Teradata Certification

Exam Objectives

The Data Science Exam covers the features and functionality of Vantage 1.1 including the Advanced SQL Engine through release 16.20.

Data Management and Governance – 15%

- Given a graphic representation of data, identify supporting statistical evidence of the distribution, skew, and outliers.
- Given a Moments table from the UnivariateStatistics function, identify assumptions about the population.
- Given a graphic or a set of numbers, identify complex data quality issues.
- Given a description of a complex quality issue, identify the SQL code snippet that should be used to correct the problem.
- Given a scenario with missing data, identify the correct metric to remediate the missing data issue.
- Given a text analytic task, identify the correct sequence of preprocessing functions to prepare the data to accomplish the task.
- Given a data set with a specific distribution, identify the sampling strategy that should be used.
- Identify the characteristics of random sampling and stratified sampling.
- Identify the purpose of SAX.
- Given a complex scenario, identify the CASE WHEN statement that should be used to accomplish the task.
- Identify when aggregate windowing functions should be used.
- Given a data set and Teradata's R and Python packages, identify the appropriate loading statement.

Statistical Techniques – 20%

- Identify the definition of heteroscedasticity and describe its effects.
- Identify the definition of monotonicity and why it is important.
- Given a data set, identify how an outlier affects the modeling approach that should be used.
- Identify the expected behavior from a model if outliers are not removed.
- Identify the use for PCA and identify the analytics workflow that uses PCA.
- Identify the relationship between PCA and multicollinearity.
- Given a scatter plot matrix, identify the level of correlation of the elements.
- Given multiple distributions, identify the appropriate hypothesis test method.
- Given a target variable type, identify the approaches that should be used to model it.
- Given a data set and independent data variable types, identify the model that should be used.
- Given a data set that has a nonlinear relationship, identify the data manipulation function that allows linear modeling.
- Identify the risks associated with assuming linearity.
- Given a model output, identify the interpretation of GLM coefficients.

- Given a model's goodness of fit test statistics, identify the interpretation of the results.
- Given a purpose, identify the visualization that should be used.
- Given an output from a data function, identify the visualizations that can be created in Teradata AppCenter.

Data Analytics Methods and Algorithms – 36%

- Given a complex text mining task, identify the combination of functions that should be used to complete the task.
- Identify the steps to implement a custom dictionary.
- Identify uses for Parts of Speech (POS) Tagger and lemmatization.
- Identify the purpose of LDA and when it should be used.
- Given a NaiveBayes model text classification output, interpret the probability of document classification.
- Identify the meaning of TD-IDF and its utility.
- Given a complex npath statement, identify how the function will operate.
- Given a complex output, identify the npath statement that created the output.
- Given a complex scenario, identify the sessionize statement that created the output.
- Identify the usage and characteristics of supervised and unsupervised Hidden Markov Models (HMM).
- Identify how Shapley values are used as inputs to attribution functions.
- Identify the various model inputs to the attribution functions that affect the outputs.
- Identify how the VARMAX model extends the ARIMA model.
- Given a SQL snippet using the ARIMA function, identify the parameters.
- Given two survival distributions, identify a description of survival probabilities for the two populations.
- Given a survival analysis scenario, identify how to order the COX functions for a viable solution.
- Identify the usage of a Period data type.
- Identify the functions that can be performed on Period Data Types.
- Identify how to use Interpolator to process missing time series data.
- Identify the function of ChangePointDetection.
- Given a scenario, identify the geospatial function that should be used.
- Identify the calculations that can be performed with geospatial data types.
- Identify the benefits of using ADABoost on a classification problem relative to alternate methods.
- Identify the difference between LARs and PCA.
- Identify the meaning of metrics in the CFilter function.
- Match the cluster topics including KMeans, Gaussian Mixture Models (GMM), Canopy, MinHash with their definitions, uses or characteristics.
- Match graph centrality functions with their definitions, uses or characteristics.
- Identify the characteristics of the PageRank function.
- Identify the characteristics, benefits, and uses of TDPLYR and TeradataML.

Validation and Evaluation – 16%

- Given a complex graphic, interpret the results.
- Identify the characteristics of Type I and Type II errors.

- Identify why crossvalidation is used.
- Given a model output or data, identify the business value in terms that can be interpreted by business leaders.

Productionalization – 13%

- Given a SQL code snippet, identify the proper syntax to use a training output table in a given scoring function.
- Identify key characteristics for model management.
- Identify how Kubernetes and Docker are used in the Teradata Vantage architecture.
- Identify how QueryGrid enables ecosystem architecture with Teradata Vantage.
- Identify the principles of scalability with data science in the Teradata Vantage platform.
-