edquest

# Recap

## EXPLORATORY DATA ANALYSIS (EDA)


edquest

## Q1: The syntax df.shape returns a tuple representing ___ of the dataset.

A. correlation and variance

B. mean and median

C. number of rows and columns

D. size in memory


edquest

## Q1: The syntax df.shape returns a tuple representing ___ of the dataset.

A. correlation and variance

B. mean and median

C. number of rows and columns

D. size in memory


edquest

## Q2: To check the percentage of missing values column-wise, we use df.isnull().mean(), which gives us values between ____ and ____.

A. 1 and 100

B. 0 and 1

C. -1 and 1

D. 0 and 100


edquest

## Q2: To check the percentage of missing values column-wise, we use df.isnull().mean(), which gives us values between ____ and ____.

A. 1 and 100

B. 0 and 1

C. -1 and 1

D. 0 and 100


edquest

## Q3: The line

**df = df.drop(columns=null_percent[null_percent > 0.3].index) is used to remove columns with more than ____% missing values.**

A. 3

B. 10

C. 30

D. 50


edquest

## Q3: The line

**df = df.drop(columns=null_percent[null_percent > 0.3].index) is used to remove columns with more than ____% missing values.**

A. 3

B. 10

C. 30

D. 50


edquest

## Q4: To examine the summary statistics of numerical features, the code `df[num_cols].describe().T[:5]` is used. The .T stands for ____.

A. table format

B. trimming values

C. transpose

D. top values


edquest

## Q4: To examine the summary statistics of numerical features, the code df[num_cols].describe().T[:5] is used. The .T stands for ____.

A. table format

B. trimming values

C. transpose

D. top values


edquest

## Q5: The function sns.boxplot() is used to detect ___ and visualize distribution characteristics.

A. duplicates

B. outliers

C. nulls

D. categories


edquest

## Q5: The function sns.boxplot() is used to detect ____ and visualize distribution characteristics.

A. duplicates

B. outliers

C. nulls

D. categories


edquest

## Q6: A histogram with KDE (Kernel Density Estimation) is plotted using `sns.histplot(df['SalePrice'], kde=True)`. The KDE line helps to understand the ___ of the data.

A. completeness

B. sparsity

C. central tendency

D. probability density


edquest

## Q6: A histogram with KDE (Kernel Density Estimation) is plotted using `sns.histplot(df['SalePrice'], kde=True)`. The KDE line helps to understand the ___ of the data.

A. completeness

B. sparsity

C. central tendency

D. probability density


edquest

## Q7: The df.select_dtypes(include='object') line is used to select ___ features in the dataset.

A. numerical

B. categorical

C. Boolean

D. missing


edquest

## Q7: The df.select_dtypes(include='object') line is used to select ___ features in the dataset.

A. numerical

B. categorical

C. Boolean

D. missing


edquest

## Q8: The correlation matrix is created using df.corr(numeric_only=True). Setting numeric_only=True ensures that only ____ data types are used in the computation.

A. float and bool

B. object and float

C. int and float

D. datetime and object


edquest

**Q8: The correlation matrix is created using df.corr(numeric_only=True). Setting numeric_only=True ensures that only ____ data types are used in the computation.**

A. float and bool

B. object and float

C. int and float

D. datetime and object


edquest

## Q9: The sns.heatmap() function with annot=True overlays ____ on the heatmap cells.

A. feature names

B. colors

C. correlation values

D. histogram bars


edquest

## Q9: The sns.heatmap() function with annot=True overlays ____ on the heatmap cells.

A. feature names

B. colors

C. correlation values

D. histogram bars


edquest

## Q10: In the context of EDA, the main reason to remove columns with high null percentages is to avoid ___ during model training.

A. faster execution

B. better styling

C. bias and error

D. overfitting


edquest

## Q10: In the context of EDA, the main reason to remove columns with high null percentages is to avoid ___ during model training.

A. faster execution

B. better styling

C. bias and error

D. overfitting

edquest

# Knowledge Test Day 3

## REGULARIZATION + DEEP LEARNING OVERVIEW


edquest

## Q1. Overfitting occurs when a model performs well on training data but poorly on test data because it ____ the training patterns.

A. ignores

B. generalizes

C. memorizes

D. transforms


edquest

## Q1. Overfitting occurs when a model performs well on training data but poorly on test data because it ____ the training patterns.

A. ignores

B. generalizes

C. memorizes

D. transforms


edquest

## Q2. Underfitting typically happens when the model is too ____ to capture the underlying structure of the data.

A. large

B. simple

C. complex

D. deep


edquest

## Q2. Underfitting typically happens when the model is too ____ to capture the underlying structure of the data.

A. large

B. simple

C. complex

D. deep


edquest

## Q3. One of the key goals of L1 and L2 regularization is to prevent ____ by penalizing large weights.

A. optimization

B. dropout

C. overfitting

D. training


edquest

## Q3. One of the key goals of L1 and L2 regularization is to prevent ____ by penalizing large weights.

A. optimization

B. dropout

C. overfitting

D. training


edquest

## Q4. L1 regularization encourages ____ by forcing some weights to become exactly zero.

A. sparsity

B. overfitting

C. weight duplication

D. normalization


edquest

## Q4. L1 regularization encourages ___ by forcing some weights to become exactly zero.

A. sparsity

B. overfitting

C. weight duplication

D. normalization


edquest

## Q5. In L2 regularization, the penalty added to the loss function is proportional to the ___ of the weights.

A. mean

B. square

C. absolute value

D. maximum


edquest

## Q5. In L2 regularization, the penalty added to the loss function is proportional to the ___ of the weights.

A. mean

B. square

C. absolute value

D. maximum


edquest

## Q6. A neural network consists of input, hidden, and output layers where each neuron computes a weighted sum followed by a(n) ___ function.

A. loss

B. dropout

C. activation

D. optimizer


edquest

## Q6. A neural network consists of input, hidden, and output layers where each neuron computes a weighted sum followed by a(n) ___ function.

A. loss

B. dropout

C. activation

D. optimizer


edquest

## Q7. In deep learning, the term “deep” refers to the presence of ___ layers in the network.

A. very wide

B. many hidden

C. few output

D. multiple input


edquest

## Q7. In deep learning, the term “deep” refers to the presence of ___ layers in the network.

A. very wide

B. many hidden

C. few output

D. multiple input


edquest

## Q8. Regularization is typically added to the model’s ____ function to constrain the learning process.

A. activation

B. loss

C. dropout

D. optimizer


edquest

## Q8. Regularization is typically added to the model’s ____ function to constrain the learning process.

A. activation

B. loss

C. dropout

D. optimizer


edquest

## Q9. A key indicator of overfitting is when training accuracy is high, but ___ accuracy is low.

A. validation

B. learning rate

C. training

D. test loss


edquest

## Q9. A key indicator of overfitting is when training accuracy is high, but ___ accuracy is low.

A. validation

B. learning rate

C. training

D. test loss


edquest

## Q10. One benefit of using dropout layers in neural networks is that they prevent co-adaptation of neurons by randomly ___ some during training.

A. scaling

B. deleting

C. updating

D. deactivating