

Module 6 Quiz & Explanation

1) Which plot is best suited for understanding the distribution shape of a numerical feature?

- a) Scatter plot
- b) Histogram
- c) Heatmap
- d) Pairplot

Ans: b) Histogram

ব্যখ্যা: Histogram numerical feature-এর distribution shape যেমন skewed বা normal—এসব পরিষ্কারভাবে দেখায়। এটি bins ব্যবহার করে data-এর frequency প্রদর্শন করে।

2) What does a boxplot primarily help identify?

- a) Class imbalance
- b) Outliers and spread of numerical data
- c) Correlation between variables
- d) Category frequencies

Ans: b) Outliers and spread of numerical data

ব্যখ্যা: Boxplot median, IQR এবং whiskers দেখায়, যা দিয়ে data-এর spread ও possible outliers দ্রুত বোঝা যায়।

3) Which plot is most useful for detecting linear relationships between two numerical variables?

- a) Bar chart
- b) Boxplot
- c) Scatter plot
- d) Countplot

Ans: c) Scatter plot

ব্যখ্যা: Scatter plot দুইটি numerical variable-এর মধ্যকার pattern, trend বা linear সম্পর্ক বিশ্লেষণে সবচেয়ে কার্যকর।

4) What does a correlation heatmap help analyze?

- a) Distribution of categories
- b) Missing values
- c) Strength of linear relationships between numerical features
- d) Frequency of outliers

Ans: c) Strength of linear relationships between numerical features

ব্যখ্যা: Correlation heatmap numerical featureগুলোর মধ্যে সম্পর্ক (positive/negative strength) color intensity দিয়ে উপস্থাপন করে।

5) What does a highly imbalanced target variable indicate?

- a) Serious overfitting
- b) Need for resampling or class weighting
- c) More histograms required
- d) More categorical encoding required

Ans: b) Need for resampling or class weighting

ব্যাখ্যা: Target imbalance থাকলে modeling-এর সময় oversampling, undersampling বা class-weight adjustment দরকার হয় bias কমানোর জন্য।

6) Which EDA step helps check whether data types and ranges are reasonable?

- a) Correlation heatmap
- b) describe() and info()
- c) Pairplot
- d) Countplot

Ans: b) describe() and info()

ব্যাখ্যা: describe() summary statistics দেয় এবং info() data types ও missing value দেখায়। এগুলো দিয়ে ভুল range বা type mismatch ধরা যায়।

7) Which visualization is suitable for checking distribution differences of a numerical feature across a target (0/1)?

- a) Boxplot grouped by target
- b) Histogram without grouping
- c) Bar chart
- d) Line plot

Ans: a) Boxplot grouped by target

ব্যাখ্যা: Target অনুযায়ী ভাগ করা boxplot দুই class-এর distribution difference সহজে বুঝতে সাহায্য করে।

8) Pairplot is useful for which purpose?

- a) Showing missing value patterns
- b) Exploring pairwise relationships among multiple numerical variables
- c) Encoding categorical variables
- d) Detecting class imbalance

Ans: b) Exploring pairwise relationships among multiple numerical variables

ব্যাখ্যা: Pairplot scatterplots ও distributions একসাথে দেখিয়ে multiple numerical variable-এর pattern বুঝতে সাহায্য করে।

9) Which of the following is the BEST first step before any visualization?

- a) Perform normalization
- b) Convert all categorical variables
- c) Inspect shape, data types, and missing values
- d) Build a predictive model

Ans: c) Inspect shape, data types, and missing values

ব্যাখ্যা: Visualization শুরু করার আগে dataset-এর structure, missing values এবং column types জানা EDA-এর মৌলিক ধাপ।

10) What is the main purpose of exploring categorical feature frequencies?

- a) To identify outliers

- b) To check if some categories have very few samples
- c) To detect linear correlation
- d) To scale the data

Ans: b) To check if some categories have very few samples

ব্যাখ্যা: Rare category modelকে noisy করতে পারে; তাই value counts দেখে category distribution বোঝা জরুরি।