**Solutions Manual – Chapter 1**

**Solutions to Multiple Choice Questions**

1. (LO 1-1) Big Data is often described by the four Vs, or

1. volume, velocity, veracity, and variability.
2. volume, velocity, veracity, and variety.
3. volume, volatility, veracity, and variability.
4. variability, velocity, veracity, and variety.

Answer: b

2. LO 1-4) Which data approach attempts to assign each unit in a population into a small set of classes (or groups) where the unit best fits?

1. Regression
2. Similarity matching
3. Co-occurrence grouping
4. Classification

Answer: d

3. (LO 1-4) Which data approach attempts to identify similar individuals based on data known about them?

1. Classification
2. Regression
3. Similarity matching
4. Data reduction

Answer: c

4. (LO 1-4) Which data approach attempts to predict connections between two data items?

1. Profiling
2. Classification
3. Link prediction
4. Regression

Answer: c

5. (LO 1-6) Which of these terms is defined as being a central repository of descriptions for all of the data attributes of the dataset?

1. Big Data
2. Data warehouse
3. Data dictionary
4. Data Analytics

Answer: c

6. (LO 1-5) Which skills were *not* emphasized that analytic-minded accountants should have?

1. Developed an analytics mindset
2. Data scrubbing and data preparation
3. Classification of test approaches
4. Statistical data analysis competency

Answer: c

7. (LO 1-5) In which areas were skills *not* emphasized for analytic-minded accountants?

1. Data quality
2. Descriptive data analysis
3. Data visualization and data reporting
4. Data and systems analysis and design

Answer: d

8. (LO 1-4) The IMPACT cycle includes all *except* the following steps:

1. perform test plan.
2. visualize the data.
3. master the data.
4. track outcomes.

Answer: b

9. (LO 1-4) The IMPACT cycle specifically includes all *except* the following steps:

1. data preparation.
2. communicate insights.
3. address and refine results.
4. perform test plan.

Answer: a

10. LO 1-1) By the year 2024, the volume of data created, captured, copied, and consumed worldwide will be 149 \_\_\_\_\_\_\_\_ .

1. zettabytes
2. petabytes
3. exabytes
4. yottabytes

Answer: a

**Solutions to Discussion and Analysis Questions**

1. The accounting function is one of being an information provider. To the extent that data is available to address accounting questions, be they tax, managerial, audit or financial questions. With such rich available data, and software tools to prepare and analyze the data, data analytics will continue to be an important tool for accountants to use.
2. Data analytics is defined as the process of evaluating data with the purpose of drawing conclusions to address business questions. Indeed, effective Data Analytics provides a way to search through large structured and unstructured data to identify unknown patterns or relationships.

A university might learn from the analyzing the demographics of its current set of students in order to attract its future student recruits. Did they come from cities or high schools that were close by? Were their parents alumni of the university? Did they score high on certain parts of the ACT? Were those offered a scholarship more likely to attend, etc.? Was social media effective in attracting new, potentially stronger students? By analyzing this type of data, previously unknown patterns will emerge that will make recruiting students more effective.

1. There are many potential answers. For example, Monsanto may use mathematical and statistical models to plot out the best times to plant both male and female plants and where to plant them to maximize yield. (<https://www.cio.com/article/3221621/analytics/6-data-analytics-success-stories-an-inside-look.html#tk.cio_rs>)
2. There are many potential answers. Data analytics gives both internal and external auditors additional tools to examine every accounting transaction and assess for compliance with GAAP. The audit process is changing from a traditional process toward a more automated one, which will allow audit professionals to focus more on the logic and rationale behind data queries and less on the gathering of the actual data. No longer will they be simply checking for errors, material misstatements, fraud, and risk in financial statements or merely be reporting their findings at the end of the engagement. Instead, audit professionals will now be collecting and analyzing the company’s data similar to the way a business analyst would help management make better business decisions. In this way, data analytics offers value to the audit function.
3. There are many potential answers. For example, data analytics associated with financial reporting may help accountants determine if any of their inventory obsolete? It may also help the company benchmark on the financial statements and financial reporting of other similar companies and understand their accounting practices to help infer their own.
4. Management accountants address the information needs of management. They will often see what questions management has, find applicable data to address those questions, conduct analysis of the data, and report the results to management to help them make data-driven decisions. This is consistent with the data analytics process and the IMPACT model.
5. The IMPACT cycle suggests an order of 1) Identifying the Questions; 2) Mastering the Data; 3) Performing the test plan; 4) Addressing and refining results; 5) Communicating insights and 6) Tracking outcomes. The cycle starts with a question and then identifying data and test plan that might address that question. The results of the data analysis are communicated and tracked which may lead to additional, possibly more refined questions that then restart the cycle.
6. Data analysis is most effective when a question is identified that needs to be addressed. That will focus the analysis on which data and which test method might be most effective in addressing or answering the question.
7. Mastering the data requires one to know what data is available and whether it might be able to help address the business problem. We need to know everything about the data, including how to access it, its availability, how reliable it is (if there are errors), and what time periods it covers to make sure it coincides with the timing of our business problem, etc.
8. Facebook uses link prediction to predict a relationship between two people when it suggests people that one likely knows due to similar other friends, extended family, high schools, college or work locations, etc.
9. While sampling is useful, it is still just that, sampling. By looking at all of the transactions and testing them in a way that will highlight the ones that are the biggest dollar items, or are most unusual, that will allow auditors to focus on specific items that might be of material significance.
10. There are several correct answers. One data approach might be regression analysis where, given a balance of total accounts receivable held by a firm, how long it has been outstanding, if they have paid debts in the past all will help predict the appropriate level of allowance for doubtful accounts for bad debts.
11. The Debt-to-Income ratio might suggest to **LendingClub** that the person asking for the loan was simply asking for too big of a loan and they would have little ability to repay it. The lower the credit score, the less likely the potential borrower would be able to repay the loan.
12. There are many other potential predictors of whether the **LendingClub** would pay a loan. Here are a few possibilities: What other debt do they have? How much is their disposable income? Do they have a clean criminal record? Have they had a loan with **LendingClub** before and did they repay it? Do they rent or own their house?

**Solutions to Problems**

**Note: Some problems and solutions may be altered in Connect for auto grading purposes.**

1. (LO 1-4) Match each specific Data Analytics test to a specific test approach, as part of performing test plan:

* Classification
* Regression
* Similarity Matching
* Clustering
* Co-occurrence Grouping
* Profiling
* Link Prediction
* Data Reduction

|  |  |
| --- | --- |
| **Specific Data Analytics Test** | **Test Approach** |
| 1. Predict which firms will go bankrupt and which firms will not go bankrupt. | Classification |
| 1. Use stratified sampling to focus audit effort on transactions with greatest risk. | Data Reduction |
| 1. Work to understand normal behavior, to then be able identify abnormal behavior (such as fraud). | Profiling |
| 1. Look for relationships between related parties that are not otherwise disclosed. | Link Prediction |
| 1. Predict which new customers resemble the company’s best customers. | Similarity Matching |
| 1. Predict the relationship between an investment in advertising expenditures and subsequent operating income. | Regression |
| 1. Segment all of the company’s customers into groups that will allow further specific analysis. | Clustering |
| 1. The customers who buy product X will be most likely to be also interested in product Y. | Co-occurrence Grouping |

1. (LO 1-4) Match each of the specific Data Analytics tasks to the stage of the IMPACT cycle:

* Identify the Question
* Master the Data
* Perform Test Plan
* Address and Refine Results
* Communicate Insights
* Track Outcomes

|  |  |
| --- | --- |
| **Specific Data Analytics Task** | **Stage of IMPACT Cycle** |
| 1. Should we use company-specific data or macro-economic data to address the accounting question? | Master the Data |
| 1. What are appropriate cost drivers for activity-based costing purposes? | Identify the Question |
| 1. Should we consider using regression analysis or clustering analysis to evaluate the data? | Perform the Analysis |
| 1. Should we use tables or graphs to show management what we’ve found? | Communicate Insights |
| 1. Now that we’ve evaluated the data one way, should we perform another analysis to gain additional insights? | Address and Refine Results |
| 1. What type of dashboard should we use to get the latest, up-to-date results? | Track Outcomes |

1. (LO 1-5) Match the specific analysis need/characteristic to the appropriate Microsoft Track tool:

* Excel
* Power Query
* Power BI
* Power Automate

|  |  |
| --- | --- |
| **Specific Analysis Need/Characteristic** | **Microsoft Track Tool** |
| 1. Basic Visualization | Excel |
| 1. Robotics Process Automation | Power Automate |
| 1. Data joining | Power Query |
| 1. Advanced visualization | Power BI |
| 1. Works on Windows/Mac/Online platforms | Excel |
| 1. Dashboards | Power BI |
| 1. Collect data from multiple sources | Power Automate |
| 1. Data cleaning | Power Query |

1. (LO 1-5) Match the specific analysis need/characteristic to the appropriate Tableau Track tool:

* Tableau Prep Builder
* Tableau Desktop
* Tableau Public

|  |  |
| --- | --- |
| **Specific Analysis Need/Characteristic** | **Tableau Track Tool** |
| 1. Advanced Visualization | Tableau Desktop |
| 1. Analyze and share public datasets | Tableau Public |
| 1. Data joining | Tableau Prep Builder |
| 1. Presentations | Tableau Desktop |
| 1. Data transformation | Tableau Prep Builder |
| 1. Dashboards | Tableau Desktop |
| 1. Data cleaning | Tableau Prep Builder |

1. (LO 1-6) Here are the predictive attributes and whether they would be applicable to predicting which loans would be delinquent and which loans will ultimately be fully repaid.

|  |  |
| --- | --- |
| **Predictive Attributes** | **Predictive? (Yes/No)** |
| 1. date (Date when the borrower accepted the offer) | No |
| 1. desc (Loan description provided by borrower) | No |
| 1. dti (A ratio of debt owed to income earned) | Yes |
| 1. grade (LC assigned loan grade) | Yes |
| 1. home\_ownership (Values include Rent, Own, Mortgage, Other) | Yes |
| 1. loanAmnt (Amount of the loan) | Yes |
| 1. next\_pymnt\_d (Next scheduled payment date) | No |
| 1. term (Number of payments on the loan) | No |
| 1. tot\_cur\_bal (Total current balance of all accounts) | Yes |

1. (LO 1-6) Navigate to the Connect Additional Student Resources page. Under [Chapter 1](file:///C:\Users\michael_mccormick\Documents\000%20Accounting\Richardson%20DAA%203e%20(Steve)\Richardson%20DAA%203e%20Connect%20Solutions%20Manuals%20and%20Labs\1c72941d063247d1a96ee739f461c897) Data Files, download and consider the rejected loans dataset of **LendingClub** data titled “DAA Chapter 1-1 Data”. Choose among these attributes in the data dictionary, and indicate which are likely to be predictive of loan rejection, and which are not.

|  |  |
| --- | --- |
| **Predictive Attributes** | **Predictive? (Yes/No)** |
| 1. Amount Requested | Yes |
| 1. Zip Code | Yes |
| 1. Loan Title | No |
| 1. Debt-To-Income Ratio | Yes |
| 1. Application Date | No |
| 1. Risk\_Score | Yes |
| 1. Employment Length | Yes |

7A) **Multiple Choice:** What is the percentage of total loans rejected in the United States that came from Arkansas?

1. Less than 1%.
2. Between 1% and 2%.
3. More than 2%.

Answer: b

Percentage of total loans rejected that live in Arkansas = 1.219%

7B) **Multiple Choice:** Is this loan rejection percentage greater than the percent of the U.S. population that lives in Arkansas (per 2010 census)?

1. Loan rejection percentage is greater than the population.
2. Loan rejection percentage is less than the population.

Answer: a

2,915,918 population in Arkansas divided by USA population of 308,745,538 = 0.9444%

The loan rejection percentage is greater than the percent of the USA population that lives in Arkansas (per 2010 census), but is reasonably close.

1. (LO 1-6) Download the rejected loans dataset of LendingClub data titled “DAA Chapter 1-1 Data” from Connect Additional Student Resources and do an Excel PivotTable by state; then figure out the number of rejected applications for each state.

8A) Put the following states in order of their loan rejection percentage based on the count of rejected loans (from high [1] to low [11]) of the total rejected loans.

|  |  |
| --- | --- |
| **State** | **Rank 1 (High) to 11 (Low)** |
| 1. Arkansas (AR) | 2 |
| 1. Hawaii (HI) | 9 |
| 1. Kansas (KS) | 5 |
| 1. New Hampshire (NH) | 10 |
| 1. New Mexico (NM) | 8 |
| 1. Nevada (NV) | 1 |
| 1. Oklahoma (OK) | 3 |
| 1. Oregon (OR) | 4 |
| 1. Rhode Island (RI) | 11 |
| 1. Utah (UT) | 6 |
| 1. West Virginia (WV) | 7 |

8B) What is the state with the highest percentage of rejected loans?

Answer: CA

8C) What is the state with the lowest percentage of rejected loans?

Answer: ND

|  |  |
| --- | --- |
| State | Loan Rejection % |
| CA | 0.13292708 |
| TX | 0.08344411 |
| NY | 0.0797736 |
| FL | 0.07688089 |
| PA | 0.04401981 |
| IL | 0.04246422 |
| OH | 0.03779744 |
| NJ | 0.03708008 |
| GA | 0.03683527 |
| VA | 0.03131478 |
| MI | 0.02718255 |
| NC | 0.02672393 |
| MA | 0.02547822 |
| MD | 0.02340048 |
| AZ | 0.02142811 |
| MO | 0.01954559 |
| WA | 0.0187585 |
| CO | 0.01812325 |
| AL | 0.0169798 |
| CT | 0.01640652 |
| SC | 0.01569535 |
| LA | 0.01450077 |
| WI | 0.01430865 |
| MN | 0.01407314 |
| KY | 0.01367649 |
| NV (1) | 0.01275305 |
| AR (2) | 0.01219062 |
| OK (3) | 0.01103943 |
| OR (4) | 0.00954581 |
| KS (5) | 0.00862547 |
| UT (6) | 0.00692579 |
| WV (7) | 0.00643153 |
| NM (8) | 0.00590939 |
| HI (9) | 0.005756 |
| NH (10) | 0.00551739 |
| RI (11) | 0.00498905 |
| DE | 0.00354346 |
| MT | 0.00284933 |
| VT | 0.00250537 |
| AK | 0.00249142 |
| DC | 0.00236128 |
| SD | 0.00223887 |
| WY | 0.00220479 |
| IN | 0.00149516 |
| MS | 0.00059962 |
| TN | 0.00055003 |
| NE | 0.00022311 |
| IA | 0.00017043 |
| ME | 0.0001379 |
| ID | 8.0568E-05 |
| ND | 4.6482E-05 |

8D) **Analysis:** Does each state’s loan rejection percentage roughly correspond to their relative proportion of the U.S. population (by 2010 U.S. census at <https://en.wikipedia.org/wiki/2010_United_States_census>)?

The loan rejection percentage roughly corresponds with the population of each state. However, there is still substantial variation between the rejection percentage of each state.

**For** [**Problems 9**](file:///C:\Users\michael_mccormick\Documents\000%20Accounting\Richardson%20DAA%203e%20(Steve)\Richardson%20DAA%203e%20Connect%20Solutions%20Manuals%20and%20Labs\1fc54fe8367d46b3a1793d47c3ab03c1)**,** [**1**](file:///C:\Users\michael_mccormick\Documents\000%20Accounting\Richardson%20DAA%203e%20(Steve)\Richardson%20DAA%203e%20Connect%20Solutions%20Manuals%20and%20Labs\0e7386e2a12b48b8a3a6b3d606fbdaa1)**0, and** [**1**](file:///C:\Users\michael_mccormick\Documents\000%20Accounting\Richardson%20DAA%203e%20(Steve)\Richardson%20DAA%203e%20Connect%20Solutions%20Manuals%20and%20Labs\6d987c73c43642f9bc71bb805680239c)**1, we will be cleaning a data file in preparation for subsequent analysis.**

Here is the pivot table by risk score grouping:

|  |  |
| --- | --- |
| **Row Labels** | **Count of Loan Title** |
| Excellent | 2931 |
| Fair | 236669 |
| Good | 83543 |
| Poor | 189621 |
| Very Bad | 145322 |
| Very Good | 11907 |
| **Grand Total** | **669993** |

|  |  |
| --- | --- |
| **Question** | **Group** |
| Which group had the most observations? | Fair |
| Which group had the least observations? | Excellent |

The Excellent category had the smallest group, whereas the Fair group had the biggest group. Arguably there is a greater population of Fair, even though Very Bad has a smaller count, it is clearly the worst of the group.

The results from the 2013 data are similar to the data from 2007-2012.

Here is the pivot table by Debt-to-Income (DTI) grouping:

|  |  |
| --- | --- |
| **Row Labels** | **Count of Amount Requested** |
| High | 340615 |
| Low | 159464 |
| Medium | 169914 |
| **Grand Total** | **669993** |

|  |
| --- |
|  |
|  | |  |  | | --- | --- | | **Question** | **Group** | | Which group had the most observations? | High | | Which group had the least observations? | Low | |

Low DTI is the smallest grouping whereas High DTI has the largest grouping.

The higher the debt as compared to income, the less likely the potential borrower would be able to pay back the loan from Lending Club. Since this dataset is for rejected loans, the higher the DTI ratio, the more likely to be rejected for a loan.

11.

Here is the pivot table for the loans with excellent risks but high debt-to-incomes, by years of employment:

|  |  |
| --- | --- |
| **Row Labels** | **Count of Amount Requested** |
| **Excellent** | **2931** |
| **High** | **1190** |
| 0 | 942 |
| 1 | 12 |
| 2 | 14 |
| 3 | 11 |
| 4 | 12 |
| 5 | 92 |
| 6 | 9 |
| 7 | 15 |
| 8 | 9 |
| 9 | 6 |
| 10 | 68 |

|  |  |  |
| --- | --- | --- |
| **Question** | **Level** | |
| Which employment year group had the most observations to go along with excellent risk scores but high debt-to-income? | 0 | employment years |
| Which employment year group had the least observations to go along with excellent risk scores but high debt-to-income? | 9 | employment years |

Perhaps those with excellent credit just asked for too big of a loan given their existing debt and that is why they are rejected. This PivotTable analysis suggests those with excellent credit asked for a larger loan given the debt they already had as compared to any of the others, suggesting a reason why even those potential borrowers with excellent credit were rejected. There weren’t a lot of them, but there were certainly some!