Predictive Analytics for Business Strategy, 1e **(Prince)**
Chapter 1 The Roles of Data and Predictive Analytics in Business

1) Which of the following business strategies has a strong business analytics focus?
A) Using data on historical sales to build a recommendation service for your product portfolio.
B) Building a market presence through franchising.
C) Building a reputation for high quality through informative advertising.
D) Ensuring high quality through vertically integrating.

Answer: A
Explanation: Companies routinely collect sales data by product or service at various frequencies, e.g., daily, weekly, monthly, quarterly, and annually. Thus, analysts can use this data to move beyond theoretical arguments or "instincts" and physically measure what products or services in their product portfolio customers like, purchase, and continue to purchase. Using this type of information allows companies to understand their customers' buying preferences and then target appropriate audiences with appropriate marketing campaigns thus increasing overall sales. An example of this is Amazon's "customers who purchased this item also purchased these other items…" The other answer responses require more complex business analytical measurements, and typically appropriate data are not readily available.
Difficulty: 2 Medium
Topic: Defining Data and Data Uses in Business
Learning Objective: 01-01 Explain how predictive analytics can help in business strategy formulation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

2) An appealing alternative to using "rules of thumb" and "gut feelings" in formulating business strategies includes:
A) not deviating from the firm's previously established practices.
B) being an early adopter of business strategies of industry leaders.
C) business strategies justified from evidence-based business analytics.
D) None of these choices are correct.

Answer: C
Explanation: Business strategies supported by empirical based evidence, i.e., data, become more compelling and convincing to key business decision makers.
Difficulty: 1 Easy
Topic: Defining Data and Data Uses in Business
Learning Objective: 01-01 Explain how predictive analytics can help in business strategy formulation.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

3) A plan of action designed by a business practitioner to achieve a business objective would be best described as:
A) business analytics.
B) predictive analytics.
C) business strategy.
D) a database.

Answer: C
Explanation: Business and predictive analytics inform business strategies. They provide evidenced-based measurement that identifies appropriate business strategies that business practitioners can use to achieve specific business outcomes such as increased sales. The resulting business plan of action details how to execute the business strategy(s) to achieve that business objective.
Difficulty: 1 Easy
Topic: Defining Data and Data Uses in Business
Learning Objective: 01-01 Explain how predictive analytics can help in business strategy formulation.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

4) A necessary foundational element for a business to organize the collection of data that the firm will use to analyze is:
A) the HR department.
B) cloud-based storage.
C) the IT department.
D) a database.

Answer: D
Explanation: Business analysts build digital databases based on the relational model of the data using a defined table(s) of information. Analysts can then access or reassemble the information in many ways without having to reorganize the database tables.
Difficulty: 1 Easy
Topic: Defining Data and Data Uses in Business
Learning Objective: 01-01 Explain how predictive analytics can help in business strategy formulation.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

5) Which of the following statements best summarizes the use of *business analytics*?
A) The use of data analysis to aid in business decision making.
B) Using of cluster analysis to create customer categories.
C) Forecasting stock prices.
D) Using "big data" methods to predict sales.

Answer: A
Explanation: The use of data analysis to aid in business decision making is the underlying argument for the use of business analytics, i.e., empirical-based measurement that provides powerful and convincing evidence for use by key business decision makers. The other answer options are approaches to data analysis used to conduct business analytics.
Difficulty: 1 Easy
Topic: Defining Data and Data Uses in Business
Learning Objective: 01-01 Explain how predictive analytics can help in business strategy formulation.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

6) Suppose you collect a series of pictures from Facebook and wish to perform analysis on them. The collection of pictures you've gathered is an example of what type of data?
A) Unstructured data
B) Incomplete data
C) Structured data
D) Censored data

Answer: A
Explanation: A series of pictures do not provide well-defined units of observation of which corresponding information is identifiable. Thus, one cannot classify the data as incomplete, structured, or censored, given the lack of a well-defined unit of observation.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-02 Distinguish structured from unstructured data.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

7) Suppose you collect a series of texts from CEOs' public speaking engagements and wish to perform analysis on them. The collection of texts you've gathered is an example of what type of data?
A) Unstructured data
B) Incomplete data
C) Structured data
D) Censored data

Answer: A
Explanation: A series of text does not provide well-defined units of observation of which corresponding information is identifiable. Thus, one cannot classify the data as incomplete, structured, or censored, given the lack of a well-defined unit of observation.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-02 Distinguish structured from unstructured data.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

8) Suppose you collect monthly sales figures for each store location in your company as well as the wages paid to employees for each store, and that you have a complete history for the last ten years for each store. Provided this information is in table format, the information you've gathered is an example of what type of data?
A) Unstructured data
B) Incomplete data
C) Structured data
D) Censored data

Answer: C
Explanation: One can define the unit of observation for this data as monthly sales figures and employee wages by store location. Store location is the defined unit of observation; thus, the data are structured.
Difficulty: 1 Easy
Topic: Data Features
Learning Objective: 01-02 Distinguish structured from unstructured data.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

9) Suppose you collect a series of texts from statements made of a sample of potential customers as they watch an advertisement for your company's product. The collection of texts you've gathered from these statements is an example of what type of data?
A) Structured data
B) Incomplete data
C) Unstructured data
D) Panel data

Answer: C
Explanation: A series of text does not provide well-defined units of observation of which corresponding information is identifiable. Thus, one cannot classify the data as structured, incomplete, or panel data, given the lack of a well-defined unit of observation.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-02 Distinguish structured from unstructured data.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

10) Use the following dataset to answer the following question.

Name	Month	Commission	Tenure	Travel Expense
Sophie Castro	1	$48,000	4 years	$11,000
Travis Turner	1	$53,000	7 years	$21,000
Elias Hansen	2	$67,000	10 years	$7,000
Amanda Garza	2	$72,000	8 years	$14,000

What is the unit of observation?
A) Month-tenure
B) Person-month
C) Month
D) Tenure

Answer: B
Explanation: The dataset contains commission, tenure, and travel expenses across names (cross-section of people), by month (over time). Thus, the unit of observation is person by month or person-month.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

11) Use the following dataset to answer the following question.

Name	Month	Commission	Tenure	Travel Expense
Sophie Castro	1	$48,000	4 years	$11,000
Travis Turner	1	$53,000	7 years	$21,000
Elias Hansen	2	$67,000	10 years	$7,000
Amanda Garza	2	$72,000	8 years	$14,000

What type of dataset is this?
A) Cross-sectional
B) Time series
C) Pooled cross-section
D) Panel

Answer: C
Explanation: The dataset collects commission, tenure, and travel expense for various company people (cross-section of people by name) by month (over time). Therefore, it is a pooled cross-section dataset. It is not a panel dataset because month 2 collects data on different people than month 1.
Difficulty: 1 Easy
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

12) Use the following dataset to answer the following question.

Firm	Year	# of Employees	Sales	Profits
Jim's Auto	2014	23	$741,000	$52,000
Mechanic Zone	2014	15	$510,000	-$72,000
Jim's Auto	2015	31	$1,081,050	$101,000
Mechanic Zone	2015	27	$811,000	$28,000

What is the unit of observation?
A) Firm
B) Year
C) Year-Profits
D) Firm-Year

Answer: D
Explanation: The dataset contains data on # of employees, sales, and profits across firms (cross-section of firms) by year (over time). Thus, the unit of observation is Firm-Year.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

13) Use the following dataset to answer the following question.

Firm	Year	# of Employees	Sales	Profits
Jim's Auto	2014	23	$741,000	$52,000
Mechanic Zone	2014	15	$510,000	-$72,000
Jim's Auto	2015	31	$1,081,050	$101,000
Mechanic Zone	2015	27	$811,000	$28,000

What type of dataset is this?
A) Cross-sectional
B) Time-series
C) Pooled cross-section
D) Panel

Answer: D
Explanation: The dataset contains data on # of employees, sales, and profits across firms (cross-section of firms) by year (over time). It contains data on the same set of firms for both 2014 and 2015. Thus, it is a panel dataset.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

14) Use the following dataset to answer the following question.

State	Year	Sales Tax	Sales per Store	Employees per Store
Indiana	2014	7.25	$650,000	50
Illinois	2014	7.75	$625,000	65
Indiana	2015	8.00	$800,050	75
Illinois	2015	8.75	$525,000	55

What is the unit of observation?
A) Panel
B) State
C) Sales-employee per store
D) State-year

Answer: D
Explanation: The dataset contains data on sales tax, sales per store, and employees per store across states (cross-section of states) by year (over time). Thus, the unit of observation is State-Year.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

15) Use the following dataset to answer the following question.

State	Year	Sales Tax	Sales per Store	Employees per Store
Indiana	2014	7.25	$650,000	50
Illinois	2014	7.75	$625,000	65
Indiana	2015	8.00	$800,050	75
Illinois	2015	8.75	$525,000	55

What type of dataset is this?
A) Pooled cross-section
B) Panel
C) Time Series
D) State-year

Answer: B
Explanation: The dataset contains data on sales tax, sales per store, and employees per store across states (cross-section of states) by year (over time). It contains data on the same set of states for both 2014 and 2015. Therefore, it is a panel dataset.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

16) Use the following dataset to answer the following question.

State	Year	Salary	Location	Sales Workshop?
Aiden Rosen	2017	$98,000	Seattle	Y
Sadie Hubbard	2017	$83,000	Chicago	N
Jen Jordan	2018	$77,000	Seattle	N
Rob Jackson	2018	$79,000	Chicago	N

What is the unit of observation?
A) Person-salary
B) Person-year
C) Person-year-location
D) Salary

Answer: B
Explanation: The dataset contains data on salary, location, and sales workshop by name (cross-section of people) and year (over time). Therefore, the unit of observation is Person-Year.
Difficulty: 3 Hard
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

17) Use the following dataset to answer the following question.

State	Year	Salary	Location	Sales Workshop?
Aiden Rosen	2017	$98,000	Seattle	Y
Sadie Hubbard	2017	$83,000	Chicago	N
Jen Jordan	2018	$77,000	Seattle	N
Rob Jackson	2018	$79,000	Chicago	N

What type of dataset is this?
A) Panel
B) Time series
C) Pooled cross-section
D) Cross-section

Answer: C
Explanation: The dataset contains salary, location, and sales workshop attendance across people (cross-section of people) by year (over time). Therefore, it is a pooled cross-section dataset. It is not a panel dataset because 2018 collects data on different people than year 2017.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

18) Use the following dataset to answer the following question.

Stock	Day	Stock Price
Apple	5-13-18	$189
Apple	5-14-18	$187
Apple	5-15-18	$185
Apple	5-16-18	$188

What is the unit of observation?
A) Apple-stock price
B) Day-price
C) Day
D) Stock price

Answer: C
Explanation: The dataset contains stock price by day (over time) for Apple. Thus, it does not have a cross-sectional aspect to it. Therefore, the unit of observation is Day.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

19) Use the following dataset to answer the following question.

Stock	Day	Stock Price
Apple	5-13-18	$189
Apple	5-14-18	$187
Apple	5-15-18	$185
Apple	5-16-18	$188

What type of dataset is this?
A) Panel
B) Unstructured
C) Pooled cross-section
D) Time series

Answer: D
Explanation: The dataset contains stock price by day (over time) for Apple (no cross-section, only one firm). Therefore, the dataset contains a well-defined structure given its unit of observation (over time). Thus, it is time series data.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

20) Use the following dataset to answer the following question.

Name	Year	Salary	Tenure	MBA?
Dmitry Haas	2017	$98,000	5 years	Y
Jackie Bay	2017	$83,000	3 years	N
Renee Topkis	2017	$77,000	12 years	Y
Maria Val	2017	$79,000	8 years	N

What is the unit of observation?
A) Person-salary
B) Person
C) Person-Tenure
D) Salary

Answer: B
Explanation: The dataset contains data on salary, tenure, and whether an employee has an MBA across people (cross-section of people) for 2017 only (point in time). Thus, the unit of observation is Person.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

21) Use the following dataset to answer the following question.

Name	Year	Salary	Tenure	MBA?
Dmitry Haas	2017	$98,000	5 years	Y
Jackie Bay	2017	$83,000	3 years	N
Renee Topkis	2017	$77,000	12 years	Y
Maria Val	2017	$79,000	8 years	N

What is the type of data set?
A) Pooled cross-section
B) Panel
C) Unstructured
D) Cross-section

Answer: D
Explanation: The dataset contains data on salary, tenure, and whether an employee has an MBA across people (cross-section of people) for 2017 only (point in time). Therefore, it is cross-section.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

22) Suppose you collect monthly sales figures for each store location in your company as well as the wages paid to employees for each store, and that you have a complete history for the last ten years for each store. What is the type of this data set?
A) Pooled cross-section
B) Unstructured data
C) Censored data
D) Panel

Answer: D
Explanation: This dataset contains one company's data on sales and employee wages across store location (cross-section by store location) by month for ten years (over time). Implied in the question is that data collected involves the same store locations each month and year (i.e., the company did not close or open stores). Therefore, this is a panel dataset. In addition, the data have a well-defined unit of observation, which is store location by month and year; thus, it is a structured dataset.
Difficulty: 1 Easy
Topic: Data Features
Learning Objective: 01-03 Differentiate units of observation.
Bloom's: Understand
AACSB: Reflective Thinking

Accessibility: Keyboard Navigation

23) If commercial real estate costs of Kroger affect demand for Kroger Milk through the product prices Kroger charges to its customers, this is an example of what type of relationship?
A) A direct causal effect of real estate costs on product demand
B) An indirect causal effect of product prices on product demand
C) An indirect causal effect of product demand on product prices
D) An indirect causal effect of real estate costs on product demand

Answer: D
Explanation: A change in Kroger's real estate costs causes a change in Kroger's milk price, which then causes a change Kroger milk demand. Therefore, one can say that a change in Kroger's real estate costs indirectly causes a change in Kroger milk demand. This is an indirect causal relationship.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-04 Outline a data-generating process.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

24) If advertising exposure, disposable income and price affect demand for a product, and advertising exposure affects the price of the product, what type of relationship is ***not*** present?
A) A direct causal effect of advertising exposure on product demand
B) A direct causal effect of advertising exposure on product prices
C) An indirect causal effect of disposable income on product demand
D) A direct causal effect of product prices on product demand

Answer: C
Explanation: The only indirect causal relationship defined in the question is that a change in advertising exposure causes a change in product price, which then causes a change in product demand. Thus, advertising exposure indirectly causes a change in product demand. The question states that disposable income directly affects product demand; therefore, a change in disposable income causes a change in product demand, which is a direct causal relationship.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-04 Outline a data-generating process.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

25) In describing the data-generating process for "click-throughs" for one of your firm's advertising campaigns you've assumed the following relationship: $Y_i = f(H_i, J_i) + U_i$, where Y_i is if individual i clicked through, H_i is individual i's household income, J_i is an indicator for whether or not individual i has a job. What type of factors might be contained in U_i?
A) A factor that affects J_i but does not affect Y_i.
B) A factor that affects H_i but does not affect Y_i.
C) A factor that affects H_i and Y_i.
D) None of the answers is correct.

Answer: C
Explanation: To build a formal model of the data generating process of what causes people to both view and add and click on it, one asserts that click-throughs are a function of household income and whether an individual in that household works. However, there are other factors that may affect click-throughs either directly or indirectly through household income, and U_i will contain those factors. For example, a household member completed a college degree and subsequently received a promotion at work, thus raising household income. Therefore, a change in education caused a change in household income, which then caused a change in click-throughs. Thus, a change in education indirectly caused a change in click-throughs.
Difficulty: 2 Medium
Topic: Data Features
Learning Objective: 01-04 Outline a data-generating process.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

26) Suppose you've assumed the following two data-generating processes: (1) $Y_i = f(H_i, J_i)$ and (2) $J_i = g(X_i, Z_i)$. What do these assumptions imply?
A) J has a direct causal effect on H.
B) Z has a direct causal effect on Y.
C) X has an indirect causal effect on J.
D) Z has an indirect causal effect on Y.

Answer: D
Explanation: A change in Z causes a change in J, which then causes a change in Y. Thus, a change in Z indirectly causes a change in Y. Therefore, Z has an indirect causal effect on Y.
Difficulty: 2 Medium
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

27) Providing a *graphical presentation of past trends* of a set of critical indicators for a company is an example of what data presentation format?
A) Query
B) Report
C) KPIs (key performance indicators)
D) Dashboard

Answer: D
Explanation: A dashboard is a graphical presentation of the current standing and historical trends for a company's variables of interest, typically KPIs. Analysts design dashboards to give some indication of the company's performance at a point in time or over time. Query outcomes may result in inputs to dashboards as well. Reports are typically more formal, structured presentations of information contained in datasets.
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

28) Presenting an assessment of variables of interest against a given benchmark for a company is an example of what data presentation format?
A) Query
B) Scorecard
C) KPIs (key performance indicators)
D) Dashboard

Answer: B
Explanation: A scorecard is any structured assessment of variables of interests, typically KPIs, against a given benchmark. Scorecards "score" how a firm is performing along various dimensions.
Difficulty: 2 Medium
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

29) Identifying distinctive relationships between observations in a dataset is an example of what sort of data analysis?
A) Query
B) Pattern discovery
C) Mean
D) Covariance

Answer: B
Explanation: The pattern in a dataset is any distinctive relationship between variables in the dataset. Pattern discovery is the process of identifying those distinctive relationships between observations in the data. Thus, the process of pattern identification is the discovery aspect of this type of analysis using various analytical methods such as data mining, linear regression, association analysis, cluster analysis, and outlier detection.
Difficulty: 2 Medium
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

30) In examining the career earnings of different undergraduate majors, identifying Michael Jordan as an atypical instance amongst Geography majors would be an example of what sort of data analysis?
A) Dashboard
B) Pattern discovery
C) Outlier detection
D) Data mining

Answer: C
Explanation: Outlier detection: Small subsets of observations that contain information far different from the vast majority of observations in the dataset. Michael Jordan's career earnings are far different than most college undergraduate majors, and his career earnings meet the criteria of outlier amongst different undergraduate majors.
Difficulty: 2 Medium
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

31) The process of identifying distinctive relationships between observations in a data set, or pattern discovery, within a very large dataset is typically known as?
A) Data mining
B) Outlier detection
C) Query
D) Pattern discovery

Answer: A
Explanation: Data mining is a subset of pattern discovery, and it typically involves pattern discovery in large datasets. Outlier detection is pattern discovery that looks at small subsets of observations. Queries are not part of pattern discovery data analytics because they focus on information retrieval and summary.
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

32) In grouping customers into separate types according to their spending, internet browsing, payment method, and age a firm is engaging in what sort of analysis?
A) Data mining
B) Cluster analysis
C) Outlier detection
D) Linear regression

Answer: B
Explanation: Grouping customers into categories having similarities is cluster analysis. Observations in the same group are more similar than observations across groups. Pattern discovery for cluster analysis usually involves setting thresholds, e.g., Euclidean distance, to define a cluster or category.
Difficulty: 2 Medium
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

33) In examining the career earnings of individuals with different levels of college/graduate attainment, identifying Bill Gates as an atypical instance amongst the population with less than a bachelor's degree, would be an example of what sort of data analysis?
A) Query
B) Pattern discovery
C) Data mining
D) Outlier detection

Answer: D
Explanation: Outlier detection: Small subsets of observations that contain information far different from the vast majority of observations in the dataset. Bill Gates's career earnings are far different than most of the population with less than a bachelor's degree; thus, his career earnings meet the criteria of outlier amongst the population with less than a bachelor's degree.
Difficulty: 2 Medium
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

34) Which of the following would ***not*** be an example of summary statistic of a variable within dataset?
A) Mean
B) Variance
C) Interquartile range
D) None of the above.

Answer: D
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

35) Any request for information from a database is an example of:
A) pattern discovery.
B) data mining.
C) query.
D) association analysis.

Answer: C
Explanation: Queries are any request for information from a database to retrieve information and provide summaries. Data mining and association analysis are examples of pattern discovery, which is the process of identifying distinctive relationships between observations in a dataset. Some will argue that pattern discovery is no different than queries. The difference focuses on the active word distinctive. Pattern discovery imposes criteria or rules that require patterns to meet a threshold that determines or defines the pattern, which distinguishes them from simple information retrieval or summary. Queries do not.
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

36) Quantitative measures meant to summarize and interpret properties of a dataset are instances of:
A) pivot tables.
B) descriptive statistics.
C) linear regressions.
D) outliers.

Answer: B
Explanation: Analysts define descriptive statistics broadly as quantitative measures that summarize and interpret properties of a dataset. Descriptive statistics are standard queries for data analysis. Pivot tables are data summarization tools that provide different views of a given dataset, while linear regressions and outliers are examples of pattern discovery.
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

37) Use the following dataset to answer the following question.

State	Year	Sales	Capital Expenditures	Wages
Western	2016	$750,000	$130,000	$330,000
Eastern	2016	$645,000	$225,000	$430,000
Western	2017	$770,050	$145,000	$230,000
Eastern	2017	$925,000	$125,000	$530,000

Using a spreadsheet software, such as Excel, to preview different views of this dataset, such as Sales in the Eastern part of the country, would be aided by what?
A) Linear regression
B) Cluster analysis
C) Pivot table
D) Scorecard

Answer: C
Explanation: Pivot tables are data summarization tools that provide different views of a given dataset, and they are an example of a standard query. Linear regression and cluster analysis are examples of pattern discovery. Scorecard is any structured assessment of variables of interest, typically KPIs, against a given benchmark.
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

38) Use the following presentation to answer the following question.

State	Year	Sales per Store	Target	Performance
Western	2016	$750,000	$700,000	Good
Eastern	2016	$645,000	$680,000	Nearing Acceptable
Western	2017	$770,050	$745,000	Good
Eastern	2017	$925,000	$825,000	Good

This presentation is an example of what?
A) Cluster analysis
B) Association analysis
C) Pivot table
D) Scorecard

Answer: D
Explanation: Scorecard is any structured assessment of variables of interest, typically KPIs, against a given benchmark. For the years 2016 and 2017, an analyst is assessing how well eastern and western states' sales per store are doing relative to meeting or exceeding their assigned target sales.
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

39) Measures of the central tendencies of variables such as the mean, median, and mode are examples of:
A) descriptive statistics.
B) data mining.
C) covariance.
D) causal relationships

Answer: A
Explanation: Analysts define descriptive statistics broadly as quantitative measures that summarize and interpret properties of a dataset. They typically consist of measures that assess the center and spread of variables in a dataset. Descriptive statistics are standard queries for data analysis.
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

40) In grouping TV markets into separate types according to regional spending, demographics, and business activity, and cost of advertising is engaging in what sort of analysis?
A) Passive prediction
B) Cluster analysis
C) Data mining
D) Linear regression

Answer: B
Explanation: Grouping TV markets into separate market types or categories based on similarities, e.g., regional spending, demographics and business activity, and cost of advertising, is cluster analysis. Observations in the same group are more similar than observations across groups.
Difficulty: 2 Medium
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

41) Measures of the spread of variables such as the variance and range are examples of:
A) descriptive statistics.
B) data mining.
C) covariance.
D) causal relationships.

Answer: A
Explanation: Analysts define descriptive statistics broadly as quantitative measures that summarize and interpret properties of a dataset. They typically consist of measures that assess the center and spread of variables in a dataset. Descriptive statistics are standard queries for data analysis.
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

42) An analyst is attempting to understand whether the frequency of faulty units coming out of production last month is different for Eastern location plants relative to all plants across the entire company. This analyst is engaging in what type of data analysis?
A) Linear regression
B) Outlier detection
C) Cluster analysis
D) Association analysis

Answer: D
Explanation: Association analysis is associated with attempts to discover dependencies, generally in the form of conditional probabilities, between two or more variables in the data. In this case, the hypothesis is that higher frequency fault rates are associated with or conditioned on plants being in the Eastern part of the country relative to the frequency fault rates of all company plants.
Difficulty: 1 Easy
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

43) Gauging customer discontent for an airline by looking for patterns between unfavorable mentions on Twitter (a potentially large dataset) and on-time traffic departures would likely be an example of what type of data analysis?
A) Linear regression
B) Data mining
C) Outlier detection
D) Scorecard

Answer: B
Explanation: Data mining is an example of pattern discovery, usually in large datasets. In this question, an analyst is assessing whether there is a pattern between on-time departures and unfavorable mentions on Twitter for a given airline. If so, then the argument (which requires further assessment to prove) is that an airline experiencing a substantial level of unfavorable mentions on Twitter is also missing its assigned departure time frequently.
Difficulty: 2 Medium
Topic: Basic Uses of Data Analysis for Business
Learning Objective: 01-05 Describe the primary ways that data analysis is used to aid business performance.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

44) Which of the following is an example of lead information?
A) Measurement of sales in response to a recently implemented ad campaign
B) Measurement of price sensitivity of customers in a given city
C) Measurement of rate of customer complaints following an employee training program
D) None of the answers is correct.

Answer: B
Explanation: Lead information is information that provides insight about the future. If we know that customers are price sensitive to a company's product in a given city, then we know that in the future if a company raises the price of that product in that city, we will on average see a decline in quantity demanded (purchases) in that city for their product. This assumes that nothing else that affects customers' purchases of the company's product in that city has changed. The other answer options are examples of lag information, which is information about past outcomes and usually used to measure firm performance.
Difficulty: 1 Easy
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-06 Discriminate between lead and lag information.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

45) Which of the following is an example of lead information?
A) Measurement of number of employees enrolled in new HR workshop
B) Measurement of employee complaints following an employee training program
C) Measurement of wage sensitivity of employee retention in a given department
D) Measurement of number of employees hired since opening of new store location

Answer: C
Explanation: Lead information is information that provides insight about the future. Wage sensitivity in a given firm department provides information to HR and department managers regarding increasing or decreasing wages and whether employees of that department will be more likely to stay or leave given the wage increase or decrease. The other answer options are examples of lag information, which is information about past outcomes.
Difficulty: 1 Easy
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-06 Discriminate between lead and lag information.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

46) Determining the average propensity of customers to purchase a warranty plan, if the plan is under a promotional offer during the time of product purchase, would be an example of what type of information?
A) Passive prediction
B) Lag information
C) Indirect causal effect
D) Lead information

Answer: D
Explanation: Lead information is information that provides insight about the future. Knowing that customers are more likely to buy a warranty plan discounted at the time of purchase provides information that will help firms to increase future sales. As an example, firms should discount lawn mower warranty plans in the spring when lawn mowers are selling at higher rates due to higher lawn mowing requirements, if customers' average propensity to purchase a warranty plan is higher when discounted. Customers buying lawn mowers at that time would be more likely to buy the warranty, thus increasing total sales to the firm.
Difficulty: 2 Medium
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-06 Discriminate between lead and lag information.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

47) Which of the following is an example of lag information?
A) Measurement of faulty units in response to a recently implemented inventory method
B) Measurement of price sensitivity of customers in a given city
C) Measurement of rate of purchasing propensity following exposure to advertising
D) None of the answers is correct.

Answer: A
Explanation: Analysts develop lag information by using information from variables usually classified as KPIs to answer the following question, "What happened?" Measurement of faulty units in response to recently implemented inventory methods provides information regarding whether the number of faulty units has declined after implementation of an inventory method. The other answer options are examples of lead information.
Difficulty: 1 Easy
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-06 Discriminate between lead and lag information.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

48) For a health care management company, indicators such as bed utilization, overtime wages, and patients served would represent what sort of information?
A) Unstructured
B) Indirect causal relationship
C) Key performance indicators
D) None of the answers is correct.

Answer: C
Explanation: For this type of company, health care management, bed utilization, overtime wages, and patients served are KPIs (key performance indicators). They are variables that help measure firm performance in near-real time.
Difficulty: 1 Easy
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-06 Discriminate between lead and lag information.
Bloom's: Remember
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

49) Suppose you want to answer the following question: "How will Sales of version 2.0 of our product change when we release version 3.0 of our product?" What type of data analysis is crucial toward answering this question?
A) Active prediction
B) Passive prediction
C) Data mining
D) OLAP cube

Answer: A
Explanation: Active prediction is the use of predictive analytics to make predictions based on actual and/or hypothetical data for which one or more variables experience an exogenous alteration. The current data-generating process for sales of Version 2.0 does not depend on competition from Version 3.0. However, the business decision to introduce Version 3.0 is an exogenous event that may impact the demand, and therefore the sales, of Version 2.0. The question of interest is will (and how) the introduction of Version 3.0 impact the sales of Version 2.0, all other things affecting the demand for Version 2.0 held constant. To predict the consequences of introducing Version 3.0 on sales of Version 2.0 requires an active prediction method based on causal inference modeling.
Difficulty: 2 Medium
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-07 Discriminate between active and passive prediction.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

50) Suppose you want to answer the following question: "How will Sales of our product at our outdoor location change if it rains above average next week?" What type of data analysis is crucial toward answering this question?
A) Active prediction
B) Passive prediction
C) Dashboard
D) Pivot table

Answer: B
Explanation: Passive prediction is the use of analytics to make predictions based on actual and/or hypothetical data for which no variables are exogenously altered. Having a week of rainfall above average is not an exogenous event. No one is actively changing how much rainfall there is; rather we passively observe it without intervention.
Difficulty: 2 Medium
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-07 Discriminate between active and passive prediction.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

51) Using employment growth in the leisure and hospitality sector for the entire city of Indianapolis to predict yearly sales for your restaurant (in Indianapolis) is an example of:
A) passive prediction.
B) active prediction.
C) outlier detection.
D) scorecard.

Answer: A
Explanation: Passive prediction is the use of analytics to make predictions based on actual and/or hypothetical data for which no variables are exogenously altered. You are not actively changing employment growth in the leisure and hospitality sector in Indianapolis; rather you passively observe it without intervention.
Difficulty: 2 Medium
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-07 Discriminate between active and passive prediction.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

52) Predicting how moving to a just-in-time production plan will affect employee productivity will require what sort of prediction?
A) Passive prediction
B) Active prediction
C) KPI
D) Lag information

Answer: B
Explanation: Active prediction is the use of predictive analytics to make predictions based on actual and/or hypothetical data for which one or more variables experience an exogenous alteration. The introduction of a just-in-time production plan is an exogenous change to predicting (causally) employee productivity. The question of interest is will (and how) the implementation of a just-in-time production plan impact employee productivity. Active prediction based on causal modeling will answer this question.
Difficulty: 2 Medium
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-07 Discriminate between active and passive prediction.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

53) The critical distinction between the effectiveness of engaging in active prediction versus passive prediction is that one or more of the variables of interest undergoes a(n):
A) lead information.
B) trend.
C) outlier detection.
D) exogenous change.

Answer: D
Explanation: Active prediction is the use of predictive analytics to make predictions based on actual and/or hypothetical data for which one or more variables experience an exogenous alteration. Passive prediction is the use of analytics to make predictions based on actual and/or hypothetical data for which no variables are exogenously altered. The critical distinction is whether an exogenous alteration occurred to the data-generating process.
Difficulty: 2 Medium
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-07 Discriminate between active and passive prediction.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

54) Common criteria to judge competing models to be used in passive prediction include:
A) degree of exogenous variation.
B) model fit.
C) data mining.
D) pattern discovery.

Answer: B
Explanation: Passive prediction implies no degree of exogenous variation in the data-generating process. Model fit is the determination of the model that most closely fits the data according to some metrics. Analysts conduct passive prediction using a wide range of predictive analytics models, e.g., neural networks, data mining, and regression, designed for pattern discovery. Analysts then use model fit as the basis for selecting which model to use.
Difficulty: 2 Medium
Topic: Data Analysis for the Past, Present, and Future
Learning Objective: 01-07 Discriminate between active and passive prediction.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation

55) Which of the following business questions require active prediction?
A) How will employee retention change after introducing a new-hire workshop?
B) Are older or younger employees more likely to stay with the firm?
C) Which department within the company has the highest turnover rate?
D) Do high performance employees have higher tenures with the firm?

Answer: A
Explanation: Active prediction is the use of predictive analytics to make predictions based on actual and/or hypothetical data for which one or more variables experience an exogenous alteration. Introducing a new hire workshop is an exogenous change to the employee retention change data-generating process. The other answer options are examples of passive prediction.
Difficulty: 2 Medium
Topic: Active Prediction for Business Strategy Formation
Learning Objective: 01-08 Recognize questions pertaining to business strategy that may utilize (active) predictive analytics.
Bloom's: Understand
AACSB: Reflective Thinking
Accessibility: Keyboard Navigation