



Python Institute

PCED Exam

PCED: Python Institute Certified Entry-Level Data Analyst with Python
(PCED-30-02)

Exam Latest Version: 6.0

DEMO Version

Full Version Features:

- 90 Days Free Updates
- 30 Days Money Back Guarantee
- Instant Download Once Purchased
- 24 Hours Live Chat Support

Full version is available at link below with affordable price.

<https://www.directcertify.com/python-institute/pced>

Question 1. (Single Select)

You are analyzing a dataset of customer purchase history. You notice that several entries have ages listed as 'N/A'. Which of the following actions would be most appropriate for handling these missing age values in the context of building a predictive model for customer churn?

- A: Remove the rows with 'N/A' age values entirely.
- B: Replace 'N/A' with the average age of all customers in the dataset.
- C: Replace 'N/A' with the median age of all customers in the dataset.
- D: Create a new category for 'N/A' ages and treat it as a separate value.
- E: Impute the missing ages using a machine learning model trained on the available age data.

Correct Answer: E

Explanation:

Imputing missing age values using a machine learning model is the most appropriate approach in this scenario. This approach takes advantage of the existing patterns in the data to estimate the missing values, preserving more information than simply removing the rows or replacing with a single value. Removing rows may lead to data loss, and replacing with average or median may introduce bias. Creating a separate category for 'N/A' may not be informative in this context. Therefore, using a machine learning model for imputation offers the best balance of data integrity and model accuracy.

Question 2. (Multi Select)

You are analyzing a dataset with a column representing product prices. You identify that some prices are recorded as strings, such as '\$10.99' instead of numerical values. Which of the following approaches would be the most suitable for converting these string values into numerical values while maintaining data integrity?

- A: Use the method to remove the dollar sign and then convert the string to a float.
- B: Use the 'str.strip()' method to remove any leading or trailing spaces and then convert the string to a float.

C: Use the 'str.split()' method to split the string at the decimal point, convert each part to a float, and then combine them into a single float.

D: Use the function with the 'errors='coerce'S parameter to attempt conversion and set invalid values to NaN.

E: Use e 'str.replace()' method to replace the dollar sign with an empty string and then convert the string to a float, handling any errors with a 'try-except block.

Correct Answer: A, E

Explanation:

Both option A and option E are suitable approaches for converting string values into numerical values while preserving data integrity. Option A utilizes the 'str.replace()' method to directly remove the dollar sign and then converts the string to a float. This approach is straightforward and efficient. Option E also uses the 'str.replace()' method to remove the dollar sign and then converts the string to a float, but it includes a 'try- except' block to handle any potential errors during the conversion process. This approach is more robust and ensures that any invalid or unexpected values are handled gracefully. Option B is not the most suitable approach since it only removes spaces, not the dollar sign. Option C is complex and may not be necessary for this specific scenario. Option D will convert valid values to numeric but will leave invalid values as NaN, which may not be desirable in this case.

Question 3. (Multi Select)

You are analyzing a dataset of customer feedback ratings for a product. The ratings range from 1 to 5, with 5 being the highest. You notice a significant number of ratings clustered around 4 and 5, with very few ratings below 3. This indicates a potential bias in the data'. Which of the following techniques could be applied to address this bias and get a more accurate representation of customer sentiment?

A: Remove the outlier ratings below 3, as they likely represent negative reviews and can skew the analysis.

B: Normalize the ratings using a standard scaling method, such as Min-Max scaling, to adjust the range and distribution.

C: Collect additional data points, such as customer reviews and comments, to gain a deeper understanding of the reasons behind the high ratings.

D: Apply a transformation technique, such as logarithmic transformation, to compress the range

of high ratings and emphasize the lower ratings.

E: Use a robust statistical method, such as the median instead of the mean, to calculate the average rating and minimize the impact of outliers.

Correct Answer: C, D, E

Explanation:

The scenario describes a potential bias towards positive feedback. Here's why the chosen options are the most suitable: - C: Collect additional data : Gathering customer reviews and comments can provide context for the high ratings. This helps you understand if the ratings reflect genuine satisfaction or other factors like marketing bias or incentives. - D: Logarithmic Transformation : A logarithmic transformation can compress the range of high ratings and emphasize the lower ratings. This can help to create a more balanced distribution and provide a better representation of the overall sentiment. - E: Robust Statistics : Using the median instead of the mean can help minimize the impact of outliers (the high ratings) and provide a more accurate representation of the typical customer sentiment. Option A is incorrect because removing outliers can lead to data loss and distort the analysis. Option B, normalization, would rescale the data but wouldn't address the underlying bias.

Question 4. (Single Select)

You are analyzing a dataset of customer purchase history. You notice a few entries with ages over 150 years old. What is the most likely cause of this data error and how should it be addressed?

A: Data entry error, replace with the median age

B: System malfunction, remove the entries as they are unusable

C: Data inconsistency, flag the outliers and investigate further

D: Data corruption, replace the values with a random number within the acceptable range

E: Data transformation error, apply a log transformation to normalize the data

Correct Answer: C

Explanation:

The most likely cause of such extreme outliers is data inconsistency. It's crucial to investigate further to understand the reason for the error. Simply removing or replacing the data without investigation might lead to losing valuable information. Flagging the outliers and performing a deeper analysis is the best approach in this scenario.

Question 5. (Single Select)

You have a dataset with several columns containing categorical variables like 'Gender', 'Marital Status', and 'Education'. Which Python library is most suitable for encoding these categorical variables into numerical ones for use in machine learning algorithms, and why?

A: NumPy, as it provides efficient array operations for handling numerical data.

B: Pandas, as it offers powerful data manipulation capabilities including dummy variable creation.

C: Scikit-learn, as it contains a variety of machine learning algorithms for data analysis.

D: Matplotlib, as it provides functions for visualizing data, helping to identify categorical patterns.

E: Seaborn, as it simplifies data visualization with attractive statistical plots.

Correct Answer: B

Explanation:

Pandas is the most suitable library for encoding categorical variables. It offers functions like 'get_dummies' which efficiently create dummy variables (one-hot encoding) for categorical features. This transformation is crucial for most machine learning algorithms that require numerical input.



Full version is available at link below with affordable price.

<https://www.directcertify.com/python-institute/pced>

30% Discount Coupon Code: LimitedTime2025

*** 100% MONEY BACK GUARANTEED**
CERTIFICATION EXAMS
STUDY GUIDES

FREE TRIAL

*** Product Features**

- * 100% Success in the Final Exam
- * 90 Days Free Updates
- * Latest Exam Q/A
- * 24/7 Customer Support
- * Practice Exams

*** Free Demo for Practice Test & PDF**

50K Plus Satisfied Customers

VISA AMERICAN EXPRESS DISCOVER G Pay