Data Analyst Certification Study Guide

Please use this study guide to create your certification self-study plan. We've included the objectives you should meet for each assessed competency, with links to relevant skill assessments.

- **Associate Certification**
  - Exam DA101
- **Professional Certification**
  - Exams DA101 and DA201

Associate and Professional

Exam DA101: Data Management and Exploratory Analysis in SQL, Exploratory Analysis and Statistical Experimentation Theory

1.1 Perform data extraction, joining and aggregation tasks

- Aggregate numeric, categorical variables and dates by groups using PostgreSQL.
- Interpret a database schema and combine multiple tables by rows or columns using PostgreSQL.
- Extract data based on different conditions using PostgreSQL.
- Use subqueries to reference a second table (e.g. a different table, an aggregated table) within a query in PostgreSQL

1.2 Perform cleaning tasks to prepare data for analysis

- Match strings in a dataset with specific patterns using PostgreSQL.
- Convert values between data types in PostgreSQL
- Clean categorical and text data by manipulating strings in PostgreSQL.
- Clean date and time data in PostgreSQL.

1.3 Assess data quality and perform validation tasks

- Identify and replace missing values using PostgreSQL.
- Perform different types of data validation tasks (e.g. consistency, constraints, range validation, uniqueness) using PostgreSQL.
- Identify and validate data types in a data set using PostgreSQL.

Data Analyst Certification Study Guide

Related Assessment

Data Management in SQL (PostgreSQL)

2.1 Calculate metrics to effectively report characteristics of data and relationships between features

- Calculate measures of center (e.g. mean, median, mode) for variables using PostgreSQL.
- Calculate measures of spread (e.g. range, standard deviation, variance) for variables using PostgreSQL.
- Calculate skewness for variables using PostgreSQL.
- Calculate missingness for variables and explain its influence on reporting characteristics of data and relationships in PostgreSQL.
- Calculate the correlation between variables using PostgreSQL.

Related Assessment

Data Analysis in SQL (PostgreSQL)

2.2 Read and analyze data visualizations to demonstrate characteristics of data

- Distinguish between different types of data visualizations (e.g. bar chart, box plot, line graph, and histogram) in demonstrating the characteristics of data.
- Interpret the data visualizations (e.g. bar chart, box plot, line graph, and histogram) and summarize the characteristics of the data.

2.3 Read and analyze data visualizations to represent the relationship between features

- Distinguish between different types of data visualizations (e.g. scatterplot, heatmap, and pivot table) in representing the relationships between features.
- Interpret the data visualizations (e.g. scatterplot, heatmap, and pivot table) and summarize the relationship between features.

3.1 Describe statistical concepts that underpin hypothesis testing and experimentation

- Define different statistical distributions (e.g. binomial, normal, Poisson, t-distribution, chi-square, and F-distribution, etc.).
- Explain the statistical concepts in hypothesis testing (e.g. null hypothesis, alternative hypothesis, one-tailed and two-tailed hypothesis tests, etc.).

Data Analyst Certification Study Guide

- Explain the statistical concepts in the experimental design (e.g. control group, randomization, confounding variables, etc.).
- Explain parameter estimation and confidence intervals.

Professional only

Exam DA201: Data Management, Exploratory Analysis, and Statistical Experimentation in R or Python

1.1 Perform standard data import, joining and aggregation tasks

- Import data from flat files into R or Python.
- Import data from databases into R or Python
- Aggregate numeric, categorical variables and dates by groups using R or Python.
- Combine multiple tables by rows or columns using R or Python.
- Filter data based on different criteria using R or Python.

1.2 Perform standard cleaning tasks to prepare data for analysis

- Match strings in a dataset with specific patterns using R or Python.
- Convert values between data types in R or Python.
- Clean categorical and text data by manipulating strings in R or Python.
- Clean date and time data in R or Python.

1.3 Assess data quality and perform validation tasks

- Identify and replace missing values using R or Python.
- Perform different types of data validation tasks (e.g. consistency, constraints, range validation, uniqueness) using R or Python.
- Identify and validate data types in a data set using R or Python.

1.4 Collect data from non-standard formats by modifying existing code

- Adapt provided code to import data from an API using R or Python.
- Identify the structure of HTML and JSON data and parse them into a usable format for data processing and analysis using R or Python.

Data Analyst Certification Study Guide

Related Assessments

Importing and Cleaning with R

Importing and Cleaning with Python

2.1 Calculate metrics to effectively report characteristics of data and relationships between features

- Calculate measures of center (e.g. mean, median, mode) for variables using R or Python.
- Calculate measures of spread (e.g. range, standard deviation, variance) for variables using R or Python.
- Calculate skewness for variables using R or Python.
- Calculate missingness for variables and explain its influence on reporting characteristics of data and relationships in R or Python.
- Calculate the correlation between variables using R or Python.

2.2 Create data visualizations in R or Python to demonstrate the characteristics of data

- Create and customize bar charts using R or Python.
- Create and customize box plots using R or Python.
- Create and customize line graphs using R or Python.
- Create and customize histograms graph using R or Python.

2.3 Create data visualizations in R or Python to represent the relationships between features

- Create and customize scatterplots using R or Python.
- Create and customize heatmaps using R or Python.
- Create and customize pivot tables using R or Python.

Related Assessments

Data Manipulation with R

Data Manipulation with Python

3.1 Apply sampling methods to data

- Distinguish between different types of random sampling techniques and apply the methods using R or Python

Data Analyst Certification Study Guide

- Sample data from a statistical distribution (e.g. normal, binomial, Poisson, exponential, etc.) using R or Python
- Calculate a probability from a statistical distribution (e.g. normal, binomial, Poisson, exponential, etc.) using R or Python

3.2 Implement methods for performing statistical tests

- Run statistical tests (e.g. t-test, ANOVA test, chi-square test) using R or Python.
- Analyze the results of statistical tests from R or Python.

Related Assessments

Statistics Fundamentals with R

Statistics Fundamentals with Python