Solutions Manual for Business Analytics 2025 Release 1st Edition by Jaggia, Kelly, Lertwachara, Chen

ISBN: 9781266926020

Chapter 1. Introduction to Business Analytics

Solutions

1.

   a. The relevant population consists of all Americans.

   b. The estimates for averages are based on sample data.

2. The value 35 is the estimated average age of the population. It is both costly and time consuming (likely impossible) to take a census of all video game players and compute the actual average age.

3.

   a. The population is all students enrolled in the accounting class.

   b. The value 3.29 represents the population parameter since we are not choosing a sample but drawing results from the actual population.

4.

   a. The population is all recent college graduates with an engineering degree.

   b. No, the average salary is a sample statistic computed from a sample, not the population.

5.

   a. The population is all elderly people. The sample consists of 949 elderly people.

   b. 22% and 17% represent sample statistics.

6. The data are cross-sectional data.

   Note: Data will vary due to the nature of sampling.

7. The data are time series data.

   Note: Data will vary depending on the date of retrieval. These data were retrieved July 2019.

Date	Adj. Close Price
7/1/2018	192.55
8/1/2018	195.72
9/1/2018	202.97
10/1/2018	172.33

11/1/2018	176.68
12/1/2018	169.36
1/1/2019	180.90
2/1/2019	182.49
3/1/2019	189.15
4/1/2019	202.28
5/1/2019	188.53
6/1/2019	206.52

8. The front page of the New York Times website is likely to be textual (written reports) with multimedia contents (photographs, etc.). The resulting data are unstructured in that they do not conform to a predefined row-column format.

9. Data on price and fuel economy of small hybrid vehicles can be specified in a predefined row-column format, and therefore, are structured.

   Note: Data will vary due to the nature of sampling.

10. The data for Under Armour's annual revenue are structured since they are specified in a well-defined row-column format. The data are time series data.

    Note: Data will vary depending on the date of retrieval. These data were retrieved July 2019.

Year	Annual Revenue (in $ millions)
2018	$5,193
2017	$4,989
2016	$4,833
2015	$3,963
2014	$3,084
2013	$2,332
2012	$1,835
2011	$1,473
2010	$1,064
2009	$856

11. The resulting data about online social media usage have a well-defined row-column format, and therefore, are structured. The data are cross-sectional data.

    Note: Data will vary due to the nature of sampling.

12. The resulting data about living accommodations and expenses have a well-defined row-column format, and therefore, are structured. The data are cross-sectional data.

    Note: Data will vary due to the nature of sampling.

13.

State	Median Household Income in 2013-2017 ($)

Alabama	46,472
Arizona	53,510
California	67,169
Florida	50,883
Georgia	52,977
Indiana	52,182
Iowa	56,570
Maine	53,024
Massachusetts	74,167
Minnesota	65,699
Mississippi	42,009
New Mexico	46,718
North Dakota	61,285
Washington	66,174

Source: *http://www.census.gov/* ; Retrieved July 1, 2019.

These data are estimates for the period 2013-2017. The data are cross-sectional data. In this particular group of states, Massachusetts has the highest median income, whereas Mississippi has the lowest median income. Also, states in the North such as Massachusetts, Minnesota and Washington tend to have higher incomes than Southern states.

14.

a. Numerical; discrete
b. Categorical
c. Numerical; continuous

15.

a. Categorical
b. Numerical; continuous
c. Numerical; discrete

16.

a. Nominal
b. Interval
c. Ordinal

17.

a. Ratio
b. Ordinal

c. Nominal

18.

a. Ratio

b. Interval

c. Ratio

19.

a. Nominal scale; the values differ in name.

b.

Major	# of Students
Accounting	4
Economics	4
Finance	4
History	4
Management	5
Psychology	3
Statistics	3
Undecided	3

c. An inspection of the data shows that Management has the highest number of students whereas Psychology, Statistics, and Undecided have the lowest.

20.

a. Nominal

b. Interval; the observations for Year can be ranked, categorized and measured when using this kind of scale. However, there is no true zero point so we cannot calculate meaningful ratios between years.

c. Ratio; this type of scale is the strongest form of measurement. There is a true zero point which allows for the calculation of meaningful ratios between observations.

21.

a. The variable Sex is categorical whereas Income, Age, and Spending are numerical.

b. The measurement scale is nominal for Sex and ratio for Income, Age, and Spending. Recall, that the nominal and ratio scales represent the least and most sophisticated levels of measurement, respectively.

22.

a. The variable Year is measured on the interval scale because the observations can be ranked, categorized, and measured when using this kind of scale. However, there is no true zero point so we cannot calculate meaningful ratios between years.

b. The variable Quarter is measured on the nominal scale, even though it contains numbers. It is the least sophisticated level of measurement because if we are presented with nominal data, all we can do is categorize or group the data.

c. The variable Vacation is measured on the ratio scale. It is the strongest level of measurement because it allows us to categorize and rank the data as well as find meaningful differences between observations. Also, with a true zero point, we can interpret the ratios between observations.

23.

a.

Name, Sold

Mercedes, 02

Toyota, 03

Ford, 06

Hyundai, 04

b.

Name, Sold

Mercedes, 02

Toyota, 03

Ford, 06

Hyundai, 04

c.

```
<Data> <Manufacturer> <Name>Mercedes</Name> <Sold>2</Sold>
</Manufacturer> <Manufacturer> <Name>Toyota</Name>
<Sold>3</Sold> </Manufacturer ><Manufacturer> <Name>Ford</Name>
<Sold>6</Sold> </Manufacturer> <Manufacturer>
<Name>Hyundai</Name> <Sold>4</Sold> </Manufacturer>
</Data>
```

d.

```
<table>
<tr>
<th>Name</th>
<th>Sold</th>
</tr>
<tr>
<td>Mercedes</td>
<td>2</td>
</tr>
<tr>
<td>Toyota</td>
<td>3</td>
</tr>
<tr>
<td>Ford</td>
<td>6</td>
</tr>
</tr>
<tr>
<td>Hyundai</td>
<td>4</td>
</tr>
</table>
```

e.

```
{
“Manufacturer”: [
{
“Name”: “Mercedes”,
“Sold”: “2”,
},
{
“Name”: “Toyota”,
“Sold”: “3”,
},
{
“Name”: “Ford”,
“Sold”: “6”,
},
      {
“Name”: “Hyundai”,
“Sold”: “4”,
},
```

```
]
}
```

24.

a.

FirstName LastName Salary
Robert Schneider 56000
Chun Zhang 52000
Sunil Banerjee 58000
Linda Jones 60000

b.

FirstName,LastName,Salary
Robert,Schneider,56000
Chun,Zhang,52000
Sunil,Banerjee ,58000
Linda,Jones,60000

c.

```
<Data> <Employee> <FirstName>Robert</FirstName>
<LastName>Schneider</LastName> <Salary>56000</Salary>
</Employee> <Employee>
<FirstName>Chun</FirstName>
<LastName>Zhang</LastName> <Salary>52000</Salary> </Employee>
<Employee>
<FirstName>Sunil</FirstName>
<LastName>Banerjee</LastName> <Salary>58000</Salary> </Employee>
<Employee>
<FirstName>Linda</FirstName>
<LastName>Jones</LastName> <Salary>60000</Salary> </Employee>
</Manufacturer>
```

d.

```
<table>
<tr>
<th>FirstName</th>
<th>LastName</th>
<th>Salary</th>
</tr>
<tr>
<td>Robert</td>
<td>Schneider</td>
<td>56000</td>
</tr>
<tr>
<td>Chun</td>
```

```
<td>Zhang</td>
<td>52000</td>
</tr>
<tr>
<td>Sunil</td>
<td>Banerjee</td>
<td>58000</td>
</tr>
<tr>
<td>Linda</td>
<td>Jones</td>
<td>60000</td>
</tr>
</table>
```

e.

```
{
"Employee": [
{
"FirstName": "Robert",
"LastName": "Schneider",
"Salary": "56000",
},
{
"FirstName": "Chun",
"LastName":Zhang",
"Salary": "52000",
},
      {
"FirstName": "Sunil",
"LastName":Banerjee",
"Salary": "58000",
{
"FirstName": "Linda",
"LastName":Jones",
"Salary": "60000",
},
      {

]
}
```

25.

a.

Year India China
2013 1278.56 1357.38

2014 1293.86 1364.27
2015 1309.05 1371.22
2016 1324.17 1378.67
2017 1339.18 1386.40

b.

Year,India,China
2013,1278.56,1357.38
2014,1293.86,1364.27
2015,1309.05,1371.22
2016,1324.17,1378.67
2017, 1339.18,1386.40

26.

a.

Country Happiness GDP
Finland 7.769 45670
Denmark 7.600 57533
Norway 7.544 75295
Iceland 7.494 73060
Netherlands 7.488 48754

b.

Country,Happiness,GDP
Finland,7.769,45670
Denmark,7.600,57533
Norway,7.544,75295
Iceland,7.494,73060
Netherlands,7.488,48754

27.

a.

Name University
Bridget Yale
Minori Stanford
Matthew Harvard

b.

Name,University
Bridget,Yale
Minori,Stanford
Matthew,Harvard

c.

```
<Data> <Student> <Name>Bridget</Name>
<University>Yale</University> </Student> <Student>
<Name>Minori</Name> <University>Stanford</University>
```

```
<Student> <Name>Matthew</Name> <University>Harvard</University>
</Student> </Data>
```

d.

```
<table>
<tr>
<th>Name</th>
<th>University</th>
</tr>
<tr>
<td>Bridget</td>
<td>Yale</td>
</tr>
<tr>
<td>Minori</td>
<td>Stanford</td>
</tr>
<tr>
<td>Matthew</td>
<td>Harvard</td>
</tr>
</table>
```

e.

```
{
"Student": [
{
"Name": "Bridget",
"University": "Yale",
},
{
"Name": "Minori",
"University": "Stanford",
},
{
"Name": "Matthew",
"University": "Havard",
},
]
}
```

28.

a.

```
<Data> <Player>
<Name>Michael Trout</Name>
<Team>Angels</Team> <Salary>39000000</Salary> </Player> <Player>
```

```
<Name>Bryce Harper</Name>
<Team>Phillies</Team> <Salary>36500000</Salary>
<Player>
<Name> David Price </Name>
<Team>Red Sox</Team> <Salary>32000000</Salary> </Player> </Data>
```

b.

```
<table>
<tr>
<th>Name</th>
<th>Team</th>
<th>Salary</th>
</tr>
<tr>
<td>Michael Trout</td>
<td>Angels</td>
<td>39000000</td>
</tr>
<tr>
<td>Bryce Harper</td>
<td>Phillies</td>
<td>36500000</td>
</tr>
<tr>
<td>David Price</td>
<td>Red Sox</td>
<td>32000000</td>
</tr>
</table>
```

c.

```
{
"Player": [
{
      "Name": "Michael Trout",
"Team": "Angels",
      "Salary": "39000000",
},
{
      "Name": "Bryce Harper",
"Team": "Phillies",
      "Salary": "36500000",
},
{
      "Name": "David Price",
"Team": "Red Sox",
```

```
      "Salary": "32000000",
},
]
}
```

29.

a.

```
<Data> <Business Profession>
<Profession>Statistician</Profession>
<Jobs>12600</Jobs> <Salary>84060</Salary> </Business Profession>
<Business Profession>
<Profession>Mathematician</Profession>
<Jobs>900</Jobs> <Salary>103010</Salary> </Business Profession>
</Data>
```

b.

```
<table>
<tr>
<th>Profession</th>
<th>Jobs</th>
<th>Salary</th>
</tr>
<tr>
<td>Statistician</td>
<td>12600</td>
<td>84060</td>
</tr>
<tr>
<td>Mathematician</td>
<td>900</td>
<td>103010</td>
</tr>
</table>
```

c.

```
{
"Business Profession": [
{
      "Profession": "Statistician",
"Jobs": "12600",
      "Salary": "84060",
},
{
      "Profession": "Mathematician",
"Jobs": "900",
      "Salary": "103010",
```

```
},
]
}
```

Possible Output for Suggested Case Studies

There are many ways to approach these case studies. We provide numerical guidance for one possible scenario.

Report 1.1

Based on the 2019-2021 rankings, Finland, Denmark, and Iceland are among the happiest countries in the world. Consistent with the past data, Scandinavian countries continue to have favorable rankings compared to other countries. In fact, Finland has occupied the top spot in the Happiness Index Report over the past five years. The report is based on aggregate data on quality of life (e.g., GDP per capita and life expectancy) and a survey of citizens in each country including questions about positive emotions (e.g., enjoyment and laughter) and negative emotions (e.g., anger and sadness). Below is the listing of the top 10 happiest countries according to the 2019-2021 report.

Rank	Country
1	Finland
2	Denmark
3	Iceland
4	Switzerland
5	Netherlands
6	Luxembourg
7	Sweden
8	Norway
9	Israel
10	New Zealand

Report 1.2 and Report 1.3

We do not provide numerical guidance for these case studies because the numbers will vary, depending on when the data are extracted.

Business Analytics Instructor Resource Guide

Chapter 1: Introduction to Business Analytics

Learning Objectives:

LO 1.1: Explain the importance of business analytics, data privacy, and data ethics.

LO 1.2: Explain the various types of data.

LO 1.3: Describe variables and the types of measurement scales.

LO 1.4: Describe different data sources and file formats.

Big Picture:

Business analytics is "the methodology of extracting information and knowledge from data that improves a company's bottom line and enhances consumer experience." The focus is on how the analyst derives information from the data that can help the organization. The focus of the analyst has shifted (somewhat) from 'how can I get more, rich data?' to 'how can I use the myriad data I have to improve the business?' To answer this question, a clear understanding of the data is key, as that will drive the methods applied.

Chapter Summary:

Section 1.1: Overview of Business Analytics

Business analytics is about using data in addition to other business knowledge to help make decisions to improve the business. It focuses attention on problem solving which is the focal point of any business venture. The 'flavors' of business analytics (descriptive, predictive, and prescriptive) are often used both separately and together to solve large problems. The way data is collected, protected, and analyzed (data privacy and ethics) is crucial to the entire business analytics workflow.

Section 1.2 Types of Data

Most data collected are a 'sample' of the population of interest. Knowledge of the sampling process is key for analyzing these data and will be discussed in future chapters. Cross-sectional data provide a snapshot in time of elements of the population, while time-series data allow for analysis of the progression of the variables. The type of question being proposed often dictates how the data must be collected. Structured data follow a traditional pre-defined tabular format, which is fairly straightforward to analyze with traditional techniques. Unstructured data are often textual and open-ended, which requires new methods of analysis that are often more complex. While big data isn't a centerpiece of the text, no analytics text would be complete without discussing the implications of big data. When referring to 'big data,' we are referring to data so immense that they are too difficult to manage, process, and analyze

in a traditional framework. For a deeper understanding of this concept, the 3Vs of big data (volume, velocity, and variety) are discussed.

Section 1.3 Variables and Scales of Measurement

How the data are described in the dataset will determine which analysis is appropriate. The first demarcation is between categorical (qualitative) and numerical (quantitative) data. Categorical data represent categories, while numerical variables represent meaningful numbers. Numerical variables can be further broken down into discrete (countable number of possible values) and continuous (uncountable number of values in an interval). Examples of discrete data are number of red cars in a lot (you cannot find 1.23 red cars) or number of employees. Examples of continuous data are height, weight, income, etc. While most continuous variables are 'discretized' (e.g., round to the nearest lb for weight), we often treat measurements with many options that are not restricted to be integers as continuous. Another way to describe data is using the scale of measurement. The four types are: nominal (unordered categorical), ordinal (ordered categorical), interval (numerical with an arbitrary zero), and ratio (traditional numerical variables).

Section 1.4 Data Sources and File Formats

Fixed-width and delimited formats are traditional spreadsheet formats most students will be familiar with. These are quite useful for tabular structured data. The markup languages (XML and HTML) are briefly discussed concerning structured data using tags. JSON is also introduced as another structured format used in sharing amongst computer programs.

Critical concepts:

- Business analytics is a confluence of qualitative reasoning (subject matter expertise) and quantitative tools to identify, solve, and communicate solutions to key business problems.
- There are three main flavors of business analytics: descriptive (what has happened), predictive (what will likely happen), and prescriptive (what should we do to maximize output).
- Understanding the use and misuse of data from a privacy and ethical point of view is essential.
- Sometimes, collecting an entire population of data is impossible (time and/or money), so a sample is utilized. Future chapters will examine the implications of sampling in the analysis.
- Cross-sectional data provide a snapshot in time (what is happening at a specific moment), while time-series or longitudinal data can show the progression of a variable over time.

- Structured data contain a predefined row-column format (standard tabular data), whereas unstructured data tend to be more open-ended, textual information.
- Big data refers to massive amounts of data that are quite difficult to manage, process, and analyze using traditional data-processing tools.
- Scale of measurement can determine appropriate analytical techniques. It is very important to understand the types of data (nominal, ordinal, numerical, etc.), as they have a downstream effect on the analyses to be performed.
- Data can be expected in myriad formats. Familiarization with the different formats can be helpful in cleaning/organizing data.

Key Terminology:

- Business analytics
    - Descriptive analytics
    - Predictive analytics
    - Prescriptive analytics
- Data privacy
- Data ethics
- Sample vs. population
- Cross-sectional vs. time-series
- Structured vs. unstructured
- Big data: volume, velocity, variety, veracity, value
- Discrete vs. continuous
- Categorical vs. numerical
- Nominal/ordinal/interval/ratio
- Fixed-length/delimited/XML/HTML/JSON

How I would teach this chapter:

This chapter provides a foundation to the entire book. Defining 'business analytics' has become a point of contention, but the definition given in the book summarizes its key points well. I would emphasize the need to understand your data at a fundamental level (type, structure, etc.) before beginning to integrate it into a larger framework.

Section 1.1 centers around the purpose of the field as a whole. Depending on the length/structure of your specific course, you may only focus on one or two of the types of analytics, which is possible due to the structure of the book. Keep in mind how the descriptive, predictive, and prescriptive techniques often inform each other to provide a consistency in the business analytics framework. The figure on page 5 gives a nice overview of this process. My focus would be on *understanding* the problem and seeking techniques (taught throughout the

book) to solve that problem. The problem-solving framework introduced must center on the business case and how the analysis will benefit the business. Data privacy and ethics should also be emphasized here as this is becoming increasingly important in the current world.

Section 1.2 is about understanding data and how the type of data is important in your future analyses. Courses that are shorter in length may omit unstructured data analysis, but an analyst should be aware of when the data is unstructured vs. structured. An understanding of sampling vs. census (population level data) is important. Big data discussions are more peripheral for a short course. I believe knowledge of big data's 'existence' is nice, but this text isn't focused on the true study of big data.

Section 1.3 looks at scales of variable measurement: discrete vs. continuous, cross-sectional vs. time-series, and categorical vs. numerical. These broad categories drive all future analyses so all students should have a clear understanding of these differences. In describing the hierarchy of nominal, ordinal, interval, and ratio, I would spend my time focused on students understanding the difference between nominal, ordinal, and ratio, as interval data is fairly rare (temperature is the classic example).

Section 1.4 contains more than most will need for introductory business analytics courses. The simple information about fixed-width and delimited formats should be discussed in all classes. I would reserve more complex XML, HTML, and JSON discussions for classes with a heavier computer science prerequisite or data science major.

This chapter is essential to any business analytics course as an introduction to the course. I would include Section 1.1–1.3 in any business analytics course. Section 1.4's discussion of fixed-width format and delimited format is useful for a lower level course, but the remainder can be skipped. In a shorter course, I would also skip 'big data.'