

Week Activity - DS bootcamp 🧑🧑

1 play · 25 players

أكاديمية طويق
TUWAIQ ACADEMY

🔒 A public kahoot

Questions (31)

1 - Quiz

What statistical measure is unaffected by outliers in a dataset?



Mode



Median



Mean



Average



2 - Quiz

Which of the following is NOT a valid method to handle missing data in EDA?



Removing rows with missing values



Imputing missing values with the mean of the column



Ignoring missing values during analysis



Imputing missing values with a random value from the same column



3 - Quiz

What is the purpose of using a box plot in EDA?



- ☐ To visualize the spread and skewness of the data ✓
- ☐ To identify outliers and their magnitude ✓
- ☐ To compare the means of different groups within the dataset ✗
- ☐ To show the correlation between two variables ✗

4 - Quiz

What does the `dropna()` function in Pandas do?



- ☐ Drops columns with missing values ✗
- ☐ Drops both rows and columns with missing values ✗
- ☐ Drops rows with missing values ✓
- ☐ Fills missing values with a specified value ✗

5 - Quiz

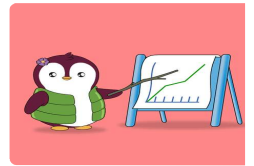
In EDA, what does the interquartile range (IQR) represent?



- ☐ The range between the first and third quartiles ✓
- ☐ The range between the maximum and minimum values ✗
- ☐ The spread of data around the mean ✓
- ☐ The standard deviation of the dataset ✗

6 - Quiz

What is the primary purpose of a correlation matrix in EDA?



To identify outliers in the dataset



To assess the normality assumption of a dataset



To visualize the distribution of data



To measure the strength and direction of relationships between variables



7 - Quiz

What does the `groupby()` function in Pandas do?



Aggregates data based on specified functions



Sorts the DataFrame based on specified columns



Groups rows based on unique values in specified columns

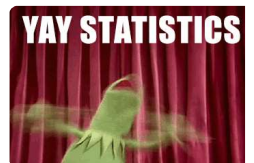


Filters rows based on a specified condition



8 - Quiz

Which measure of central tendency is most affected by outliers?



Median



Mean



Range



Mode



9 - Quiz

Which type of plot is suitable for visualizing the relationship between two continuous variables?



Box plot



Pie chart



Scatter plot

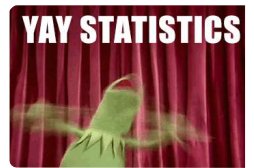


Bar plot



10 - Quiz

Which of the following is NOT a valid method for detecting outliers in EDA?



Z-score method



Mode method



IQR method

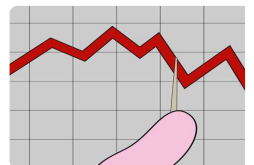


Box plot



11 - Quiz

In a histogram, what does the height of each bar represent?



Frequency



Relative frequency



Cumulative frequency



Percentage



12 - Quiz

Which Pandas function is used to check for missing values in a DataFrame?



check_missing()



missing_values()



isnull()



isna()



13 - Quiz

What is meant by the completeness of data?



The degree to which the data is accurate



The consistency of data across different sources



The extent to which all necessary data is available

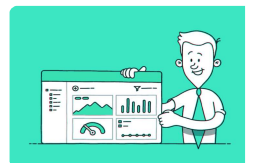


The reliability of the data analysis techniques used



14 - Quiz

Which data visualization technique is used to visualize the distribution of a single continuous variable?



Box plot



Scatter plot



Histogram



Bar plot



15 - Quiz

Which aspect of data quality refers to the accuracy and consistency of data over time?



Reliability



Consistency



Timeliness



Relevance



16 - Quiz

How much insights a box plot can deliver?



6



7



8



9



17 - Quiz

What is the primary objective of data cleaning in the context of data analysis?



Removing all missing values from the dataset



Ensuring that the dataset is error-free and consistent



Applying advanced statistical techniques to the dataset



Adding noise to the dataset to improve accuracy



18 - Quiz

What does data consistency check for in the context of data quality?



The conformity of data values to predefined rules or standards



The uniqueness of data values within a dataset



The relevance of data to the intended analysis

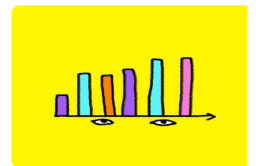


The accuracy of data values compared to a trusted source



19 - Quiz

Which type of chart is suitable for visualizing the distribution of a categorical variable?



Scatter plot



Histogram



Bar plot

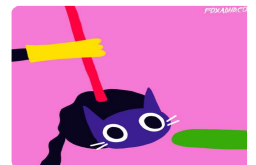


Line plot



20 - Quiz

Which of the following is NOT a common data cleaning task?



Imputing missing values



Removing duplicates



Renaming columns



Standardizing data formats



21 - Quiz

What does a box plot visualize in a dataset?

Central tendency and spread



Correlation between two variables



Distribution of data points



Outliers and quartiles



22 - Quiz

Which of the following is a consequence of incomplete data?

Improved reliability of data insights



Increased efficiency of data processing



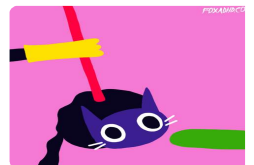
Enhanced scalability of data infrastructure



Increased probability of bias data analysis results



23 - Quiz

What is the purpose of outlier detection and treatment in data cleaning?

To increase the number of outliers in the dataset



To remove errors from the dataset



To enhance the variability of the dataset



To improve the accuracy and reliability of data analysis



24 - Quiz

What is the purpose of calculating the standard deviation of a dataset?



- ☐ To measure the central tendency of the data
- ☐ To identify outliers in the dataset
- ☒ To assess the spread or dispersion of data points
- ☐ To detect the presence of skewness in the dataset



25 - Quiz

What is the primary objective of data profiling?



- ☐ To validate the accuracy of the dataset
- ☒ To identify and understand the structure, content and quality of dataset
- ☒ To collect statistical information on the dataset
- ☐ To visualize the data distribution



26 - Quiz

What does a negative skewness value indicate about a dataset?



- ☐ The dataset is perfectly symmetrical
- ☐ The dataset is skewed to the right
- ☐ The dataset has a uniform distribution
- ☒ The dataset is skewed to the left



27 - Quiz

In which phase of analytics are statistical models built to be insightful and foresightful?



Descriptive Analytics



Diagnostic Analytics



Predictive Analytics



Prescriptive Analytics



28 - Quiz

Which data quality dimension focuses on ensuring that all necessary information for analysis is available?



Accuracy



Completeness



Uniqueness

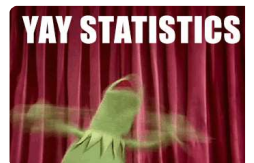


Relevance



29 - Quiz

What does a parameter represent in statistics?



A characteristic of a sample



A characteristic of a population



A measure of central tendency



A measure of spread or dispersion



30 - Quiz

What does the third quartile (Q3) represent in a dataset?



The value below which 25% of the data falls



The value below which 50% of the data falls



The value below which 75% of the data falls

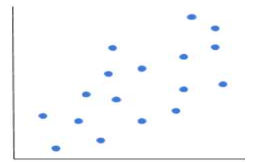


The median of the dataset



31 - Quiz

What's the correlation type in the graph below:



Weak correlation



Negatively correlated



Strong correlation



Not correlated



Resource credits ^