Test_1_Propuesto

May 13, 2025

Creado por:

Isabel Maniega

1 Test 1

1.1 Question 1

```
Patricia is analyzing consumer trends in the fashion industry.
She needs to collect data on the types of clothing items that are currently popular.
Which method should Patricia use to gather this information?

a) Conducting in-deph interviews with fashion designers.
b) Analyzing sales data from major retail fashion stores.
c) Surveying consumers on their recent clothing purchases.
d) Reviewing fashion magazines and trend reports.
```

1.2 Solution 1

Solution:

1.3 Question 2

You're responsible for presenting the quarterly performance of your company's social media channels to a group of executives with limited technical knowledge. The key metrics include Total Followers, Engagement Rate, and Click-Through Rate. Here's the performance data for the year:

Quater	Total Followers	Engagement Rate (%)	Clicks-Through Rate (%)
Q1 (Jan-Mar)	20000	5.0	2.5
Q2 (Apr-Jun)	23500	6.0	3.0
Q3 (Jul-Sep)	27000	6.5	3.5
Q4 (Oct-Dec)	30000	7.0	3.8

How should you present this data to ensure clear understanding and actionable insights for your non-technical audience, especially focusing on the distribution and variability of the data? Select the best approach.

a) Use box plots to display the distribution and variability of Total Followers, engagement Rate, and Click-Through Rate for each quarter,paired with clear explanations to help the audience understand the central tendency, spread and outliers in the data.

b) Use a detailed spreadsheet format in the presentation that shows every metric's quarterly data, expecting stakeholders to draw conclusions on their own, demostrating the complexity of the data.

c) Create a complex infographic that combines all metrics using advanced visualizations and technical jargon, assuming the intricate design will impress and engage the executive.

d) Utilize simple line and bar graphs to display trends in Total Followers, Engagement Rate, and Click-Trough Rate, paired with brief explanations that provide context and highlight key takeaways, ensuring the data is easy to understand and actionable.

1.4 Solution 2

Solution:

1.5 Question 3

```
Sophia is tasked with combining quarterly financial data from different departments into a sing
To ensure data accuracy and consistency, what is the most important step Sophia should take?

a) Prioritize data from the department with the hihest revenue.
b) Standardize numerical formats and check for discrepancies in departmental data before mergi
c) Merge all data first and then address any discrepancies afterward.
d) Use the most recent quarter's data template for all department.
```

1.6 Solution 3

Solution:

1.7 Question 4

You conducted a study and calculated a 95% confidence interval for the mean difference between two groups as (2, 8). What does this confidence interval indicate?

a) The true mean difference between the two groups is likely between 2 and 8.

b) The true mean difference the two groups is exactly 2.

c) There is a 95% chance that the true mean difference is between 2 and 8.

d) The study results are inconclusive about the mean difference between the two groups,

1.8 Solution 4

Solution:

1.9 Question 5

```
Michael is studying the migration patterns of birds in a coastal area.
He needs to collect data on the number of different bird species present throughout the year.
What method should Michael employ to collect this data?

a) Reviewing literature on bird migration patterns.
b) Using autommated cameras to record bird movements.
c) Conducting monthly surveys through direct observation.
d) Analyzing satellite images of the area.
```

1.10 Solution 5

Solution:

1.11 Question 6

```
A dataset contains dates of customer transactions, and some dates are listed as '31st February

a) Incomplete data, as the year is not specified.
b) Irrelevant data, as the date may not be needed for analysis.
c) Erroneous data, due to an impossible date.
d) Duplicate data, if the same incorrect date appears multiple times.
```

1.12 Solution 6

Solution:

1.13 Question 7

You are working with a Pandas DataFrame df containing a column named Scores with student scores ranging from 0 to 100. You want to normalize these scores to lie in the range of [0, 1]. Which of the following lines of code will accomplish this task correctly?

a) `df['Scores'] = df['Scores'] / df['Scores'].max()`

b) `df['Scores'] = df['Scores'] / 100`

c) `df['Scores'] = df['Scores'] - df['Scores'].min()) / df['Scores'].max()`

d) `df['Scores'] = df['Scores'].apply(lambda x : (x-df['Scores'].min()) / (df['Scores'].max() - df['Scores'].min()))`

1.14 Solution 7

Solution:

1.15 Question 8

In cleaning a dataset for a health study, a data analyst notices several instances where the 'a
How should these data points be treated?

a) As incomplete data, because the exact ages are not known.
b) As valid data, considering potential recording errors.
c) As erroneous data, given the unrealistic age values.
d) As duplicated data, if the same age appears multiple times.

1.16 Solution 8

Solution:

1.17 Question 9

When integrating time-series data from multiple sensors into a single dataset, what is essentia

a) Aligning all data entries to same time scale and format.
b) Focusing on the sensor with the highest frequency of data collection.
c) Summarizing the data from each sensor before merging to reduce complexity.
d) Choosing one sensor as the primary source and discarding data from others.

1.18 Solution 9

Solution:

1.19 Question 10

Alex is integrating user interaction data from a website and a mobile app to analyze overall us

a) Aggregate all data points into a single average value for simplicity.
b) Validate and align interaction metrics from both sources to ensure they are on a similar.
c) Focus only on the platform with more user interactions for a biased analysis.
d) Manually enter website data to match the volume of app data.

1.20 Solution 10

Solution:

1.21 Question 11

You have conducted a survey on customer satisfaction and created a visualization showing the ov

a) Avoid discussing the methodology and focus on the general trends.
b) Provide a high-level summary with key statistical metrics.
c) Present detailed analysis including survey methodology and statistical significance.
d) Use visual metaphors and analogies to explain the satisfaction levels.

1.22 Solution 11

Solution:

1.23 Question 12

You have analyzed sales data for a retail company and created a visualization showing the reve

a) Present complex statistical analysis with technical jargon.
b) Use clear and concise labels on the visualization to highlight key trends.
c) Provide raw data and detailed charts for audience exploration.
d) Use advanced machine learning algorithms to explain the revenue trends.

1.24 Solution 12

Solution:

1.25 Question 13

You are developing a Python script to process a large dataset and perform various data manipul

a) Using single-letter variable names for improved readability.
b) Writing long and complex functions to minimize the number of lines.
c) Adding descriptive comments to explain the purpose of each function and major code sections
d) Using global variables extensively to avoid passing arguments between functions.

1.26 Solution 13

Solution:

1.27 Question 14

You have analyzed website traffic data and created a visualization showing the user engagement

a) Use technical terms and metrics without explanation.
b) Create a narrative around user behavior patterns and trends.
c) Provide raw data and detailed charts for audience exploration.
d) Focus on the aesthetics of the visualization rather than the insights.

1.28 Solution 14

Solution:

1.29 Question 15

```
You have analyzed a dataset containing customer purchase behavior and created a visualization s

a) Present detailed statistical analysis with technical terms.
b) Use storytelling and simple language to explain the trends and patterns.
c) Provide raw* data and complex charts for audience interpretation.
d) Focus only on the technical aspects without context or storytelling.

*raw: sin procesar
```

1.30 Solution 15

Solution:

1.31 Question 16

```
You are analyzing the performance of an email marketing campaign. Which of the following metric

a) Email Conversion Rate (Email Conversion Rate measures the perentage of recipients who take a
b) Time Spent on website.
c) Email Open Rate.
d) Bounce Rate (Bounce rate is a metric that reprensents the percentage of visitors who enter a
```

1.32 Solution 16

Solution:

1.33 Question 17

```
In a data analysis project, you are aggregating data from various external web sources using Py

a) Limit data collection to a few sources for consistency.
b) Collect data in small batches and validate each batch.
c) Use threading to speed up data collection.
d) Verify data authenticity and integrity as you ingest it.
```

1.34 Solution 17

Solution:

1.35 Question 18

In the context of an ETL (Extract, Transform, Load) process, what is the primary purpose of the

a) Creating visualizations and reports for end-users.
b) Retrieving and reading data from multiple heterogeneous data sources.
c) Loading data into a data warehouse or database.
d) Performing data cleaning and preparation for analysis.

1.36 Solution 18

Solution:

1.37 Question 19

When developing interactive web applications using Dash, how is the concept of 'Persistence' u

a) To constantly update the application's data in real-time.
b) To secure the application against unauthorized access.
c) To maintain the database connection continuosly.
d) To remember the user's choices or data entries across multiple sessions.

1.38 Solution 19

Solution:

1.39 Question 20

You are optimizing a Python script that processes a large dataset. You notice that certain func

a) Implement memoization to cache function results.
b) Rewrite the functions to perform parallel processing.
c) Increase the script's memory allocation for faster conputation.
d) Add more conditional statements to skip redundant computations.

1.40 Solution 20

Solution:

1.41 Question 21

You are analyzing sales data for a retail company, which includes daily sales figures for diffe

a) Summarizing
b) Filtering
c) Sorting

d) Grouping

1.42 Solution 21

Solution:

1.43 Question 22

You are working on a data analysis script that involves multiple data processing steps. What is

a) Writing all code in a single massive script for simplicity.
b) Using short and cryptic function names to save space.
c) Breaking down the script into smaller functions with clear and descriptive names.
d) Embedding documentation within the script's code rather than using external documentation fi

1.44 Solution 22

Solution:

1.45 Question 23

You are debugging a Python script that is producing unexpected output. After thorough examinati

a) Add print statements to log variable values.
b) Comment out the suspected conditional statement.
c) Use the step-by-step debugger with breakpoint.
d) Rewrite the contitional statement using a different syntax.

1.46 Solution 23

Solution:

1.47 Question 24

In a dataset containing customer purchase records, which data validation technique is most suitable for ensuring the accuracy of product prices?

a) Completeness validation
b) Range validation.
c) Consistency validation.
d) Cross-reference validation

1.48 Solution 24

Solution:

1.49 Question 25

```
You are working on a Python script that is running slower than expected due to inefficient code

a) Move the loop outside the main function.
b) Use a generator expression instead of a list comprehension.
c) Cache the precomputed values in a dictionary.
d) Add more nested loops for a parallel procesing.
```

1.50 Solution 25

Solution:

1.51 Question 26

You have a DataFrame named df with the following structure:

```
   ID  Name  Score  Subject
0   1   Tom     90  Math
1   2  Lisa     85  Math
2   1   Tom     92  History
3   2  Lisa     88  History
```

You want to reshape this DataFrame into a format that shows scores by Subject for each individual.
Which of the following code snippets will achieve this?

a) `df.pivot(index='Name', columns='Subject', values='Score')`

b) `df.groupby(['Name', 'Subject']).Score.sum().unstack()`

c) `df.pivot_table(index='Name', columns='Subject', values='Score', aggfunc='mean')`

d) `df.set_index(['Name', 'Subject']).unstack()`

1.52 Solution 26

Solution:

1.53 Question 27

```
Sarah is developing a Python script for data analysis and visualization. She wants to ensure th

a) Using vague variable names to encourage code exploration.
```

b) Writing functions with long and convoluted logic to minimize the number of functions.
c) Dividing the script into smaller functions with clear names and specific responsibilities.
d) Avoiding names comments in the code to keep it concise.

1.54 Solution 27

Solution:

1.55 Question 28

You are debugging a Python script that is intended to calculate the average of a list of numbe

a) Add more test cases to cover a wider range of scenarios.
b) Print the list of numbers before and after the calculation.
c) Rewrite the calculation logic using a different algorithm.
d) Step through the script using a debugger and inspect variable values.

1.56 Solution 28

Solution:

1.57 Question 29

You are working with a small dataset and want to assess your model's performance. Your colleague recommends using k-fold cross-validation. However, you are concerned about the reliability of the evaluation. What cross-validation technique would be most appropriate for your small dataset?

a) Shuffle-Split Cross-Validation
b) Time Series Cross-Validation
c) Leave-One-Out Cross-Validation
d) Stratified k-Fold Cross-Validation

1.58 Solution 29

Solution:

1.59 Question 30

You are tasked with optimizing a Python script that processes a large dataset. During testing,

a) Add print statements to track memory usage.
b) Use a profiler to analyze memory usage patterns.
c) Comments out sections of code to isolate the issue.
d) Rewrite the script using a different programming paradigm.

1.60 Solution 30

Solution:

1.61 Question 31

```
Alex is tasked with optimizing a Python script for data processing. What coding practice should

a) Using global variables extensively to simplify data sharing between functions.
b) Creating functions with overly general names to accommodate multiple functionalities.
c) Implementing meaningful variable names and organizing code into logical sections.
d) Embedding documentation only in external files separate from the code.
```

1.62 Solution 31

Solution:

1.63 Question 32

```
You are working with a dictionary in Python that stores information about students and their s

a) scores['John']
b) scores.get('John')
c) scores['scores']['John']
d) scores.get('John', 0)
```

1.64 Solution 32

Solution:

1.65 Question 33

```
You are debugging a Python script that is supposed to extract specific information from a JSON

a) Rewrite the entire script using a different programming paradigm for better error handling.
b) Utilize try-except blocks to catch and handle specific exceptions that occur during script e
c) Increase the script's memory allocation to prevent runtime errors related to memory exhausti
d) Ignore the errors and proceed with running the script to see if it resolves itself.
```

1.66 Solution 33

Solution:

1.67 Question 34

```
A data analyst is working with a dataset containing numerical values in the ‘age’ column. They

a) Regular expression matching to validate the age format.
b) Applying a lambda function to check for values outside the range.
c) Using the pd.cut() function to categorize ages into bins.
d) Using the pd.to_numeric() function with errors set to 'coerce' and then checking for NaN val
```

1.68 Solution 34

Solution:

1.69 Question 35

You conducted a correlation analysis between two variables and obtained a correlation coefficient of -0.75. What does this correlation coefficient value indicate about the relationship between the variables?

a) There is a strong positive correlation between the variables.

b) There is a moderate positive correlation between the variables.

c) There is a strong negative correlation between the variables.

d) There is no correlation between the variables.

1.70 Solution 35

Solution:

1.71 Question 36

You are tasked with validating a dataset containing ages of customers. Which data validation technique would be most appropriate to ensure the reliability and accuracy of the age data?

a) Completeness validation

b) Format validation.

c) Range validation.

d) Consistency validation

1.72 Solution 36

Solution:

1.73 Question 37

You have employed 5-Fold Cross-Validation on a binary classification problem and received the following accuracy scores for each fold: [0.8, 0.85, 0.9, 0.7, 0.6]. What should be your next course of action?

a) Consider the model to be robust as the accuracy is above 50% for all folds.

b) Increase the number of folds to 10 for a more precise evaluation.

c) Investigate the cause of the variability in the accuracy scores across folds.

d) Pick the best model from the third fold, as it has the highest accuracy.

1.74 Solution 37

Solution:

1.75 Question 38

Suppose you have a DataFrame df with columns Name, Age, Gender, and Salary. Which of the following code snippets will filter the DataFrame to include only rows where Age is more than 25 and Salary is less than 50000, and also sort the resulting DataFrame by Name?

a) `df.filter('Age' > 25 & 'Salary' < 50000).sort_values(by='Name')`

b) `df[(df['Age'] > 25) & (df['Salary'] < 50000)].sort_values('Name')`

c) `df.sort_values('Name').where(df['Age'] > 25 & df['Salary'] < 50000)`

d) `df.query('Age > 25 and Salary < 50000').sort('Name')`

1.76 Solution 38

Solution:

1.77 Question 39

Given the Python code snippet below, which utilizes Matplotlib to generate a plot, what type of chart will be produced?

```
import matplotlib.pyplot as plt

a = [2, 4, 6, 8, 10]
b = [1, 3, 5, 7, 9]

plt.plot(a, b, color='green', linestyle='solid', marker='s',
         markerfacecolor='yellow', mec='blue', linewidth=1.5, alpha=0.9,
         label='Custom Plot')
plt.xlabel('A Axis')
plt.ylabel('B Axis')
```

```
plt.title('Example Plot')
plt.legend(loc='lower left')
plt.show()
```

a) A scatter plot with square markers, green lines, and yellow marker faces.

b) A bar chart with green bars and yellow edges.

c) A line plot with square markers, green lines and yellow marker faces.

d) A pie chart with segment labled "Custom Plot".

1.78 Solution 39

Solution:

1.79 Question 40

You have the following DataFrame containing sales data:

```
df = pd.DataFrame({
    'Month': ['Jan', 'Jan', 'Feb', 'Feb', 'Mar', 'Mar'],
    'Product': ['A', 'B', 'A', 'A', 'B', 'C'],
    'Sales': [100, 150, 200, 50, 300, 400]
})
```

You want to find the total sales for each month. Which code snippet will achieve this?

a) `df.groupby('Month').sum()`

b) `df.groupby('Product').sum('Sales')`

c) `df['Month'].agg({'Sales': 'sum'})`

d) `df.groupby('Month').agg({'Sales': 'sum'})`

1.80 Solution 40

Solution:

1.81 Question 41

You have a DataFrame df with columns 'Quarter', 'Revenue', and 'Expenses'. You are tasked with calculating the quarterly profit margin as (Revenue - Expenses) / Revenue. Which of the following will correctly add a 'Profit Margin' column with these calculated values?

a) `df['Profit Margin'] = (df['Revenue'] - df['Expenses']) / df['Revenue']`

b) `df['Profit Margin'] = df['Revenue'] - df['Expenses'] / df['Revenue']`

c) `df['Profit Margin'] = df['Revenue'] / (df['Revenue'] - df['Expenses'])`

d) `df['Profit Margin'] = df.apply(lambda row: (row['Revenue'] - row['Expenses']) / row['Revenue'], axis=1)`

1.82 Solution 41

Solution:

1.83 Question 42

Your dataset contains monthly sales figures for different products over the past year. You want to visualize the sales trends for each product over time. Which type of visualization is most suitable for this task?

a) Line Plot

b) Box Plot

c) Pie Chart

d) Scatter Plot

e) Bar chat

1.84 Solution 42

Solution:

1.85 Question 43

You are analyzing a dataset containing daily temperature readings from multiple cities over a year. You want to visualize this data to compare the temperature trends across these cities. Which of the following visualization techniques would be most suitable for this purpose?

a) Barc Chart

b) Multiple Line Graphs on the same Plot

c) Scatter Plot

d) Pie Chart

e) Histogram

1.86 Solution 43

Solution:

1.87 Question 44

In your climate study, you're examining the relationship between average yearly sunshine hours and average annual temperature in different countries. Additionally, you want to emphasize countries with the highest and lowest average yearly sunshine hours. What type of data visualization would be most suitable for this analysis?

a) Histogram

b) Scatter Plot with color Coding

c) Bubble Chart with Size Variation

d) Stacked Area Chart

1.88 Solution 44

Solution:

1.89 Question 45

In the context of an ETL (Extract, Transform, Load) process, what is the primary purpose of the 'Extract' phase?

a) Loading data into a data warehouse or database

b) Retrieving and reading data from multiple heterogeneous data sources

c) Creating visualizations and reports for end-users.

d) Performing data cleaning and preparation for analysis.

1.90 Solution 45

Solution:

Creado por:

Isabel Maniega