

# **CDAP: Certified Data Analyst Professional**

**Modality: On Demand**

**Duration: 5 Hours**

## **Overview of CDAP**

Data is everywhere today! To not take advantage of it is like having an arsenal and not being able to use it in a war. Let's face it, the advent of digitalization has pushed modern organizations into a war of grabbing a customer's attention, serving him with an experience, and retaining him to become a returning customer. All this can happen, only if you thoroughly study and understand your customer. This function of knowing your customers using data is what data analysts are responsible for. The **"CDAP: Certified Data Analyst Professional"** certification validates your skills as a knowledgeable data analyst, having a commendable grip on concepts, tools and techniques of reading, understanding, visualizing, creating, and presenting data into easily understandable stories. The course curriculum covers topics around stats, math, data engineering, visualization and much more, to make you an expert in aligning and analyzing large, scattered chunks of data and pulling out actionable insights from it.

## **Who Is This Cert Suitable For?**

The certification is not for those who are planning to learn data analysis. If you have around a year or two of experience and have an overall idea of the basics of manipulating data and visualizing it, you should prepare for and take this certification exam to become a well-versed data analyst. The course curricula for this certification includes topics that can up your skillset by many folds, giving you that expert command on the functions, methodologies, disciplines, and techniques to play with data and use it to a business's advantage.

## **Pre-requisites for the Cert Exam**

The Data Analysis certification exam requires candidates planning to appear for it come with well learned fundamentals of statistics, reflecting a systematic exposure and comprehension of the principles. They must be hands-on in managing core pattern estimation problems. They will be expected to know techniques and methods and to work around dispersions and averages, permutations and combinations, and probability. In visualization, understanding of graphs, maps, histograms, and scatter plots development would help a lot in becoming friends with this certification curriculum.

## **Course Outline:**

### **Professional Data Analyst Skills**

- CRISP-DM Methodology
- Define the major components of the CRISP-DM Analytic cycle

- Explain why the steps are critical
- Confusion Matrix and Translating Business Impact
- Understand the relationship between the confusion matrix and business application being addressed
- List specific real-world examples of how the confusion matrix can be used to drive business decisions.
- The Data Science Pipeline
- List the major steps in the data science pipeline.
- Explain why the data science pipeline steps are important.
- Signal vs. Noise in Data
- Understand importance of sample size and how to identify when a signal is significant
- Identify ways to improve signal to noise ratio in data
- Preprocessing Techniques
- Explain the objective of preprocessing
- List techniques that can be used to preprocess different data types

## **Data Engineering Skills**

- Understanding Small Data
- Define the difference between big and small data
- Understand the algorithms used for big and small data
- SQL
- Analyze different types of queries for completeness and accuracy
- Understand how different SQL clauses interact and affect query results
- Data Encoding and File Formats

- List different file formats
- Understand pros and cons of different compression formats
- Understand structured vs unstructured data
- Amazon S3
- Select the benefits of S3 storage over standard storage.
- Understand methods of encryption and access for s3 files
- Connecting to datastores with R and Python
- List packages that can be used to access s3 and SQL in R and Python
- Select steps needed to connect to s3 and SQL in R and Python
- Understanding Big Data
- Define the difference between big and small data
- Understand the algorithms used for big and small data
- Spark
- Explain how Spark handles big data
- Describe a use case for Spark
- Hadoop
- Explain how Hadoop handles big data
- Describe a use case for Hadoop
- Data Lake
- Explain how Data Lakes handles big data
- Describe a use case for Data Lakes
- Connecting to datastores with R and Python
- List packages that can be used to access spark and hadoop in R and Python
- Select steps needed to connect to spark and hadoop in R and Python

## **Data Visualization Skills**

- Delivering Impactful Presentations
- Analyze presentations for errors and biases
- Describe impactful analytical figures and graphs
- Plotting Histograms
- Understand what a histogram represents
- Understand how a histogram is constructed
- Plotting Line Graphs
- Understand what a line graph represents
- Understand how a line graph is constructed
- Plotting Scatter Plots
- Understand what a scatter plot represents
- Understand how a scatterplot is constructed
- Visualizing with R and Python
- List visualization libraries in R and Python
- Understand basic visualization code in R and Python

## **Mathematics and Statistics**

- Frequentist Statistics
- Understand p value
- Understand bootstrapping
- Linear Algebra
- Understand Vectors
- Understand PCA transformations
- Variance, Mean, Median and Mode

- Explain concepts of Variance, Mean, Median and Mode
- Describe when to use Variance, Mean, Median and Mode
- Central Limit Theorem
- Describe at a high level what this concept is
- Identify how the Central Limit Theorem benefits Machine Learning
- Bias/Variance Tradeoff
- Explain the tradeoff between bias and variance
- Explain what bias and variance are in terms of modeling
- Bayesian Statistics
- Describe Bayes Rule
- Describe how machine learning updates a belief to fit a model
- Sampling Methods and Assumptions
- Describe Random, Stratified and Clustered Sampling
- Understand why and when to use different methods
- Confidence and Credible Intervals
- Understand what a confidence interval is
- Understand what a credible interval is
- Basic Calculus
- Explain what differentiation and integration are
- Cite examples of Calculus being used in Optimization
- Important metrics for Data Science
- Describe Classification metrics
- Describe Regression metrics