

# **Google Data Analytics Professional Certificate**

*Course Notes*

JESUS CARDENAZ

# 1 Concepts

The course will focus on effective questioning, problem-solving, types of data, structured thinking, and communication strategies.

- Effective and Ineffective Questions:
  - Effective questions lead to insights and help solve business problems
  - Ineffective questions hinder the data analysis process.
- Data Analysis Phases
  1. Ask: Defining the problem, understanding stakeholder expectations.
  2. Prepare: Gathering and organizing data for analysis.
  3. Process: Cleaning, transforming, and formatting data.
  4. Analyse: Applying statistical and analytical techniques.
  5. Share: Presenting findings and insights to stakeholders.
  6. Act: Taking action based on the analysis.
- Structured Thinking
  - Structured thinking involves recognizing the current problem, organizing information, identifying gaps and opportunities, and determining options.
  - It breaks down complex problems into smaller steps to find logical solutions.
- Data Collection Methods
  - First-party data: Data collected by an individual or group using their own resources.
  - Second-party data: Data collected by a group directly from its audience and then sold.
  - Third-party data: Data collected from outside sources that were not collected directly.
- Data Collection Considerations
  - How the data will be collected
  - Choose data sources
  - Decide what data to use
  - Select the right data type
  - Determine the time frame
- Types of Data
  - Qualitative Data: Descriptive data that cannot be easily expressed using number, such as movie titles and cast members.
  - Quantitative Data: Measurable or countable data expressed as numbers, such as budget and box office revenue.
  - Discrete Data: Counted data with a limited number of values.
  - Continuous Data: Measured data that can be expressed as decimals with multiple decimal places.
  - Nominal Data: Qualitative data categorized without a set order, such as responses to a yes/no question.
  - Ordinal Data: Qualitative data with a set order or scale, such as ranking a movie from 1 to 5.
  - Internal Data: Data that resides within a company's own systems and is collected using their own methods.
  - External Data: Data generated outside of an organization, often collected from various sources.
  - Structured Data: Organized in a specific format, such as spreadsheets or databases, making it easily searchable and analysis-ready.

- Unstructured Data: Data that lacks a clear organization, such as audio and video files, and does not fit into rows and columns.
- Data Modeling
  - Data Models are visual representations that organizes data elements and their relationships, ensuring consistency and providing a framework for understanding the data.
- Types of data modeling
  - Conceptual: Gives a high-level view of the data structure, such as how data interacts across an organization. It does not contain technical details.
  - Logical: Focuses on the technical details of a database such as relationships, attributes and entities.
  - Physical: Depicts how a database operates. A physical data model defines all entities and attributes used.
- Bias
  - Sampling Bias: When a sample isn't representative of the population as a whole.
  - Unbiased Sampling: When a sample is representative of the population being measured.
  - Observer Bias: The tendency for different people to observe things differently
  - Interpretation Bias: The tendency to always interpret ambiguous situations in a positive or negative way.
  - Confirmation Bias: Is the tendency to search for or interpret information in a way that confirms pre-existing beliefs.
- Identifying good data sources, a good data source needs to be:
  - Reliable
  - Original
  - Comprehensive
  - Current
  - Cited
- Data Ethics
  - Data Ethics: Well-founded standards of right and wrong that dictate how data is collected, shared, and used.
  - Ownership: The idea that individuals own the raw data they provide and have control over its usage, processing, and sharing.
  - Transaction transparency: All data processing activities and algorithms should be explainable and understood by the individuals who provide their data to ensure fairness and address concerns of data bias.
  - Consent: The right of individuals to know explicit details about how and why their data will be used before agreeing to provide it.
  - Currency: Awareness of financial transactions resulting from the use of personal data and the ability to opt out.
  - Privacy: Protecting and securing personal information and data.
  - Openness: Making the data process transparent and accessible to individuals.
- Data Subject
  - An individual whose personal data is being collected, processed, or stored.
- Data often anonymized
  - Telephone numbers
  - Names

- License plates
- Social Security numbers
- IP addresses
- Medical records
- email addresses
- photographs
- Account numbers
- Databases
  - Database Normalization: Is the process of organizing data in a relational database, reating tables, and establishing relationships between those tables.
  -
- Metadata
  - Descriptive metadata: Describes a piece of data and can be used to identify it at a later point in time. (ID, Names)
  - Structural metadata: Indicates how a piece of data is organized and whether it is part of one, or more than one, data collection (Categories, Chapters, Relationships)
  - Administrative metadata: Indicates the technical source of a digital asset. (File Type, Date, etc)
- Metadata Repositories - A database specifically created to store metadata
  - Describes the state and location of themetadata
  - Describes the structures of the tables inside
  - Describes how the data flows through the repository
  - Keep track of who accesses the metadata and when
- Data integrity
  - The accuracy, completeness, consistency, and trustworthiness of data throughout its lifecycle.
- Verification
  - A process to confirm that a data-cleaning effort was well-executed and the resulting data is accurate and reliable.
- Big Picture when verifying data-cleaning
  1. Consider the business problem.
  2. Consider the goal of the project.
  3. Consider the data.
- The 4 Phases of Analysis
  - Organize data: Involves arrangin dat in a structured manner to facilitate further exploration and understanding.
  - Format and adjust: Data is modified and prepared for analysis by applying filters, sorting, or other adjustments to make it more easily digestible and accessible.
  - Get Input from others: Seeking feedback, opinions, and perspectives from other individuals to gain insights and considerations that may help in the analysis process.
  - Transform Data: Relationships and patterns within the data are identified, and calculations are made based on the available data to derive meaningful insights and make informed decisions.
- Elements of a successful visualization:
  - Information/Data: The underlying data or information being visualized.

- Story: The narrative or concept that adds meaning and interest to the data.
- Goal: The purpose or function of the visualization, making the data useful and usable.
- Visual form: The visual design and structure of the visualization.
- Elements for effective visuals
  - Clear meaning: Clearly communicating the intended insight or message to the audience.
  - Sophisticated use of contrast: Utilizing visual context and contrast to highlight the most important data in a visualization
  - Refined execution: Attention to detail and effective use of visual elements, such lines, shapes, colors, value, space and movement, in a visualizations
- Data Storytelling steps
 

Data Storytelling is communicating the meaning of a dataset with visuals and narrative that is customized for a particular audience.

  - Engage your audience
  - Create compelling visuals
  - Tell the story in an interesting narrative.
- The McCandless Method
  1. Introduce the graphic by Names
  2. Answer obvious questions before they're asked
  3. State the insight of your graphic
  4. Call out data to support that insight
  5. Tell your audience why it matters
- What not to do in a presentation
  - No story, no logical flows
  - No titles
  - Too much text
  - Hard to understand
  - Uneven and inconsistent format, no theme
  - No conclusion or recommendation
- Responding to possible objections
  - Communicate any assumptions
  - Explain why your analysis might be different than expected
  - Acknowledge that those objections are valid and take steps to investigate further
- Types of Objections
  - About the data
    - \* Where you got the data?
    - \* What systems it came from?
    - \* What transformations happened to it?
  - About your analysis
    - \* Is your analysis reproducible?
    - \* Who did you get feedback from?
  - About your findings
    - \* Do these findings exist in previous time periods?
    - \* Did you control for the differences in your data?