

Name/ID:

Questions: True (A) of False (B)

Q1: Length, weight, and density are all examples of numerical type and ratio measurement scales.

Q2: Examples of measurement levels are categorical and numerical.

Q3: Variables arising from a counting process are called continuous.

Q4: Categorical data, where ordering is not important, are nominal.

Q5: The Second Quartile is always equal to the Median.

Q6: The most common measure of central tendency that is sensitive to extreme values is the Mean.

Q7: When a sample data has a normal distribution, then the measures of central tendency are all equal.

Q8: Measures of variation include Variance, Range, and Standard Deviation.

Q9: Before model evaluation, a data analyst should ensure the data set is balanced.

Q10: Data dredging is an essential step in the EDA phase.

Q11: Data science methodology follows the statistical inference approach.

Q12: When a sample has a Median much smaller than its Mean, then the distribution is Right-Skewed.

Q13: The following Box Plot is Left-Skewed.



Q14: A model has an accuracy of 98%. This accuracy can be misleading when the data is imbalanced.

Q15: A learning model that has No Output(Target) is called a Supervised Model.

Q16: Which of the Python Libraries is powerful for manipulating data frames?

- a) Numpy b) Seaborn c) Pandas d) DataFrame

Q17: The default index in Python starts from

- a) 0 b) 1 c) label d) head

Q18: Which of the following is False about Pandas Series? A Pandas Series is/has

- a) Multi-dimensional b) Homogeneous Data c) Integer/Label indexing d) Elementwise operations

Q19: To read a csv Data file in Python, we use

- a) `pd.read_csv('Data')` b) `pd.read_csv('Data.csv')` c) `pd.read('Data', csv)`

Q20: To obtain a random selection of 10 rows of a data frame (Df), we write

- a) `Df.head(10)` b) `Df.sample(10)` c) `Df.rows(10)` d) `Df.index(10)`

Q21: For a data frame (Df), the data shape, number of NaN, names of columns, and types of each column can be viewed using

- a) `Df.describe()` b) `Df.head()` c) `Df.tail()` d) `Df.info()`

Q22: To view only the columns: 'ID', 'Name', 'Grade', from a data frame (Df), we write

- a) `Df[['ID', 'Name', 'Grade']]` b) `df(('ID', 'Name', 'Grade'))` c) `Df.columns('ID', 'Name', 'Grade')`

Q23: To view the first 10 rows for the ID and Grade columns, we write

- a) `Df.loc[:10, 'ID', 'Grade']` b) `Df.loc[:9, ['ID', 'Grade']]` c) `Df.iloc[:10, 'ID', 'Grade']`

Q24: To filter the rows from a data frame (Df) based on the condition: Age >20, we write

- a) `Df['Age']>20` b) `Df[Df['Age']>20]` c) `Df['Age']>20`

Q25: Which of the following the commands: Group by Department and aggregate Salary using mean?

- a) `Df.groupby('Department')['Salary'].mean()`
b) `Df.groupby('Department', 'Salary', mean)`
c) `Df.groupby['Department', 'Salary'].mean()`
d) `Df.groupby('Department')('Salary').mean()`

Use the given contingency table to answer the following questions:

Q26: The number of the data rows is

- a) 335 b) 65 c) 400 d) 800

Q27: The number of records that correspond to Large Amount

- a) 65 b) 70 c) 140 d) 335

Q28: Of invoices with errors, the proportion of the Small Amount is

- a) 30.77 b) 5.00 c) 10.53 d) 170

Q29: The proportion of invoices with no errors is

- a) 16.25 b) 50.75 c) 83.75 d) 335

Q30: Of Medium Amount invoices, the proportion that has errors is

- a) 71.43 b) 61.54 c) 10.00 d) 28.57

	No Errors	Errors	Total
Small Amount	170	20	190
Medium Amount	100	40	140
Large Amount	65	5	70
Total	335	65	400