Data Analysis Course 2: Ask Questions to Make Data-Driven Decisions

Course 2, Module 1

Questions to ask in six data analysis phases:

1. Ask
   1. What are my stakeholders saying their problems are?
   2. Now that I've identified the issues, how can I help stakeholders resolve their questions?
2. Prepare
   1. What do I need to figure out how to solve this problem?
   2. What research do I need to do?
3. Process
   1. What data errors or inaccuracies might get in my way of getting the best possible answer to the problem I am trying to solve?
   2. How can I clean my data so the info I have is more consistent?
4. Analyze
   1. What story is my data telling me?
   2. How will my data help me solve this problem?
   3. Who needs my company's product or service? What type of person is most likely to use it?
5. Share
   1. How can I make what I present to the stakeholders engaging and easy to understand?
   2. What would help me understand this if I were the listener?
6. Act
   1. How can I use the feedback I received during the share phase to actually meet the stakeholder's needs and expectations?

Structured thinking: breaking down the data analysis process into smaller, manageable parts. This process involves four basic activities:

1. Recognizing the current problem or situation
2. Organizing available information
3. Revealing gaps and opportunities
4. Identifying your options

Common problem types

1. Making predictions: using data to make informed decisions about how things may be in the future.
2. Categorizing things: assigning information to different groups or clusters based on common features.

3. Spotting something unusual: identify data the is different from the norm.
4. Identifying themes: grouping categorized information into broader concepts.
5. Discovering connections: finding similar challenges faced by different entities and combining data and insights to address them.
6. Finding patterns: using historical data to understand what happened in the past and is therefore likely to happen again.

SMART Questions
Bad questions:
– Leading questions
– Closed-ended questions
– Questions that are too vague and lack context

– Specific: simple, significant, focused on a single topic or a few closely related ideas.
– Measurable: can be quantified and assessed.
– Action oriented: encourage change.
– Relevant: matter, are important, and have significance to the problem you are trying to solve.
– Time bound: specify the time to be studied.

Terms and definitions for Course 2, Module 1
Action-oriented question: A question whose answers lead to change

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

Data analysis process: The six phases of ask, prepare, process, analyze, share, and act whose purpose is to gain insights that drive informed decision-making

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

Leading question: A question that steers people toward a certain response

Measurable question: A question whose answers can be quantified and assessed

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

Relevant question: A question that has significance to the problem to be solved

SMART methodology: A tool for determining a question's effectiveness based on whether it is specific, measurable, action-oriented, relevant, and time-bound

Specific question: A question that is simple, significant, and focused on a single topic or a few closely related ideas

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

Time-bound question: A question that specifies a timeframe to be studied

Unfair question: A question that makes assumptions or is difficult to answer honestly

Course 2, Module 2

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

Potential dangers of relying entirely on data-driven decision-making can include overreliance on historical data, a tendency to ignore qualitative insights, and potential biases in data collection and analysis

- Data-inspired decision-making can avoid some of the pitfalls that data-driven decisions might be prone to.

Quantitative data: specific and objective measures of numerical facts.
Qualitative data: subjective or explanatory measures of qualities and characteristics.

- Usually, qualitative data can help analysts better understand their quantitative data by providing a reason or more thorough explanation. In other words, quantitative data generally gives you the what, and qualitative data generally gives you the why.

Report: static collection of data given to stakeholders periodically.
Dashboards: monitors live, incoming data.

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

Data vs. Metrics
Metric: a single, quantifiable type of data that can be used for measurement.

Mathematical thinking: looking at a problem and logically breaking it down step-by-step, so you can see the relationship of patterns in your data, and use that to analyze your problem.

Small data: specific, short time period, day-to-day (spreadsheets)

Big data: large and less specific, long time-period, big decisions (SQL)
  – volume: the amount of data.
  – variety: the different kinds of data.
  – velocity: how fast the data can be processed.
  – veracity: the quality and reliability of the data.

Terms and definitions for Course 2, Module 2
Algorithm: A process or set of rules followed for a specific task

Big data: Large, complex datasets typically involving long periods of time, which enable data analysts to address far-reaching business problems

Dashboard: A tool that monitors live, incoming data

Data-inspired decision-making: The process of exploring different data sources to find out what they have in common

Metric: A single, quantifiable type of data that is used for measurement

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

Pivot chart: A chart created from the fields in a pivot table

Pivot table: A data summarization tool used to sort, reorganize, group, count, total, or average data

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

Qualitative data: A subjective and explanatory measure of a quality or characteristic

Quantitative data: A specific and objective measure, such as a number, quantity, or range

Report: A static collection of data periodically given to stakeholders

Return on investment (ROI): A formula that uses the metrics of investment and profit to evaluate the success of an investment

Revenue: The total amount of income generated by the sale of goods or services

Small data: Small, specific data points typically involving a short period of time, which are useful for making day-to-day decisions

Course 2, Module 3
Function: a preset command that automatically performs a specific process or task using the data.

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

Train your brain to think structurally.
Structured thinking: the process of recognizing the current problem or situation, organizing available information, revealing gaps and opportunities, and identifying the options.

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

Scope of work (SOW): an agreed-upon outline of the work you're going to perform on a project.
  – deliverables
  – timelines
  – milestones
  – reports

Descriptive
Diagnostic
Predictive
Prescriptive

Context: the condition and circumstances that surround and give meaning to the data in data analytics.
  – Can turn raw data into meaningful information.
  – Give data perspective by defining it.

Terms and definitions for Course 2, Module 3
AVERAGE: A spreadsheet function that returns an average of the values from a

selected range

Borders: Lines that can be added around two or more cells on a spreadsheet

Cell reference: A cell or a range of cells in a worksheet typically used in formulas and functions

COUNT: A spreadsheet function that counts the number of cells in a range that meet a specific criteria

Equation: A calculation that involves addition, subtraction, multiplication, or division (also called a math expression)

Fill handle: A box in the lower-right-hand corner of a selected spreadsheet cell that can be dragged through neighboring cells in order to continue an instruction

Filtering: The process of showing only the data that meets a specified criteria while hiding the rest

Header: The first row in a spreadsheet that labels the type of data in each column

Math expression: A calculation that involves addition, subtraction, multiplication, or division (also called an equation)

Math function: A function that is used as part of a mathematical formula

MAX: A spreadsheet function that returns the largest numeric value from a range of cells

MIN: A spreadsheet function that returns the smallest numeric value from a range of cells

Open data: Data that is available to the public

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

Order of operations: Using parentheses to group together spreadsheet values in order to clarify the order in which operations should be performed

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

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

Report: A static collection of data periodically given to stakeholders

Return on investment (ROI): A formula that uses the metrics of investment and profit to evaluate the success of an investment

Revenue: The total amount of income generated by the sale of goods or services

Scope of work (SOW): An agreed-upon outline of the tasks to be performed during a project

Sorting: The process of arranging data into a meaningful order to make it easier to understand, analyze, and visualize

SUM: A spreadsheet function that adds the values of a selected range of cells

Course 2, Module 4

Questions to ask at the start of a project:
1. Who are the primary and secondary stakeholders?
2. Who is managing the data?
3. Where can you go for help?

Who is your audience?
What do they know?
What do they need to know?
How can I communicate effectively with them?

Learn as you go and ask questions.

Flag problems as early as possible for stakeholders.
Set realistic timelines and goals.

Understand the why, the context
– Takes time to get to the bottom of the story the data tells.

Limitations of data
– Incomplete or nonexistent data
– Don't miss misaligned data
– Deal with dirty data
– Tell a clear story
– Be the judge

Terms and definitions for Course 2, Module 4
Cloud: A place to keep data online, rather than a computer hard drive

Reframing: Restating a problem or challenge, then redirecting it toward a potential resolution

Turnover rate: The rate at which employees voluntarily leave a company