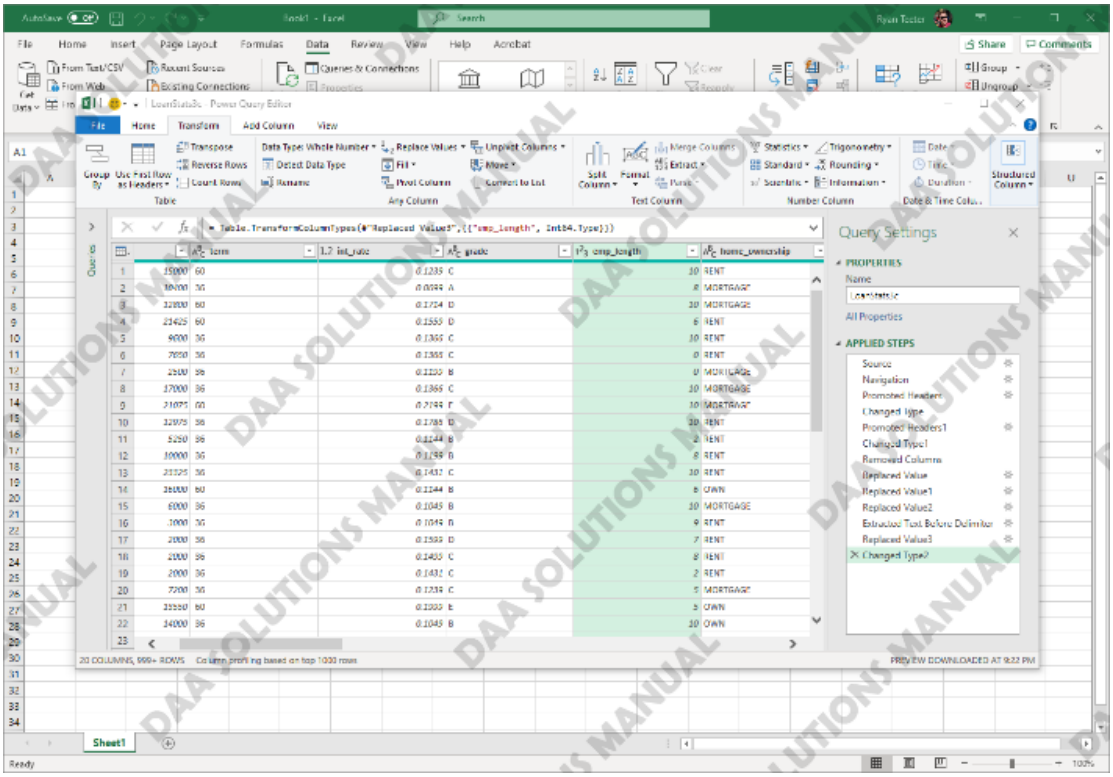
Lab 2-3 Submit Your Screenshot Lab Document

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.

2-3MA

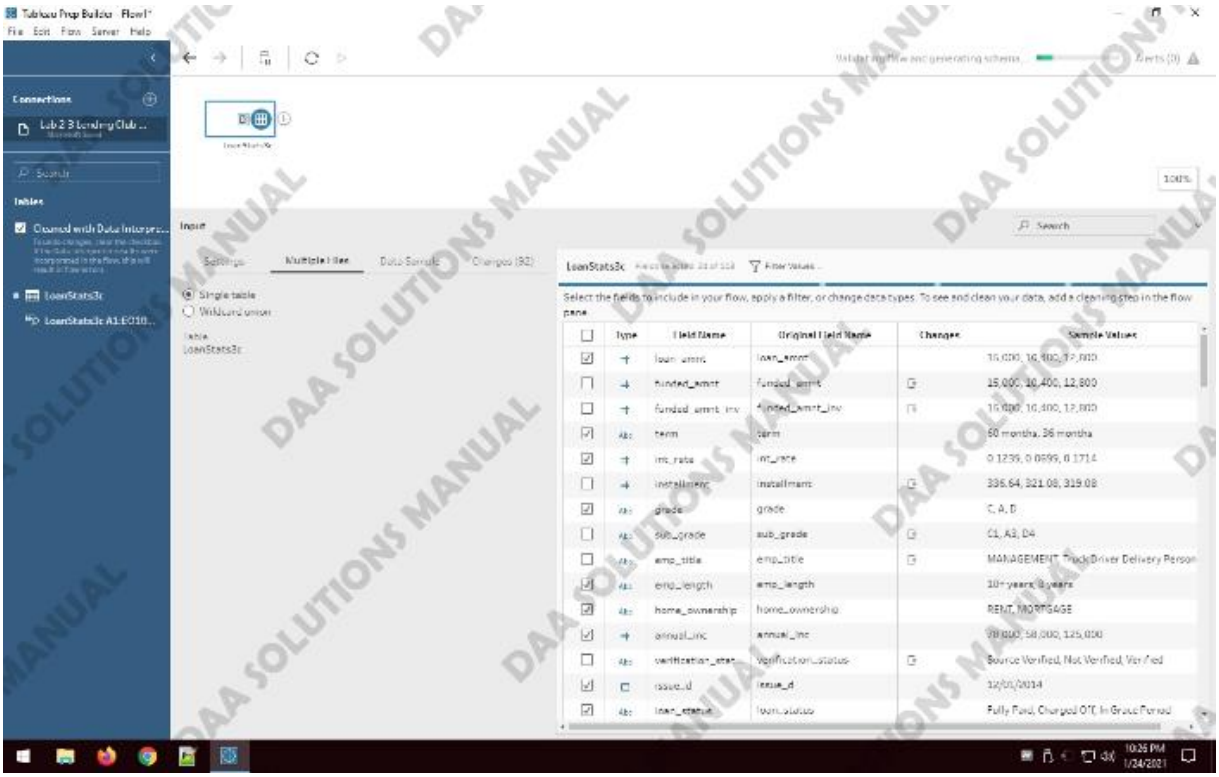


2-3MB

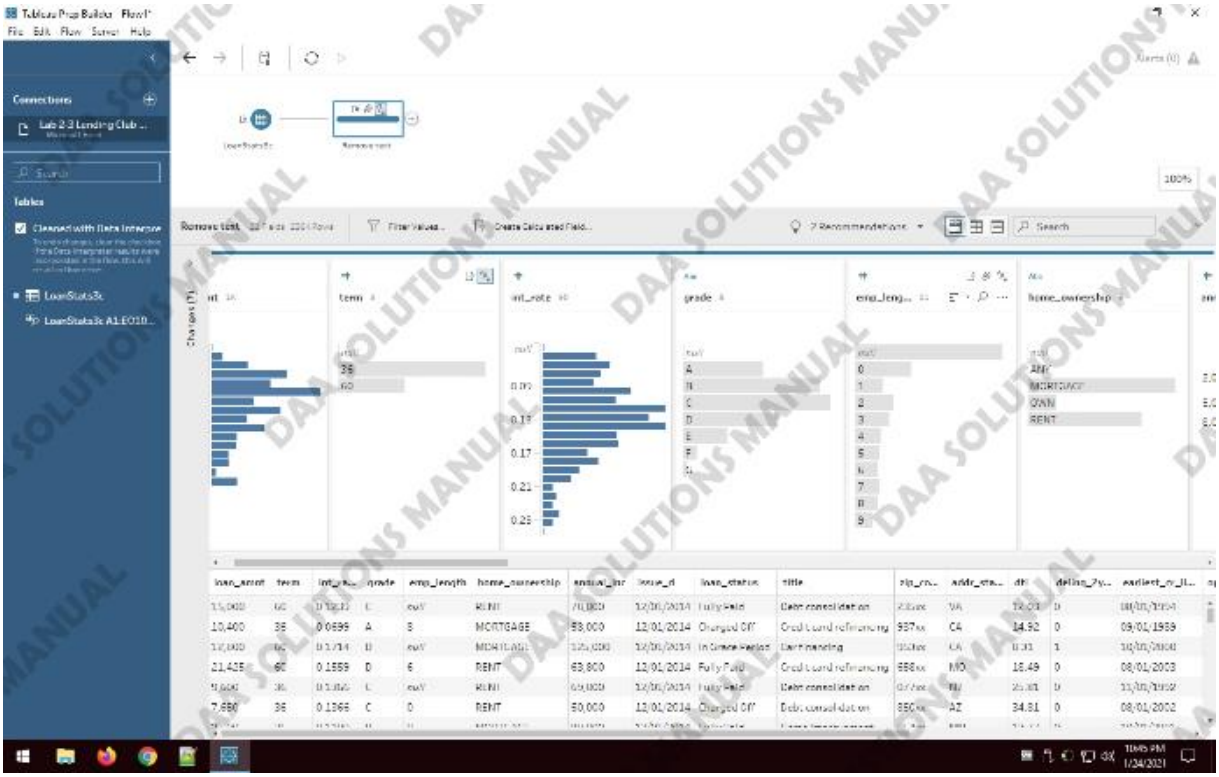


Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-3TA



2-3TB



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.

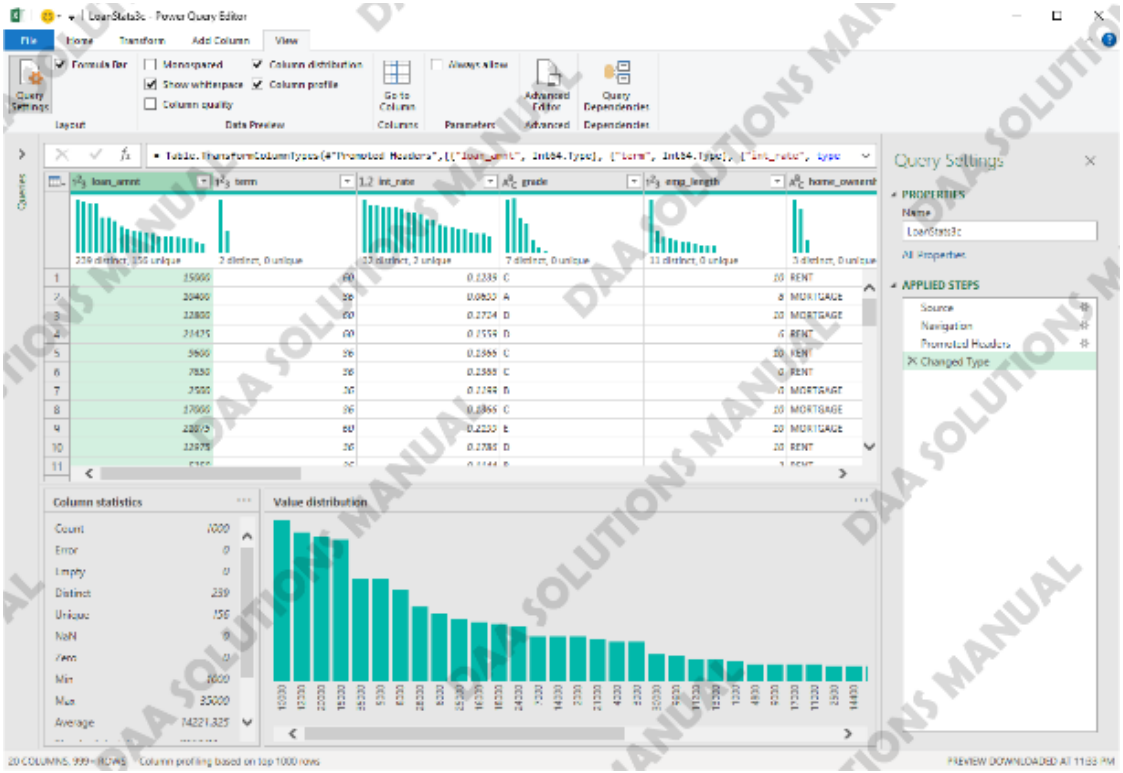
AQ4. Compare and contrast: What are some of the summary statistics measures that are unique to Power Query? To Tableau Desktop?

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

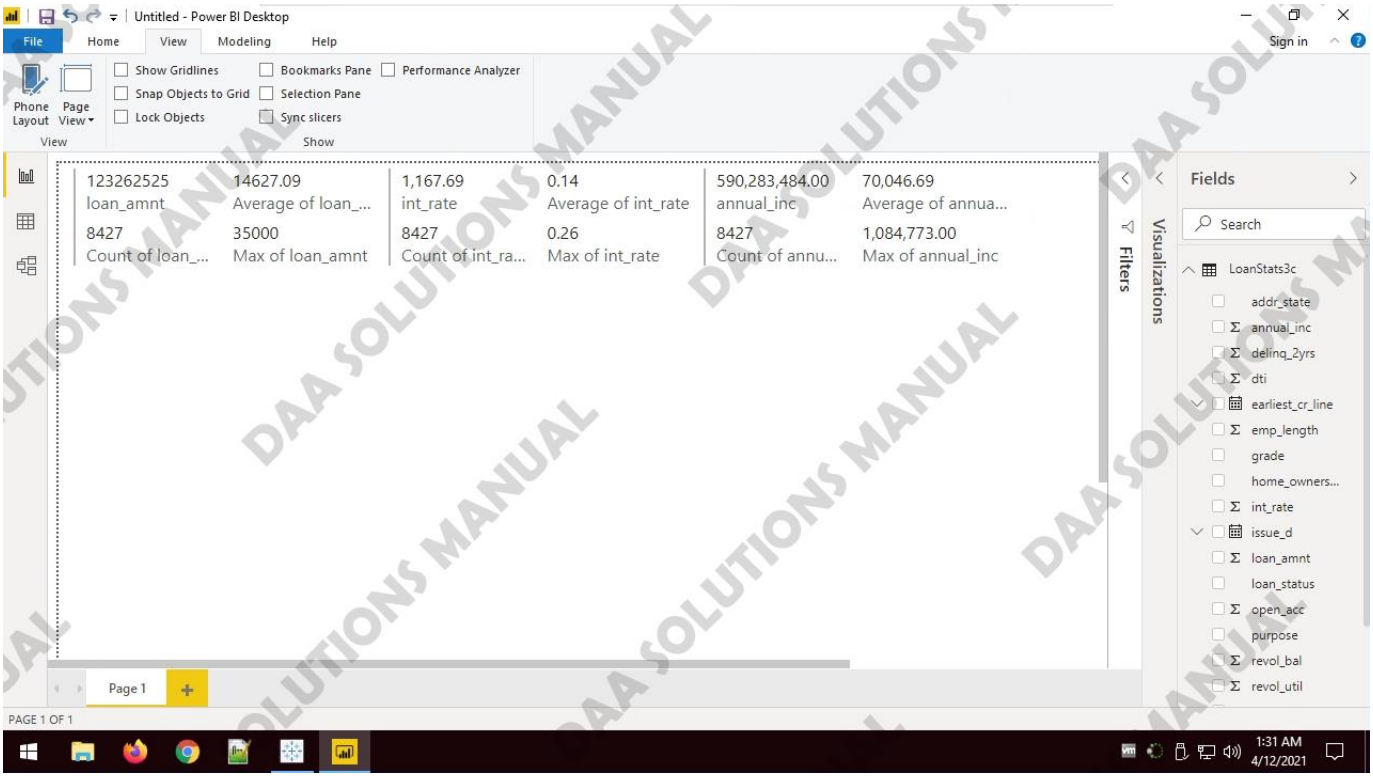
Lab 2-4 Submit Your Screenshot Lab Document

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.

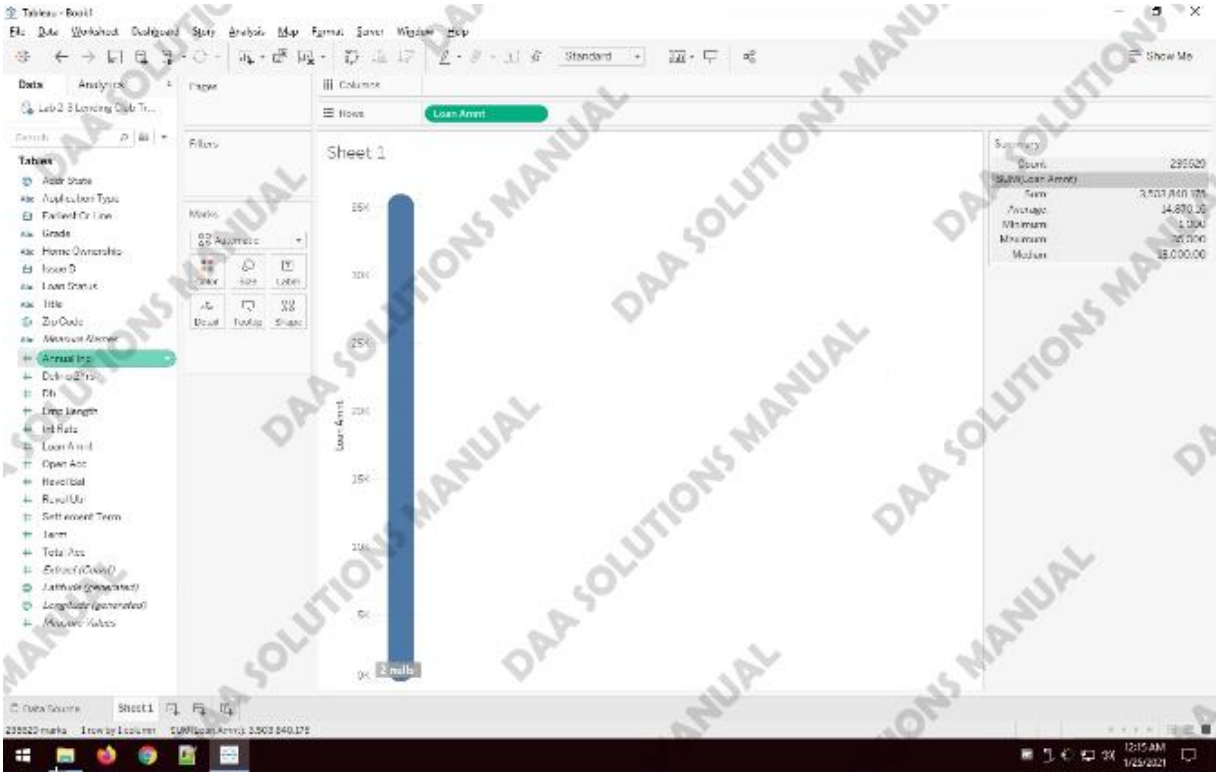
2-4MA



2-4MB



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.

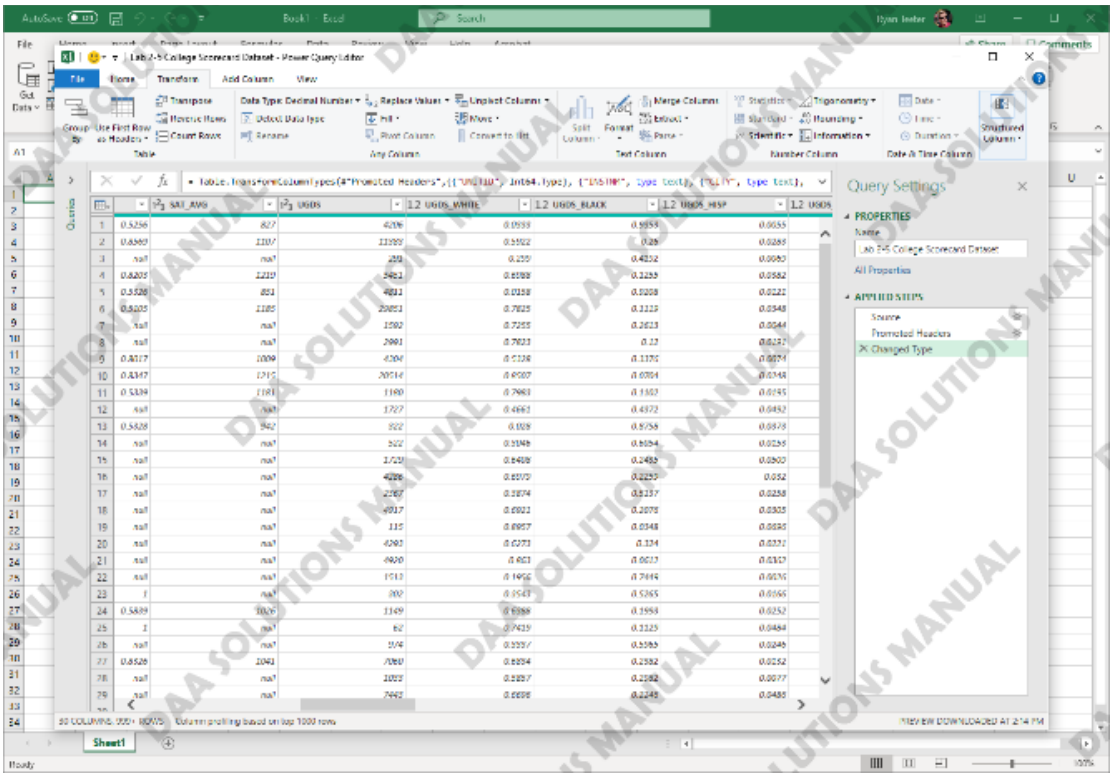
AQ3. What would happen if we excluded all schools that don't report an average SAT score?

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.

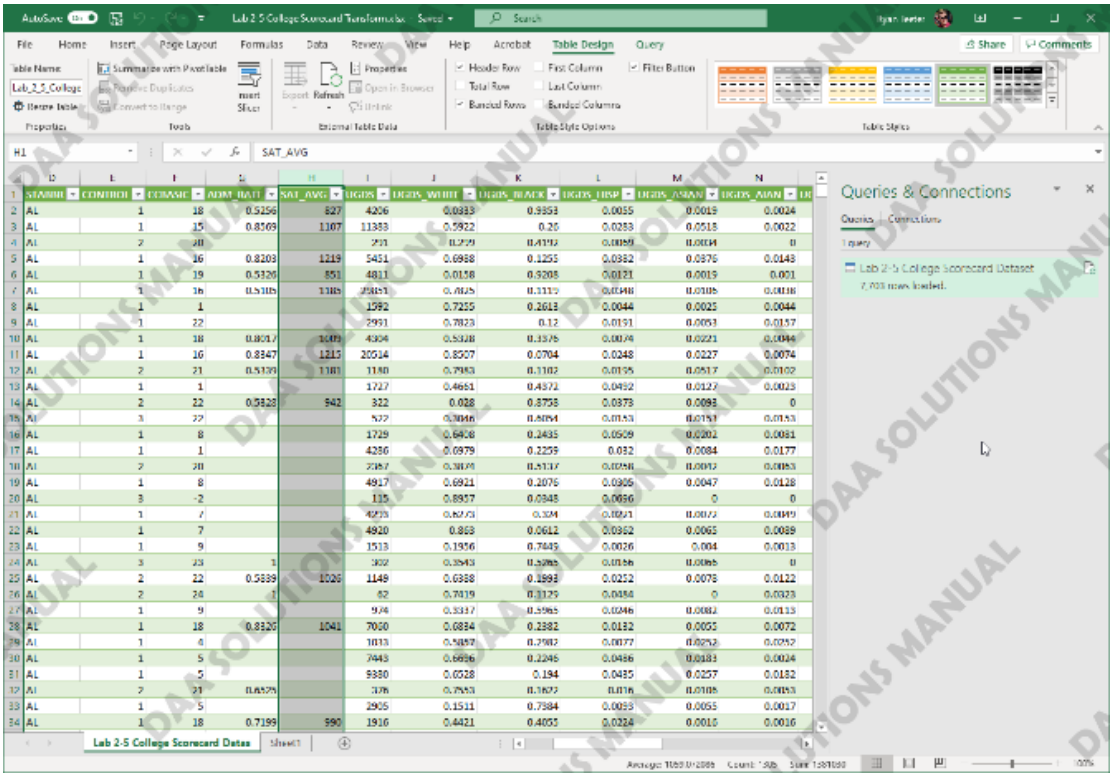
Lab 2-5 Submit Your Screenshot Lab Document

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.

2-5MA

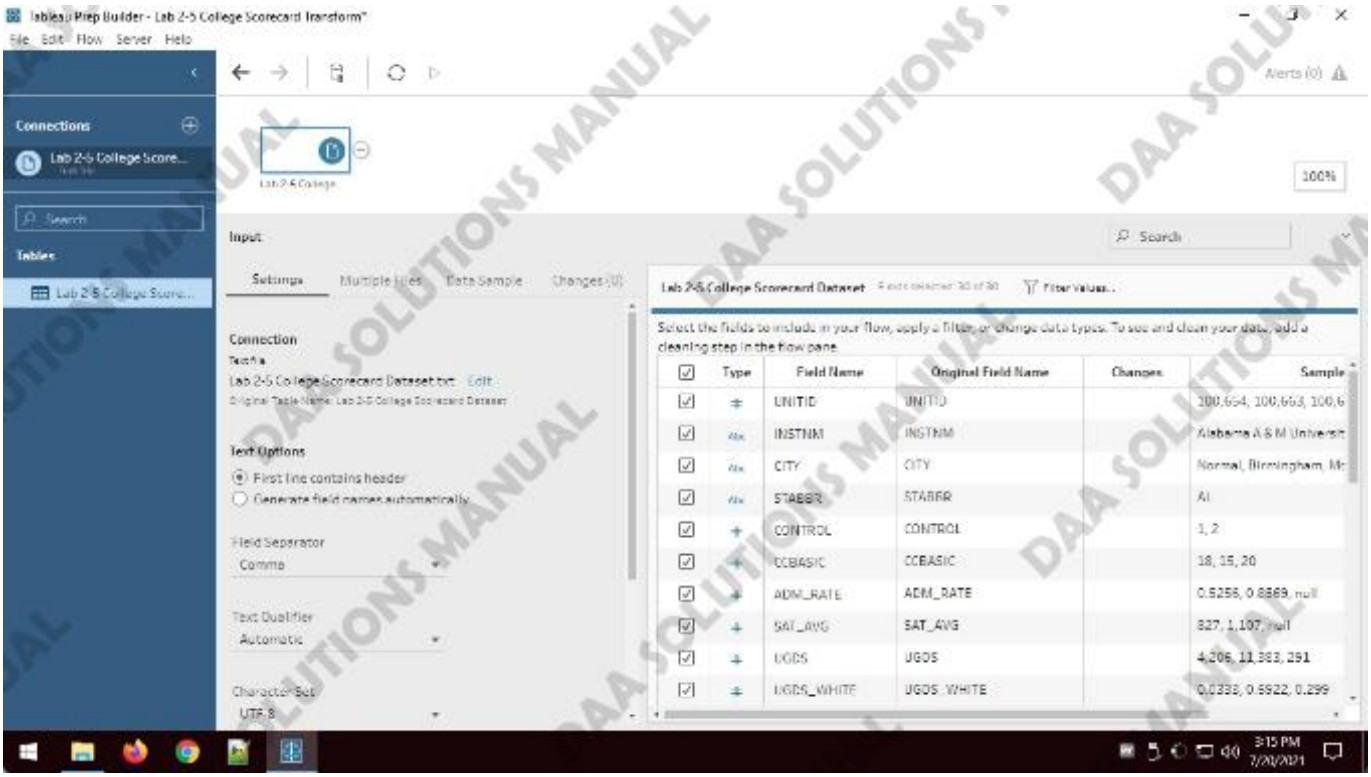


2-5MB

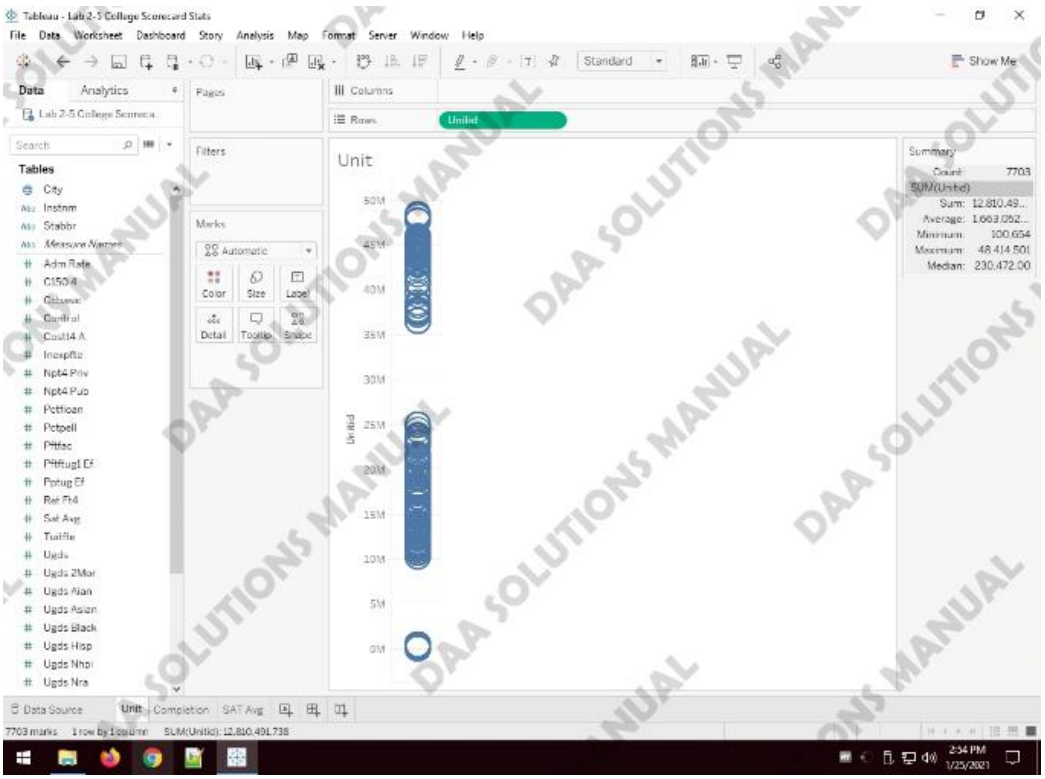


Richardson, Teeter, Terrell – Data Analytics for Accounting, 3e

2-5TA



2-5TB



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.

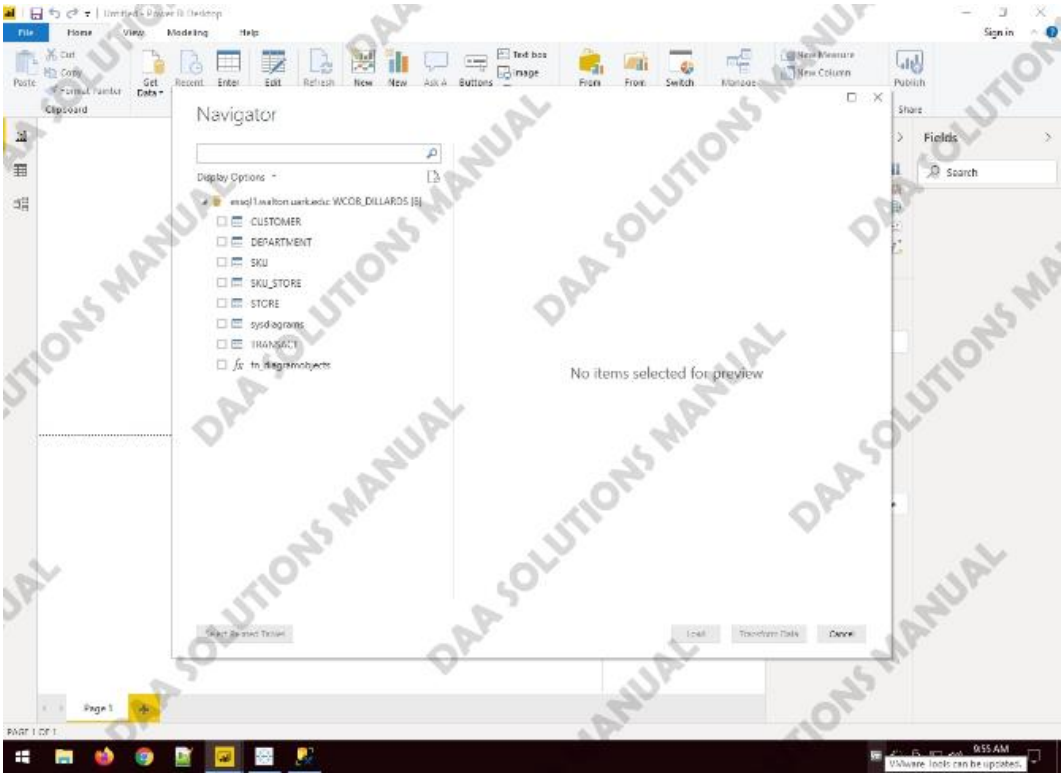
AQ4. Compare and Contrast: Compare the methods for viewing (and creating) relationships in Tableau versus Power BI. Which is easier to work with? Which provides more insight and flexibility?

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

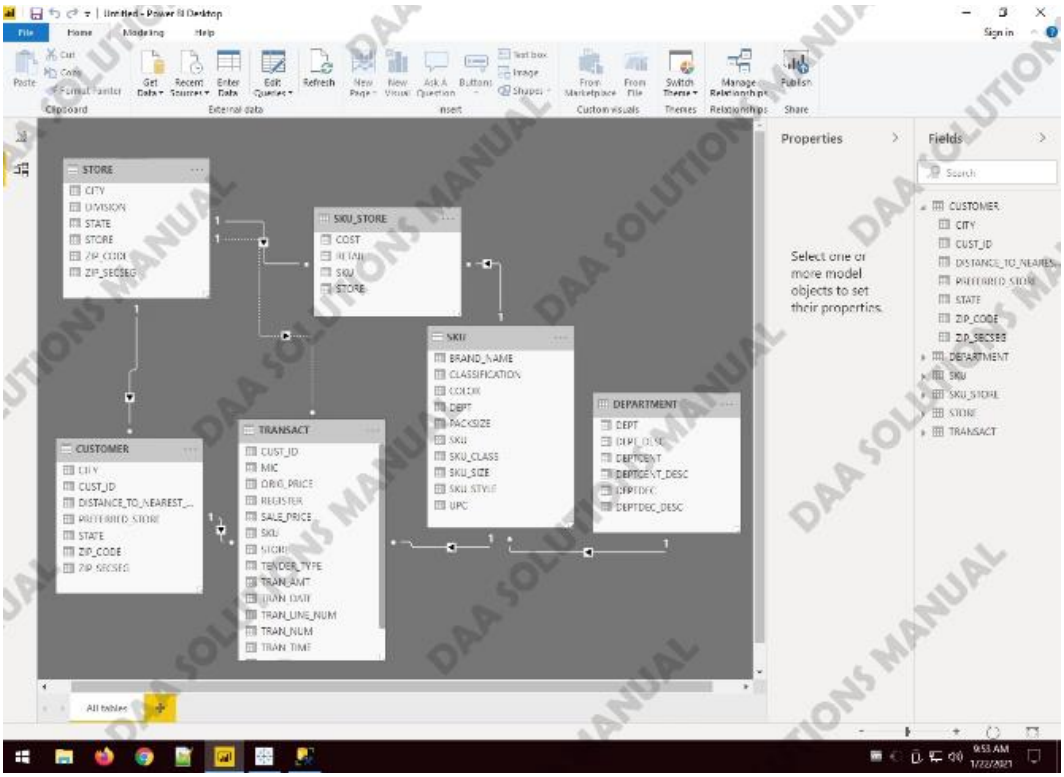
Lab 2-6 Submit Your Screenshot Lab Document

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.

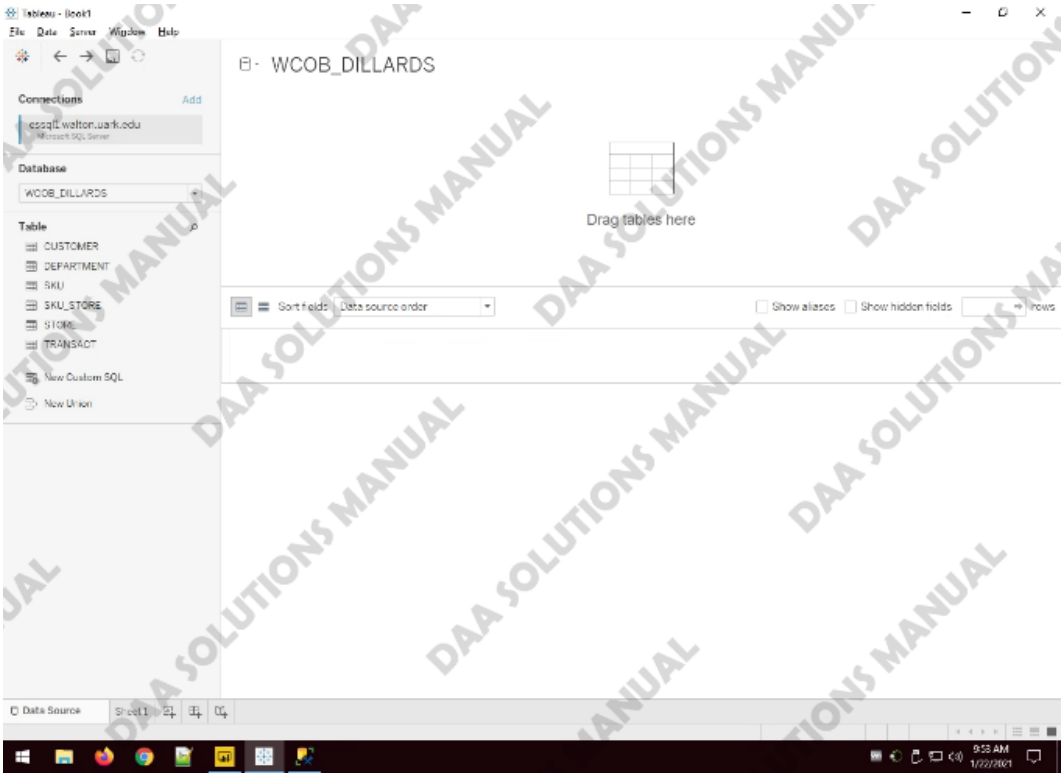
2-6MA



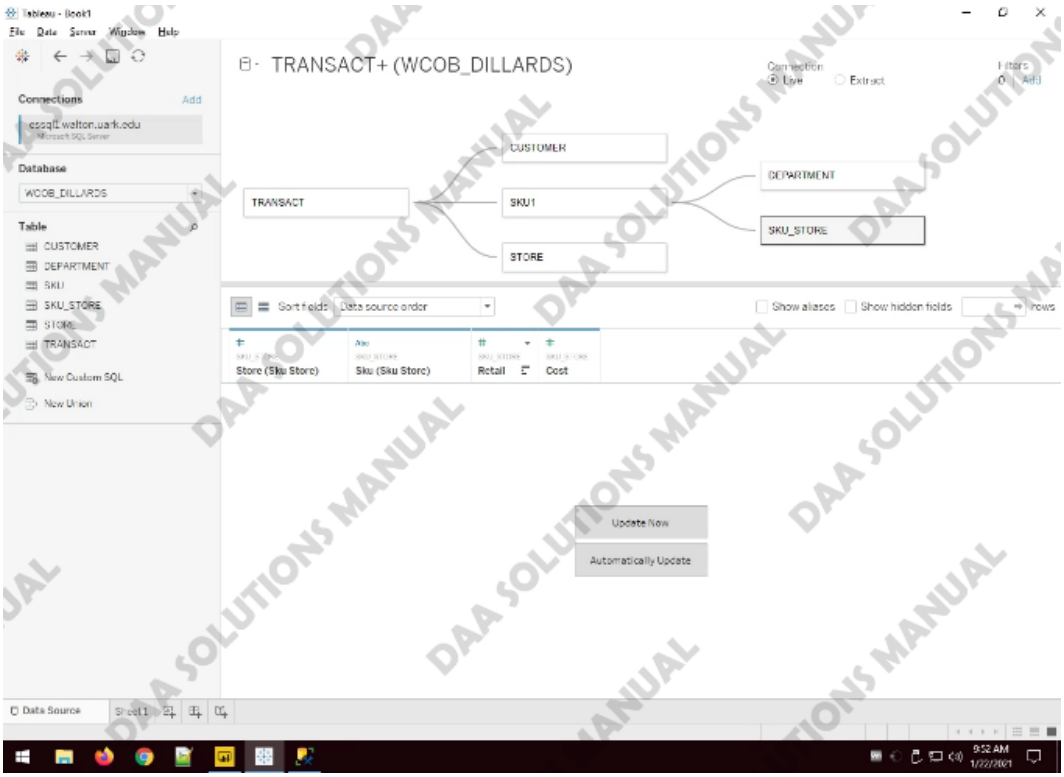
2-6MB



2-6TA



2-6TB



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

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.

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.

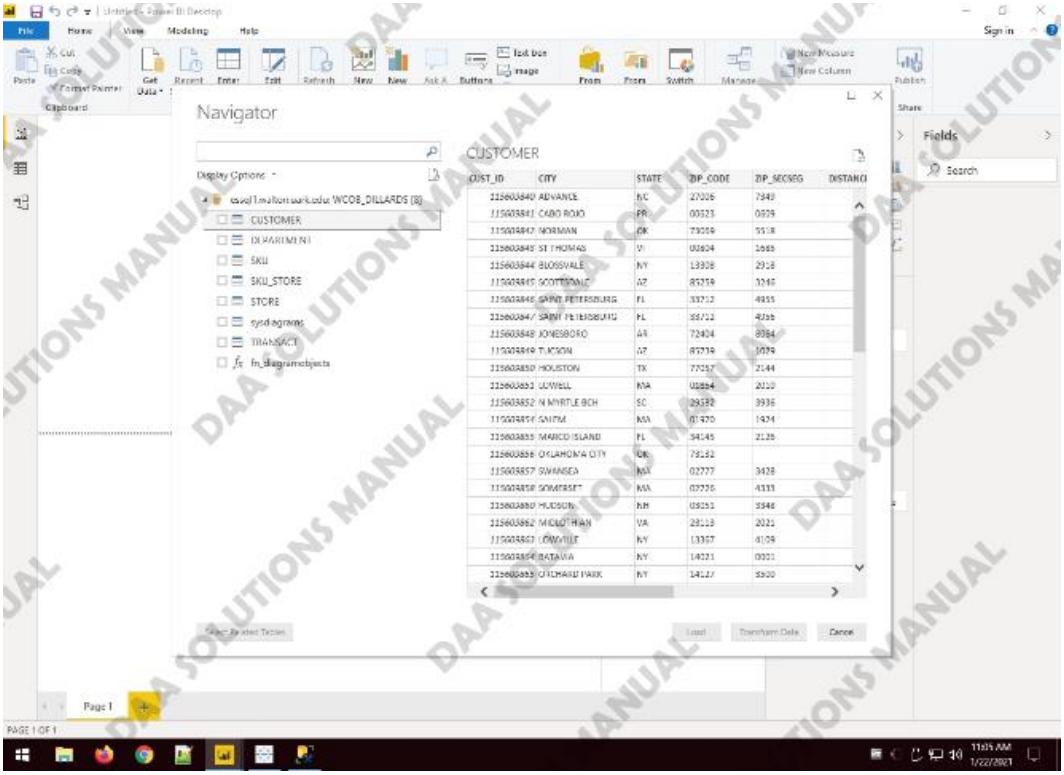
AQ3. Compare and Contrast: Compare viewing the data distribution and filtering data in Tableau Prep's Clean step to Microsoft Power BI's Data Profiling options. Which method is easier to interact with?

Answer: Student responses will vary.

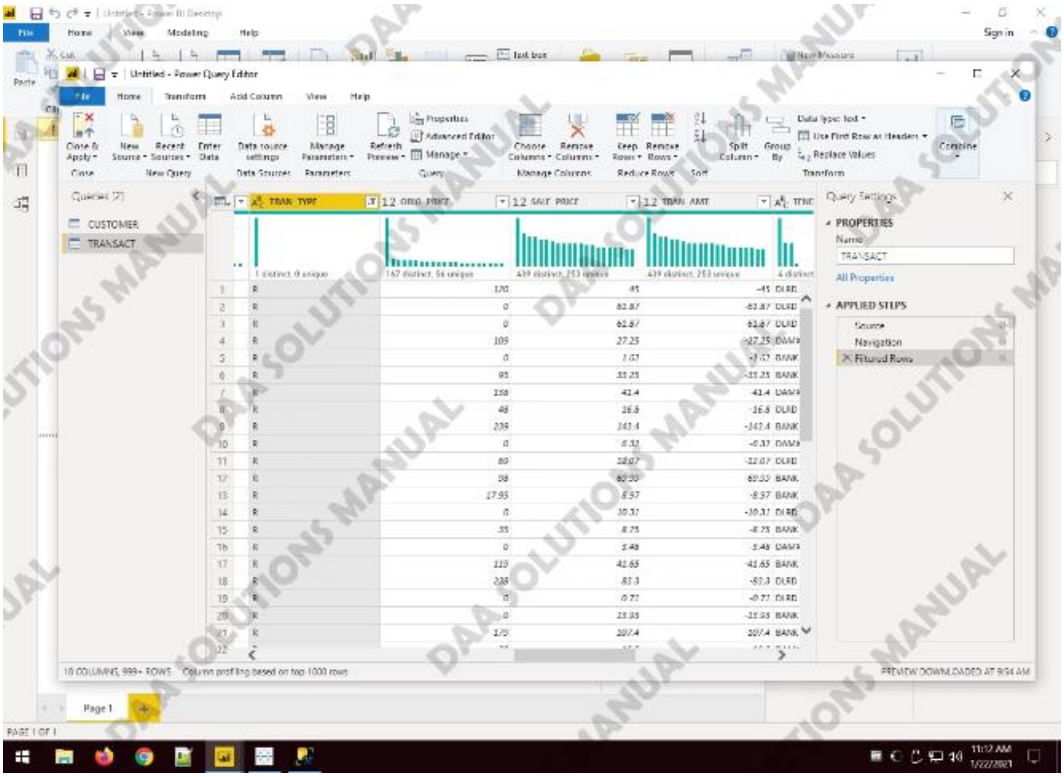
Lab 2-7 Submit Your Screenshot Lab Document

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.

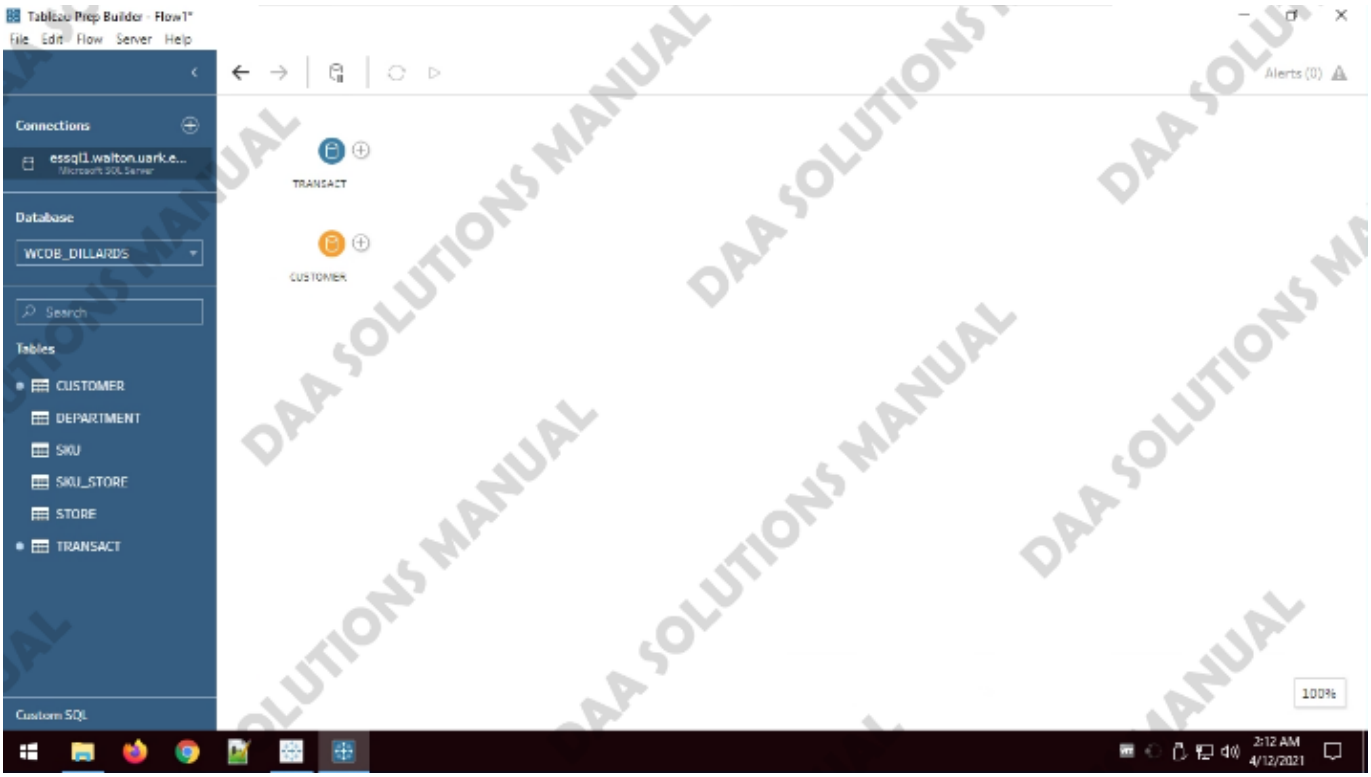
2-7MA



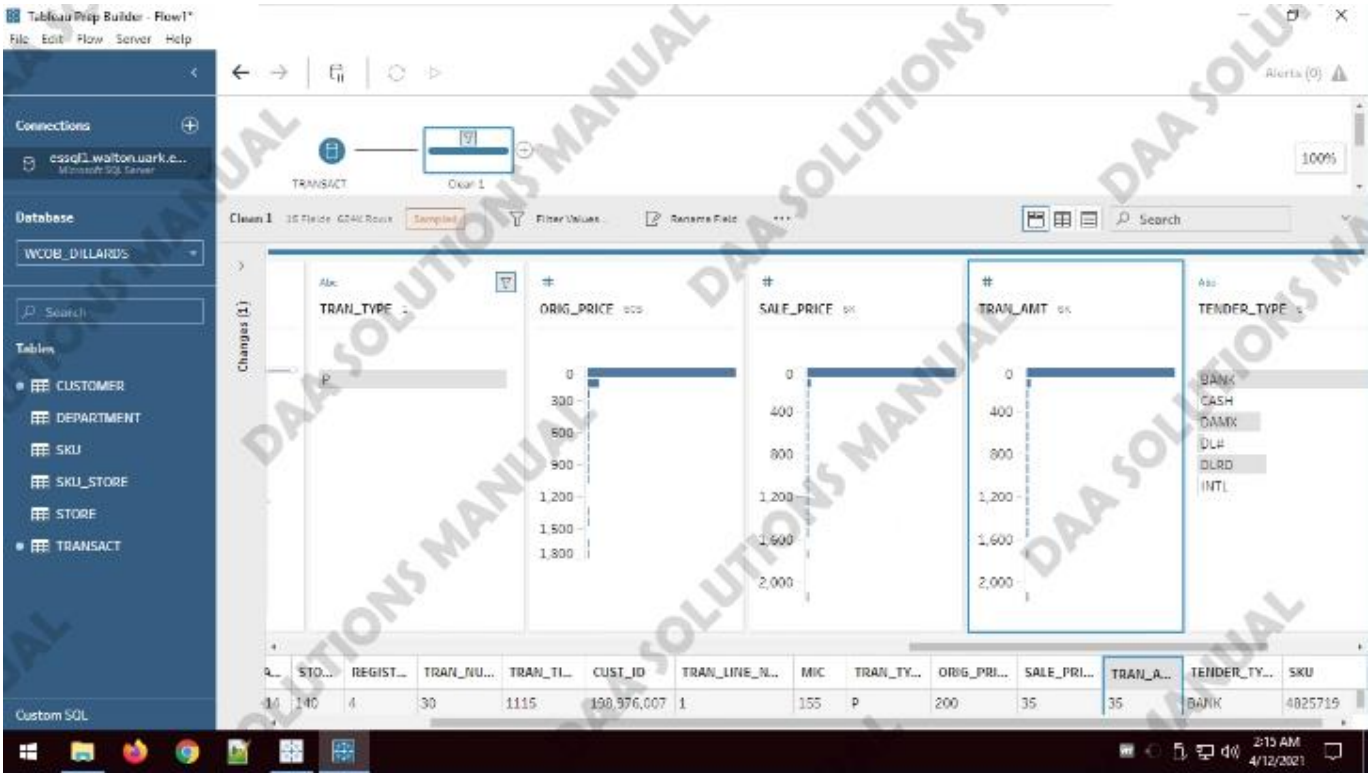
2-7MB



2-7TA



2-7TB



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,

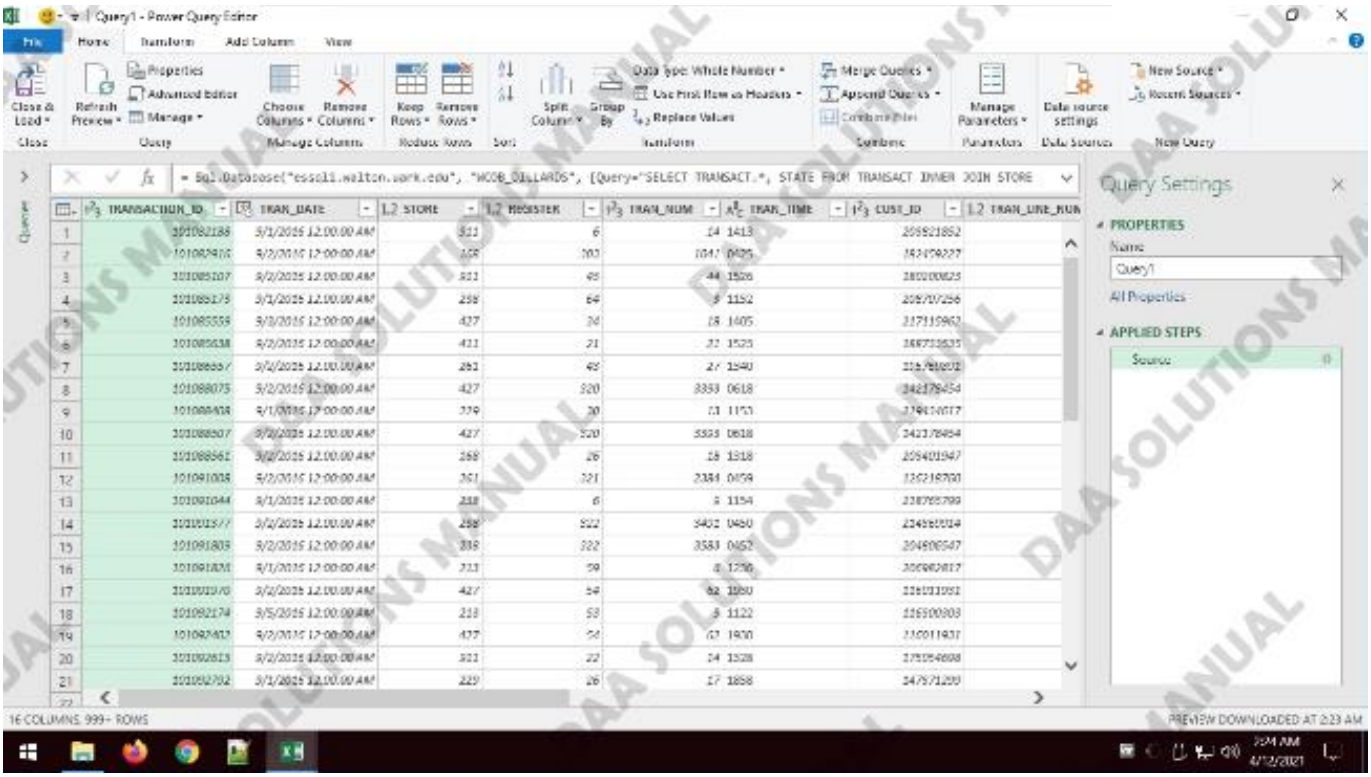
AQ3. Compare and Contrast: Compare the methods for analyzing transactions across states in Excel versus Tableau. Which tool was more intuitive for you to work with? Which provides more interesting results?

Answer: Student responses will vary.

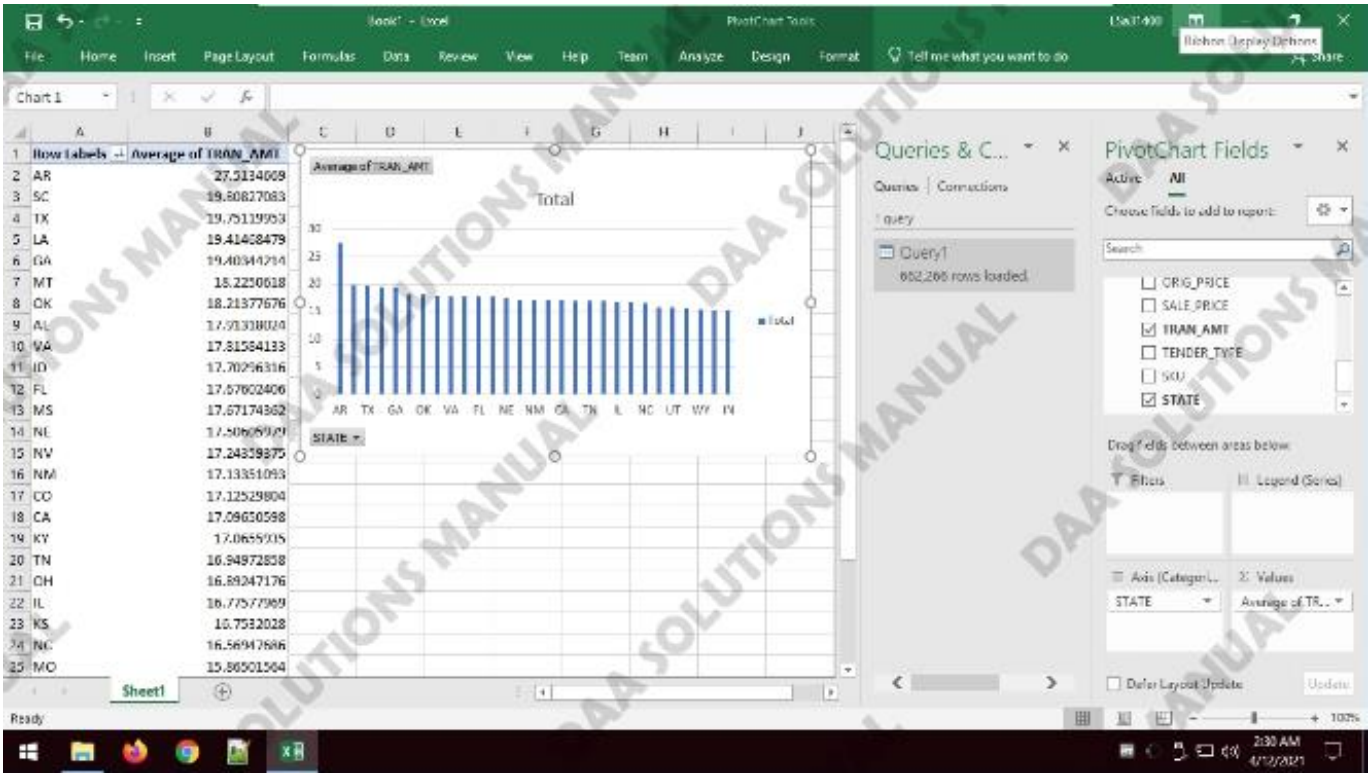
Lab 2-8 Submit Your Screenshot Lab Document

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.

2-8MA



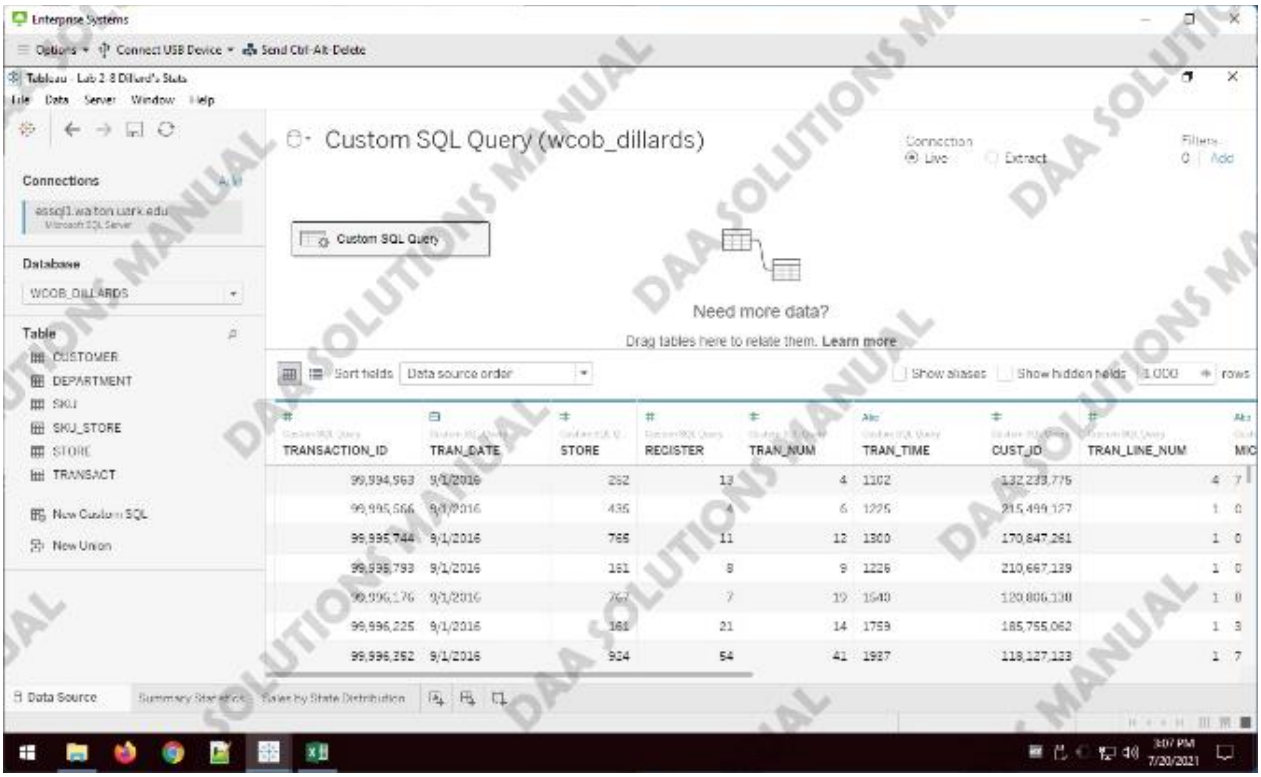
2-8MB



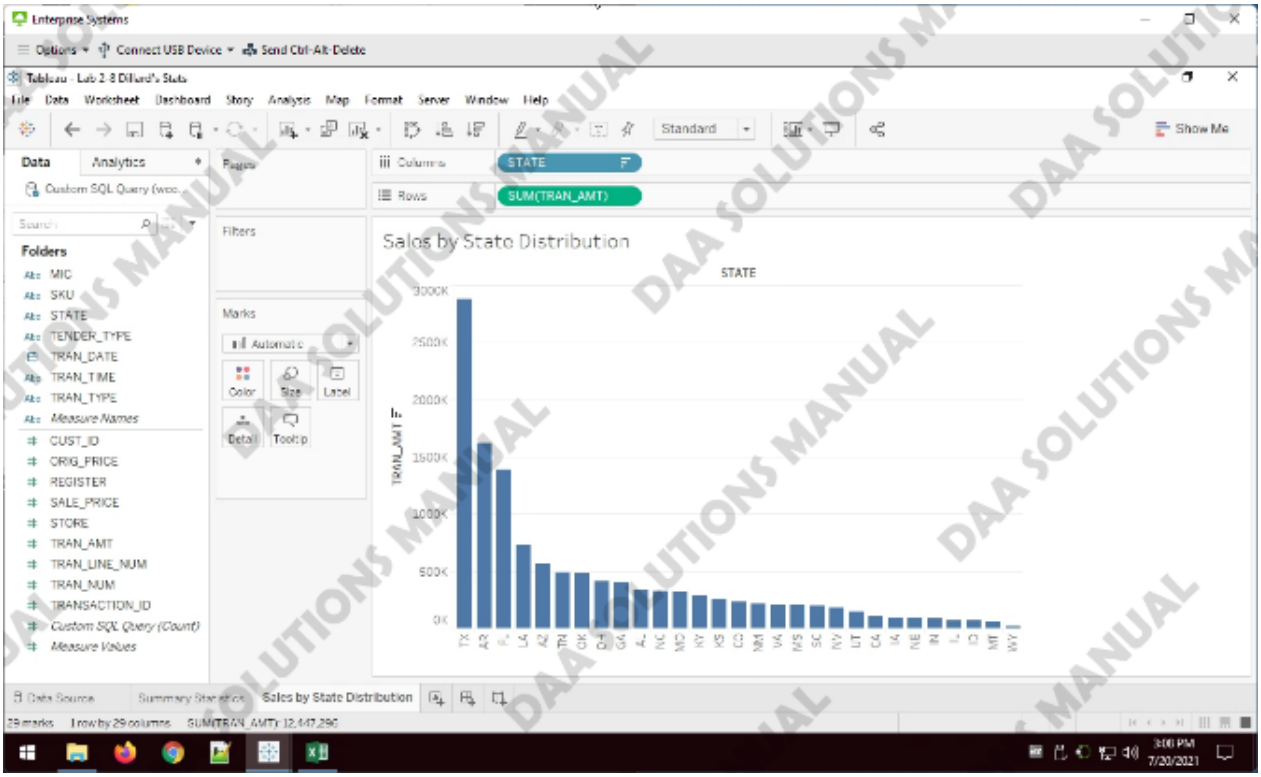
2-8MC

2-8MD

2-8TA



2-8TB



2-8TC