PassQueen
a.k.a dumpsqueen.com

Dell Data Science Foundations

EMC D-DS-FN-23

Version Demo

Total Demo Questions: 7

Total Premium Questions: 79

Buy Premium PDF

https://passqueen.com

support@passqueen.com

Topic Break Down

Topic	No. of Questions
Topic 1, Data Science and Data Analytics	11
Topic 2, Data Analytics Lifecycle	7
Topic 3, Data Acquisition, Preparation, and Exploration	33
Topic 4, Machine Learning	21
Topic 5, Model Evaluation and Deployment	4
Topic 6, Mix Questions	3
Total	79

QUESTION NO: 1 - (SIMULATION)

Match each task to its description.

ANSWER: See the explanation for the answer

Explanation:

The simulation prompt only states: *"Match each task to its description."* No task names, description choices, or simulation image/interface options were provided. Therefore, the required matches cannot be determined from the supplied JSON.

Please provide the task and description options (or a screenshot of the simulation) to identify the correct matches.

QUESTION NO: 2

Which SQL set operator returns rows that exist in the first SELECT statement answer set but not in the second SELECT statement?

A. EXCEPT

B. UNION

C. UNION ALL

D. INTERSECT

ANSWER: A

Explanation:

EXCEPT returns the set difference between two query result sets: it outputs rows returned by the first `SELECT` that do not occur in the second `SELECT`. The queries used with this operator must produce the same number of columns, with corresponding columns having compatible data types. Conceptually, it is useful when comparing two populations or lists, such as finding records present in a source dataset that have not yet appeared in a target dataset. Set operators apply distinct-result behavior by default, so `EXCEPT` returns unique rows rather than duplicate occurrences. The SQL standard defines `EXCEPT` as the operation for subtracting the right-hand query result from the left-hand query result. This direction matters: the first query supplies the candidate rows, and the second query identifies rows to remove from that candidate set. For SQL Server-specific syntax and behavior, see Microsoft Learn: EXCEPT and INTERSECT. General PostgreSQL documentation describing set-combining operations, including `EXCEPT`, is available in the PostgreSQL SQL documentation.

QUESTION NO: 3

Refer to the exhibit.

What is the approximate R-squared value for a linear regression model fitted to the data associated with this scatterplot?

A. 4

B. 0.96

C. 0.25

D. 16

ANSWER: B

Explanation:

0.96 is the appropriate approximate R-squared value. R-squared, also called the coefficient of determination, measures the proportion of variation in the dependent variable that is accounted for by the fitted regression model. Its usual interpretation is as a value from 0 to 1, with values closer to 1 indicating that the observations lie very closely around the regression line. In the scatterplot, the data display a strong linear pattern with relatively little spread around the fitted line, so the model explains nearly all of the observed variation. An R-squared value of 0.96 means that approximately 96% of the variation in the response values is explained by the linear relationship with the predictor. This is consistent with a visually tight, strongly linear scatterplot. R-squared is a goodness-of-fit summary for the model and is commonly calculated as one minus the ratio of residual sum of squares to total sum of squares. For additional detail, see the scikit-learn documentation on the R^2 score and the NIST discussion of the coefficient of determination.

QUESTION NO: 4

When should you consider using multinomial logistic regression over binary logistic regression?

A. Dependent variable is continuous or dichotomous

B. Dependent variable is continuous or categorical

C. Dependent variable has more than two categories

D. Dependent variable is continuous only

ANSWER: C

Explanation:

Multinomial logistic regression is appropriate when the dependent (target) variable is nominal categorical and has more than two mutually exclusive outcome categories. It extends binary logistic regression, which models the probability of one of two possible classes, by estimating the relative probability of each category against a selected reference category. For example, it can model a customer outcome classified as “purchase,” “no purchase,” or “defer,” provided these categories do not have an inherent order. The model uses predictor variables to estimate class probabilities, and the predicted category is commonly the category with the greatest estimated probability. This makes it useful for multiclass classification problems where the outcome is categorical rather than continuous. Careful category definition and selection of a meaningful reference class help make model coefficients interpretable. If the outcome categories have a natural ranking, an ordinal logistic model may instead be appropriate because it can explicitly use that ordering. For the stated condition—an outcome with more than two categories—multinomial logistic regression is the relevant logistic-regression approach. See the scikit-learn logistic regression documentation and the IBM documentation on multinomial logistic regression.

QUESTION NO: 5

An analyst creates an R data frame containing customer ID values, account balances, account-opening dates, and customer segments.

What characteristic allows this data frame to store these columns together?

A. Each column can contain a different data type

B. Each row must contain only numeric values

C. All columns must use the same atomic type

D. The structure must contain exactly two columns

ANSWER: A

Explanation:

Each column can contain a different data type is correct. A data frame is a list-like tabular structure in R in which columns can have different types, such as numeric, character, logical, factor, or date-related classes. Values within an individual

column are generally of a common type. This differs from a matrix or array, whose elements are stored as one common atomic type.

QUESTION NO: 6

Refer to the exhibit.

To predict whether or not a customer will renew their annual property insurance policy, an insurance company built and operationalized a naïve Bayes classification model. In the model, there are two class labels, renewal and non-renewal, that are assigned to each customer based on their attributes.

A subset of the key attributes, their values, and corresponding conditional probabilities are provided in the exhibit.

A customer has the following attributes:

● Age is greater than 65 years

● Owns their own home

● Renewal month is August

If 20% of customers do not renew the police every year, what is the **score for a renewal** in the naïve Bayesian model for the customer described above?

A. 0.0022

B. 0.0027

C. 0.0270

D. 0.0216

ANSWER: D

Explanation:

The correct score for a renewal is **0.0216**. A naïve Bayes classifier calculates an unnormalized class score by multiplying the prior probability of the class by the conditional probability of each observed attribute given that class. Since 20% of customers do not renew, the prior probability that a customer renews is 80%, or 0.8. For the described customer, the exhibit provides the renewal-class conditional probabilities: 0.3 for being older than 65, 0.9 for owning a home, and 0.1 for having an August renewal month. Under the naïve Bayes conditional-independence assumption, the renewal score is therefore 0.8 × 0.3 × 0.9 × 0.1, which equals 0.0216. This is a class score, rather than necessarily a final normalized posterior probability. To produce a posterior probability of renewal, the model would also calculate the corresponding non-renewal score and normalize the two scores so that they sum to 1. The multiplication approach used here is the standard naïve Bayes decision rule for categorical features. See the scikit-learn Naive Bayes documentation and the Naive Bayes classifier overview.

QUESTION NO: 7

What is a key consideration when preparing a presentation intended for analysts?

A. Describe how to implement the model

B. Provide talking points to promote or evangelize the project

C. Emphasize the business benefits of implementing the model

D. Focus on clean simple-to-understand visuals

ANSWER: D

Explanation:

Focus on clean simple-to-understand visuals is the key consideration for an analyst-oriented presentation. Analysts need to rapidly inspect data, recognize trends, compare measures, identify outliers, and form or validate hypotheses. Clear visual design makes those analytical tasks more efficient by presenting the relevant relationships without unnecessary visual clutter. Charts should use meaningful labels, appropriate scales, concise titles, and a layout that makes the main finding and supporting evidence easy to locate. Simplicity does not mean omitting analytical substance; it means structuring the visual story so that the audience can accurately interpret the model outputs, data patterns, assumptions, and performance measures. This supports reproducible discussion because analysts can clearly see what the data is showing before deciding whether to investigate further or act on the findings. The U.S. Census Bureau's guidance on data visualization emphasizes clarity, accurate representation, and design choices that help audiences understand data: U.S. Census Bureau Data Visualization. Similarly, the Data Visualization Society identifies effective communication and audience understanding as central goals of visualization practice: Data Visualization Society.