Glossary

Data Analytics

Terms and Definitions

A

Action-oriented question: A type of question that generates answers that can lead to action or change (Example: What design features will make our packaging easier to recycle?)

Algorithm: A process or set of rules to be followed for a specific task

Analytical skills: Qualities and characteristics associated with solving problems using facts

Analytical thinking: Identifying and defining a problem and then solving it by using data in an organized, step-by-step manner

Attribute: A characteristic or quality of data used to label a column in a table

AVERAGE (function in spreadsheets): Returns an average of the values from a selected range of cells

B

Bed occupancy rate: Total # of inpatient days for a given period x 100 / (available beds x # of days in the period)

Big data: Focuses on larger, less specific data, during a longer period of time, and helps companies make big decisions

Borders (in spreadsheets): Borders are the lines formed by the boxes around each cell which can be made visible so each piece of data is clearer, or hidden

Business task: The question or problem data analysis answers for a business

C

Cell reference: A cell or a range of cells in a worksheet that can be used in a formula and can be automatically updated when the formula is copied to a new cell

Cloud: A place to keep data online, rather than a computer hard drive

Context: The condition in which something exists or happens

COUNT (function in spreadsheets): Returns the number of cells in a range that contain values

D

Dashboard: Monitors live, incoming data

Data: A collection of facts

Data analysis: The collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making

Data analysis process: Carrying out the six phases of ask, prepare, process, analyze, share, and act in order to gain data-driven insights that lead to informed decisions

Data analyst: Someone who collects, transforms, and organizes data in order to help make informed decisions

Data analytics: The science of data

Database: A collection of data stored in a computer system

Data design: How you organize information

Data-driven decision making: Using facts to guide business strategy

Data ecosystems: The various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data

Data-inspired decision-making: Explores different data sources to find out what they have in common

Data life cycle: The sequence of stages that data experiences, which include plan, capture, manage, analyze, archive, and destroy

Data science: Creating new ways of modeling and understanding the unknown by using raw data

Data strategy: The management of the people, processes, and tools used in data analysis

E

Equations: Calculations that involve addition, subtraction, multiplication, or division, such as 3 -1 = 2, 15 + 8 ÷ 2 = 19, or 846 x 513 = 433,998 (refer also to **Math Expressions**)

F

Fairness: Ensuring that your analysis doesn't create or reinforce bias

Fill handle (in spreadsheets): A small box that appears in the lower right corner of a selected cell that, when dragged to other cells in the same column or row, fills those cells with the formula or function in the selected cell

Filtering (in spreadsheets): Focuses on and displays only the data that matches the variables selected

Formula: A set of instructions that performs a specific calculation using the data in a spreadsheet

Function: A preset command that automatically performs a specific process or task using the data in a spreadsheet

G

Gap analysis: A method for examining and evaluating how a process works currently in order to get where you want to be in the future

H

Headers (in spreadsheets): Typically the first row in a spreadsheet that labels the type of data in each column (refer also to **Variables**)

L

Leading question: A question that encourages people to respond in a certain way (Example: These are the best sandwiches ever, aren't they?)

M

Math expressions: Calculations that involve addition, subtraction, multiplication, or division, such as 3 - 1 = 2, 15 + 8 ÷ 2 = 19, or 846 x 513 = 433,998 (refe r also to **Equations**)

Math functions (in spreadsheets): Functions that are used as part of a math formula (Examples: SUM, AVERAGE, COUNT, MIN, and MAX)

MAX (function in spreadsheets)**:** Returns the largest numeric value from a range of cells

Measurable question: A question that generates answers that can be quanti fied and assessed (Example: How many times was our video shared on social media the first week it was posted?)

Metric: Single, quantifiable type of data that can be used for measurement

Metric goal: A measurable goal set by a company and evaluated using metrics

MIN (function in spreadsheets): Returns the smallest numeric value from a range of cells

O

Observation: All of the attributes for something contained in a row of a data table

Open data: Data that is made available to the public

Operator: A symbol that names the type of operation or calculation to be performed

Order of operations (in spreadsheets): Values in cells in a spreadsheet are grouped together using parentheses to make clear the order in which operations are to be performed Example: (B2+C2+D2+E2)/4 shows all four values are added together before dividing by 4.

P

Pivot chart: A chart created from fields selected from a pivot table

Pivot table: A data summarization tool that is used in data processing. Pivot tables are used to summarize, sort, reorganize, group, count, total or average data stored in a database

Problem domain: The specific area of analysis that encompasses every activity affecting or affected by the problem

Problem types: The different types of tasks that data analysts encounter including: making predictions, categorizing things, spotting something unusual, identifying themes, discovering connections, and finding patterns

Q

Qualitative data: Subjective or explanatory measures of qualities and characteristics

Quantitative data: Specific and objective measures of numerical facts

Query: The way we use SQL to communicate with the database

Query language: A computer programming language that allows you to retrieve and manipulate data from a database

R

Range: A collection of two or more cells in a spreadsheet

Relevant question: A question that has significance to a problem (Example: What

environmental factors changed in Durham, North Carolina that could have caused Pine Barrens tree frogs to disappear from the Sandhills Region?)

Report: Static collection of data given to stakeholders periodically

Return on investment (ROI): A formula designed using metrics that lets a business know how well an investment is doing, ROI= profit/investment

Revenue: Revenue= # of sales x sales price

Root cause: The reason why a problem occurs

S

Scope of work (SOW): An agreed-upon outline of the work you're going to perform on a project

Small data: Focuses on specific metrics, during a short time-period, and is useful for day-to-day decisions

SMART methodology: A method for determining a question's effectiveness. An effective question is Specific, Measurable, Action-oriented, Relevant, and Time-bound (SMART)

Sorting (in spreadsheets): Organizes selected data in order, such as from smallest to largest or alphabetical

Specific question: A question that is simple, significant, and focused on a single topic or a few closely related ideas (Example: How many kids achieve the recommended 60 minutes of physical activity at least five days a week?)

Spreadsheet: An electronic document in which data is arranged in the rows and columns of a grid and can be manipulated and used in calculations

SQL: Standard query language

Stakeholders: People that have invested time, interest, and resources into the projects you'll be working on as a data analyst

Structured thinking: The process of recognizing the current problem or situation, organizing available information, revealing gaps and oppo. unities, and identifying the options

SUM (function in spreadsheets): Adds the values from a selected range of cells

T

Technical mindset: The ability to break things down into smaller steps or pieces and work with them in an orderly and logical way

Third-party logistics: The partnership between a shipping company and a business to ship products when the business doesn't have its own trucks, planes, or ships

Time-bound question: A question that specifies the timeframe to be addressed in the answers (Example: What environmental factors changed in Durham, North Carolina, *between 1983 and 2004* that could cause Pine Barrens tree frogs to disappear from the Sandhills Region?)

U

Unfair question: A question that makes assumptions or is difficult to answer honestly (Example: What do you love most about our exhibits?)

V

Variables (in spreadsheets): Typically the first row in a spreadsheet that labels the type of data in each column (refer also to **Headers**)

Visualization: The graphical representation of information