

Google Data Analytics Certificate - Module 4

Overview

Module 4, *Process Data from Dirty to Clean*, focuses on the "Process" phase, covering data cleaning techniques to ensure data integrity. It emphasizes tools like spreadsheets and SQL for cleaning data. Designed for beginners, this module requires approximately 20 hours to complete and includes videos, readings, assignments, and quizzes.

Key Components

- **Data Cleaning:** Identifying and correcting errors (e.g., duplicates, missing values) in datasets.
- **Data Integrity:** Ensuring accuracy, consistency, and reliability of data.
- **Spreadsheet Techniques:** Using functions (e.g., VLOOKUP, IF) for data cleaning in spreadsheets.
- **Advanced SQL Queries:** Filtering, sorting, and cleaning data using SQL commands.
- **Data Validation:** Verifying data quality through checks and balances.
- **Transferable Skills:** Applying technical skills (e.g., Microsoft Excel proficiency) to clean and validate data.

Delivery and Assessment

The module includes videos, readings, assignments (e.g., cleaning a dataset in Google Sheets), and quizzes, delivered by Google instructors. It is self-paced and includes activities like writing SQL queries for data cleaning. Assessments include quizzes on data cleaning techniques.

Learning Outcomes

Participants will learn to clean and validate data using spreadsheets and SQL, ensuring data integrity. They will develop skills to prepare clean datasets for analysis, ready for the next module.

Conclusion

Module 4 equips learners with the skills to process and clean data effectively, aligning with industry practices. It prepares them for analyzing data in subsequent modules.