Google BI Certificate Notes

By Anushka Gawand, December 2024

What is Business Intelligence?

Business intelligence (BI) refers to the process of collecting, analyzing, and interpreting business data using technology and tools to generate actionable insights that help organizations make informed decisions about their operations and strategies, typically presented through dashboards and visualizations to understand past performance and identify trends for future actions; essentially, it's turning raw data into meaningful information to guide business decisions.

1 - Foundations of Business Intelligence - Project Planning

MODULE 1
- **Three Stages of BI** - Capture, Analyze, and Monitor.
- **Stakeholder Requirements Document** - What is the business problem? Who are the stakeholders and how will they use the BI tool and what are their primary requirements? What are the datasets? Who will have access to the final product?
- **Project Requirements Document** - Contains the purpose, key dependencies, stakeholder requirements, success criteria, user journeys, assumptions, compliance and privacy, accessibility, roll-out plan, and references.
- **Strategy Document** - Stakeholders discuss what metrics will be required, how metrics are calculated, and any limitations or assumptions that exist about the data. Then the BI professional provides stakeholders with a dashboard mockup to get valuable feedback

MODULE 2
- **Project Sponsor** - Has overall accountability for the project. Provides support and resources for a project and is accountable for enabling its success.
- **Developer** - A person who uses programming languages to create, execute, test, and troubleshoot software applications. Includes application and system software developers.
- **Systems Analyst** - A person who identifies ways to design, implement, and advance information systems in order to make it possible to achieve business goals.
- **Business Stakeholders** - Business stakeholders can include the following, executive team, customer facing team, and data science team.
- **Communication** - Draw out mock up dashboards and share with stakeholders. Understand Deliverables. Be aware of bias and make sure your work is fair. Understand what your audience already knows and what they need to know and how you can deliver what they need to know to them.
- **SMART Questions** - Data analysts must ask specific, measurable, action-oriented, relevant, and time-bound questions. Do not ask close-ended or leading questions.

MODULE 3
- **Metric** - A single, quantifiable data point that is used to evaluate performance
- **North Star Metric** - Intended to capture the core measurable value of a business's product or services over its entire lifetime.
- **Key Points When Choosing Metrics** - Number of metrics, alignment with business needs, tech in place to obtain metrics, and frequency of collection.

2 - The Path to Insights: Data Models and Pipelines - Data Preparation

MODULE 1
- **Database Schemas** - DB schema is a way for describing how data is organized. Star, snowflake, flat, document, key-value, wide column, and graph are all different types of DB schemas. A DB schema should include relevant data, names and data types for each column in each table, consistent formatting and unique keys.
- **Data Pipeline** - Data pipeline is a series of processes that transports data from different sources to their final destination for storage and analysis. Pulls data from multiple sources, consolidate it, and then migrate it over to its proper destination.
- **ETL** - Extract, Transform and Load. Ingesting the raw data, processing and consolidating it into categories, and dumping the data into reporting tables that users can access.

MODULE 2
- **Five Factors of DB Performance** - Workload, throughput, resources, optimization, contention.
- **Query Plan** - Order queries to ensure they run efficiently. Description of the steps the database system takes in order to execute a query
- **Indexing** - An index is an organizational tag used to quickly locate data within a database system. If the tables within a database haven't been fully indexed, it can take the database longer to locate resources.
- **Partitioning** - Data partitioning is the process of dividing a database into distinct logical parts in order to improve query processing and increase manageability. The distribution of data within the system is extremely important. Ensuring that data has been partitioned appropriately and consistently, is part of optimization too
- **Optimize SQL Queries** - Avoid outer join and use inner joins, avoid doing SELECT * and do specific columns instead. Use WHERE clauses efficiently.

MODULE 3
- **Data Quality Testing** - Checking data for defects in order to prevent system failures. Involves the following seven validation elements. Completeness, consistency, conformity, accuracy, redundancy, integrity, and timeliness. Other things to check are data mapping, inconsistencies, inaccuracies and duplications.
- **Data lineage** - The process of identifying the origin of data, where it has moved throughout the system, and how it has transformed over time.
- **Data Dictionary** - A collection of information that describes the content, format, and structure of data objects within a database, as well as their relationships
- **Schema Validation** - Check if keys are still valid, table relationships are preserved and conventions are consistent.
- **Business Rules** - Statement that creates a restriction on specific parts of a database.

3 - Decisions, Decisions: Dashboards and Reports - Dashboard Design

MODULE 1
- **Types of Dashboards** - Strategic, focuses on long-term goals and strategies at the highest level of metrics. Operational, tracks short-term performance and intermediate goals. Analytic, consists of the datasets and the mathematics used in these sets.
- **Key Elements to a Good Visualization** - Information, story, goal, and visual form.

MODULE 2
- **Pre-aggregation** - Performing calculations on the data while it's still in the database. This means reducing the number of rows or the size of the dataset before it's used in analysis or a dashboard.
- **Data Aggregation** - The process of gathering data from multiple sources in order to combine it into a single summarized collection.
- **Dimensions** - Non-numeric qualitative data such as names, locations, and observations.
- **Measures** - Numeric quantitative data such as temperature, weight, time and distance.
- **Discrete Data** - Type of quantitative data which has a limited number of values.
- **Continuous Data** - Type of quantitative data that can be measured using a timer and its value can be shown as a decimal with several places.
- **Nominal Data** - Qualitative data that's categorized without a set order. Like yes or no.
- **Ordinal Data** - Qualitative data with a set order or scale, like ranking 1-5.
- **Privacy** - Object-level permission controls the availability of a single item in a dashboard. Row-level permission controls the availability of specific rows of a table or dataset in a dashboard. Public availability allows anyone to access a dashboard.

MODULE 3
- **Questions to Ask When Designing a Dashboard** - Which parts of the dataset will enable you to address the business requirement? Which measures and dimensions will be represented in each chart? Which chart types and forms of encoding/charts/graphs will be most effective? And how many charts are required?
- **Dashboard Design Best Practices** - Place the most relevant visualizations at the top of the dash. Use hierarchy and contrast to emphasize supporting elements. Consider how to incorporate negative space between charts.

MODULE 4
- **Presentation and Communication Best Practices** - Engage the audience. Prioritize most relevant info. Use clear and concise language. Address stakeholder concerns.
- **Presentation Breakout** - 1, Introduction. 2, Project overview, goals and purpose. 3, Data and analysis, where the data came from and how it was analysed. 4, Recommendations and actionable next steps. 5, Questions and feedback from the audience.